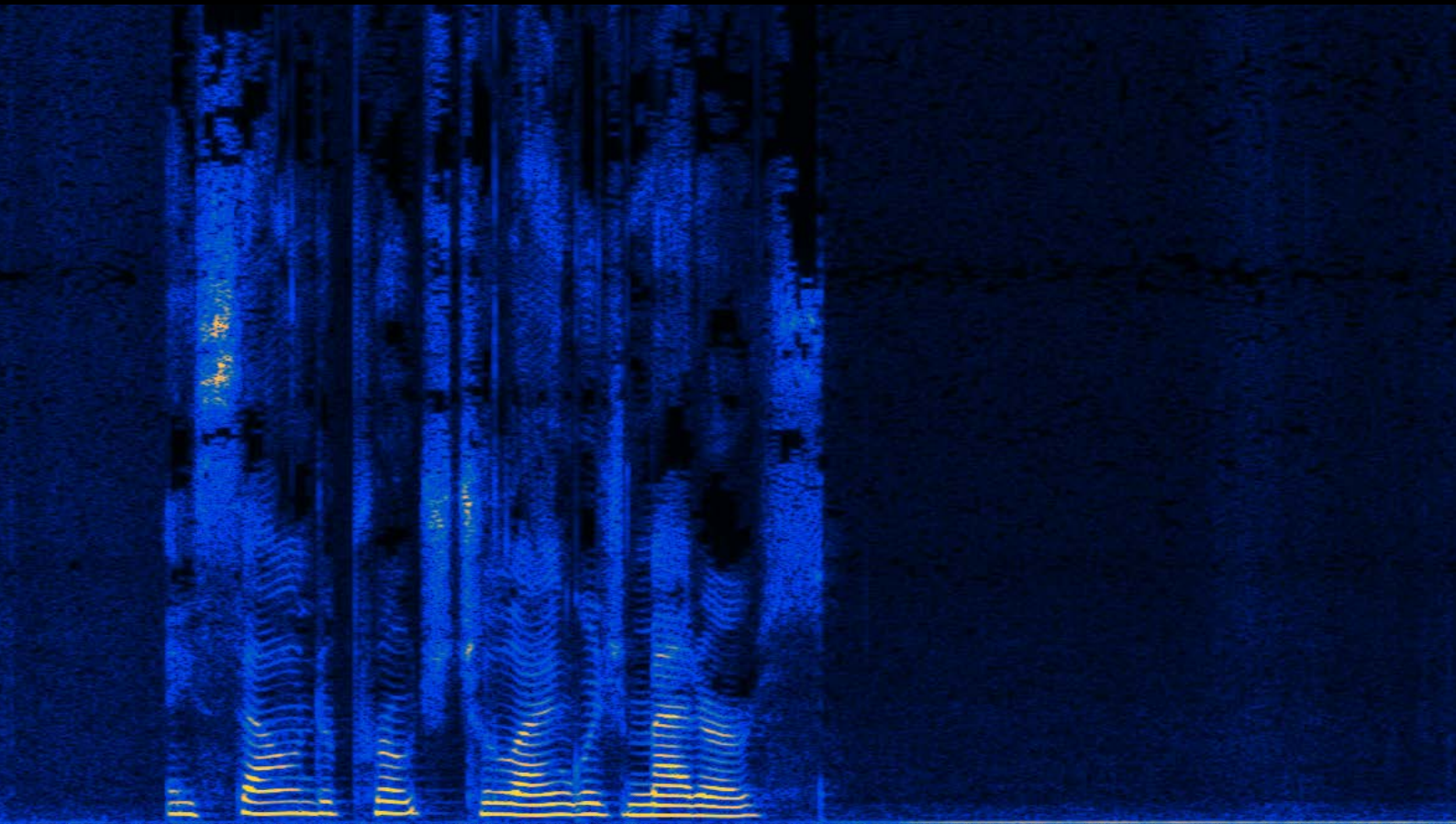


# A Sound Approach to Language Matters

In Honor of Ocke-Schwen Bohn



Edited by  
Anne Mette Nyvad  
Michaela Hejná  
Anders Højen  
Anna Bothe Jespersen  
Mette Hjortshøj Sørensen

*A Sound Approach to  
Language Matters*

*In Honor of Ocke-Schwen Bohn*





# *A Sound Approach to Language Matters*

*In Honor of Ocke-Schwen Bohn*



Edited by  
Anne Mette Nyvad, Michaela Hejná, Anders Højen,  
Anna Bothe Jespersen & Mette Hjortshøj Sørensen

Department of English  
School of Communication & Culture  
Aarhus University  
2019

# *A Sound Approach to Language Matters*

*In Honor of Ocke-Schwen Bohn*

Edited by  
Anne Mette Nyvad  
Michaela Hejná  
Anders Højen  
Anna Bothe Jespersen  
Mette Hjortshøj Sørensen

© The authors & Aarhus University, 2019

Cover: Anne Mette Nyvad & Kirsten Lyshøj  
Layout: Kirsten Lyshøj

E-ISBN: 978-87-7507-440-2

**CONTENTS** – all contributions have undergone peer review:

## **SEGMENTS**

(Handling editor: Michaela Hejná)

- Catherine Best, Cinzia Avesani, Michael Tyler & Mario Vayra:**  
PAM Revisits the Articulatory Organ Hypothesis:  
Italians' Perception of English Anterior and Nuu-Chah-Nulth  
Posterior Voiceless Fricatives 13
- Rikke Bundgaard-Nielsen & Brett Joseph Baker:**  
Paa, Paa Plack Sheep: Discrimination of L2 Stop Voicing  
Contrasts in the Absence of L1 Stop Voicing Distinctions 41
- Sidsel Rasmussen & Mengzhu Yan:**  
Acoustic Comparison of Mandarin and Danish Postalveolars 65
- D. H. Whalen:**  
Normalization of the Natural Referent Vowels 81

## **PERCEPTION OF ACCENT**

(Handling editor: Mette Hjortshøj Sørensen)

- Angélica Carlet & Juli Cebrian:**  
Assessing the Effect of Perceptual Training on L2 Vowel  
Identification, Generalization and Long-term Effects 91
- Denise Cristina Kluge:**  
Perception of Brazilian Portuguese Nasal Vowels by  
Danish Listeners 121
- Mette Hjortshøj Sørensen:**  
Accent Matters in Perception of Voice Similarity 135

## **BETWEEN SOUNDS AND GRAPHEMES**

(Handling editor: Mette Hjortshøj Sørensen)

- Henrik Jørgensen:**  
The Four Troublemakers in Danish Orthography 151



**Johanna Wood:**  
Northumbrian Rounded Vowels in the Old English Gloss to the  
Lindisfarne Gospels 167

## **PROSODY**

(Handling editor: Anna Bothe Jespersen)

**Congchao Hua, Bin Li & Ratreë Wayland:**  
Native and Non-native English Speakers' Assessment of  
Nuclear Stress Produced by Chinese Learners of English 187

**Angela Cooper, Yue Wang & Dawn M. Behne:**  
Effects of Semantic Information and Segmental Familiarity on  
Learning Lexical Tone 211

**Míša Hejrná & Anna Jespersen:**  
Focus on Consonants: Prosodic Prominence and  
the Fortis-Lenis Contrast in English 237

**Goun Lee & Allard Jongman:**  
Production and Perception of Korean Word-level  
Prominence by Older and Younger Korean Speakers 271

**Yingjie Li, Goun Lee & Joan A. Sereno:**  
Comparing Monosyllabic and Disyllabic Training in  
Perceptual Learning of Mandarin Tone 303

**Oliver Niebuhr:**  
Pitch Accents as Multiparametric Configurations of  
Prosodic Features – Evidence from Pitch-accent Specific  
Micro-rhythms in German 321

## **MORPHOLOGY & SYNTAX**

(Handling editor: Anne Mette Nyvad)

**Laura Winther Balling:**  
A Sound Approach to Text Processing: Between Experiments  
and Experience 355

**Ken Ramshøj Christensen:**  
On the Need for Experimental Syntax 373

<b>Camilla Søballe Horslund:</b> An Experimental Approach to the Conrad Phenomenon	389
<b>Johannes Kizach:</b> Ungrammatical Sentences Have Syntactic Representations too	423
<b>Sten Vikner:</b> Why German is not an SVO-language but an SOV-language with V2	437
<b>Anne Mette Nyvad:</b> The Logical Problem of Language Acquisition Revisited: Insights from Error Patterns in Typical and Atypical Development	449
 <b>SECOND LANGUAGE ACQUISITION</b> (Handling editor: Anders Højen)	
<b>Joan C. Mora &amp; Ingrid Mora-Plaza:</b> Contributions of Cognitive Attention Control to L2 Speech Learning	477
<b>James Emil Flege:</b> A Non-critical Period for Second-language Learning	501
<b>Anders Højen:</b> Improvement in Young Adults' Second-language Pronunciation after Short-term Immersion	543
<b>Linda Polka, Yufang Ruan &amp; Matthew Masapollo:</b> Understanding Vowel Perception Biases – It's Time to Take a Meta-analytic Approach	561
<b>Anja Steinlen, Thorsten Piske, Sophia Karmeli &amp; Christine Mooshammer:</b> Second and Third Language Immersion Students' Pronunciation in Foreign Language English Oral Reading	583
<b>Michael Tyler:</b> PAM-L2 and Phonological Category Acquisition in the Foreign Language Classroom	607

## **Preface**

We are immensely pleased to dedicate this Festschrift with its 27 contributions to Ocke-Schwen Bohn on the occasion of his 65<sup>th</sup> birthday, May 14, 2018, as a way for its 47 authors to honor and thank a scholar and a man who is larger than life.

Ocke's career has led him far and wide. His interest in phonetics was sparked when he studied English at Kiel University in Germany and he credits this to his teacher's talent, enthusiasm and humor. For two decades and counting, he has paid this approach forward to students at Aarhus University who appreciate exactly those qualities in him as a lecturer on phonetics and phonology. However, after working as a research assistant to Henning Wode and studying at the University of California at Berkeley for a year, Ocke set his sights on L2 syntax, on which he wrote both his Master's thesis and his PhD dissertation. The romance with syntax ended when Jim Flege offered him a postdoc position at the University of Alabama at Birmingham and taught Ocke how to carry out L2 speech research. During this time, he formed research collaborations with (among others) Linda Polka and Catherine Best, who are also contributors to this volume. He was then Director of Kiel University Language Laboratories until he was appointed Professor of English Linguistics at Aarhus University in 1996 and settled down in the city of Aarhus with his wife Annette and their three daughters.

One predominant topic that pervades all of Ocke-Schwen Bohn's scientific research activities spanning more than 35 years is speech perception, and he has left several significant marks on his research field, recently manifested in his overview chapter on "Cross-Language and Second Language Speech Perception" for *The Handbook of Psycholinguistics* (Bohn, 2018).

One of Ocke's greatest legacies is the Desensitization Hypothesis, which Bohn (1995, p. 294) explains as follows: "[W]henver spectral differences are insufficient to differentiate vowel contrasts because

previous experience did not sensitize listeners to these spectral differences, duration differences will be used to differentiate the non-native vowel contrast". As an indication of its impact, this hypothesis has repeatedly been challenged. Although the original finding at its base (Bohn & Flege, 1990) has been replicated several times, there is disagreement about the underlying mechanism. However, in typical fashion, Ocke says that it is much more important to him to provide *food for thought* than to be right all the time.

In keeping with his interest in perceptual asymmetries, Ocke has also developed the Natural Referent Vowel (NVR) framework, in collaboration with Linda Polka (Polka & Bohn, 2003). According to the NVR, vowels at the peripheries of the human vowel space have a privileged status in infant and L2 perception.

Ocke's most cited work, however, is an article that he wrote back in 1997 with Jim Flege and Sunyoung Jang, the main argument of which is that experience affects the L2 perception and production accuracy of adult learners, which also varies as a function of L1 background. This is one instance where Ocke argues against the Critical Period Hypothesis for L2 learning; Bohn (2005) is another. He passionately believes that learning continues throughout a person's lifetime, and he has played a prominent role in the field of L2 speech with regard to the debunking of the myth that adults cannot learn a new language.

These and other accomplishments have earned Ocke a great deal of respect in the international scientific community, but one of Ocke's arguably less conceptual achievements is that he is the only person who has attended every single meeting of the *International Symposium on the Acquisition of Second Language Speech (New Sounds)* since its inception in 1990, and he was a proud host and organizer of the event at Aarhus University in 2016.

The contributions in this Festschrift were written by Ocke's current and former PhD students, colleagues and research collaborators. The Festschrift is divided into six sections, moving from the smallest building blocks of language, through gradually expanding objects of linguistic inquiry to the highest levels of description – all of which have formed a part of Ocke's career, in connection with his teaching and/or his academic productions: "Segments", "Perception of Accent", "Between Sounds and Graphemes", "Prosody", "Morphology and Syntax" and "Second Language Acquisition". Each one of these illustrates a sound approach to language matters.



With this Festschrift, we would like to express our gratitude to Ocke for his significant and lasting contributions to the field of L1 and L2 speech perception and production, and for being generous with his time, his humor, comments and scholarly as well as personal advice. We would also like to thank Annette Bohn, the School of Communication and Culture, Faculty of Arts, Aarhus University, and Kirsten Lyshøj for her indispensable help with the typesetting and the lay-out of the book. Lastly, the authors and peer-reviewers of the contributions in this Festschrift deserve a special thank you.

Anne Mette Nyvad,  
on behalf of the editorial team,  
Aarhus, 2018.

### References

- Bohn, O.-S. & Flege, J.E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, 11, 303-328.
- Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279-304). Timonium, MD: York Press.
- Bohn, O.-S. (2005). A fond farewell to the Critical Period Hypothesis for non-primary language acquisition. In A. Saleemi, O.-S. Bohn & A. Gjedde (Eds.), *In search of a language for the mind-brain: Can multiple perspectives be unified?* (pp. 285-310). Århus, Aarhus Universitetsforlag.
- Bohn, O.-S. (2018). Cross-Language and Second Language Speech Perception. In E.M. Fernández & H.S. Cairns (Eds.), *The Handbook of Psycholinguistics* (pp. 213-239). Wiley.
- Flege, J. E., Bohn, O.-S. & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437-47.
- Polka, L. & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41(1), 221-231.

# SEGMENTS

Handling editor: Michaela Hejná



## **PAM Revisits the Articulatory Organ Hypothesis: Italians' Perception of English Anterior and Nuu-Chah-Nulth Posterior Voiceless Fricatives**

Catherine T. Best<sup>1</sup>, Cinzia Avesani<sup>2</sup>, Michael D. Tyler<sup>1,3</sup> and Mario Vayra<sup>4</sup>

<sup>1</sup>MARCS Institute, Western Sydney University

<sup>2</sup>Consiglio Nazionale delle Ricerche, Italia

<sup>3</sup>School of Social Sciences & Psychology, Western Sydney University

<sup>4</sup>Università di Bologna, Italia

### **Abstract**

We perceive non-native speech in terms of similarities to our native phonology, which makes many non-native contrasts difficult to discriminate (e.g., Speech Learning Model [SLM]). However, discrimination is poor mainly when contrasting non-native consonants are both mediocre exemplars of the same native consonant. Discrimination is much better if they are similar to different native consonants, and good if they are nativelike versus deviant exemplars of the same native consonant (Perceptual Assimilation Model [PAM]). The Articulatory Organ Hypothesis (AOH) offers orthogonal predictions that consonants produced by different articulators should be discriminated better than consonants using the same articulator. To compare these models, we tested Italian listeners on non-native English and Nuu-Chah-Nulth fricative contrasts differing in perceptual assimilation, articulatory organs, and articulator use in Italian. Results support PAM and pose challenges for AOH and SLM.

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 13-40). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.



## 1. Introduction

As adults we apprehend the consonants and vowels in speech with a “native ear.” This selective perceptual tuning, shaped by a lifetime of native language (L1) conversational experience, makes comprehending L1 verbal messages largely automatic and fluid, given rapid yet accurate recognition of native spoken words. But this exquisitely supportive tuning of speech perception has a dark side: it leaves us mistuned for reception of the unfamiliar phonetic properties and phonological functions of non-native consonants and vowels, i.e., speech segments that play no role in our own phonological system despite being gainfully employed by other languages. Unsurprisingly, this non-optimal perception of foreign phones hinders second language (L2) speech learning, both for L2 perception and production. And it persists in making verbal comprehension in a later-acquired language slower, more effortful and more easily disrupted than native speech comprehension, even if the listener has become reasonably fluent in the language (see Lecumberri, Cooke, & Cutler, 2010).

Theoretical and empirical investigations into native attunement of speech perception have primarily addressed how experience with a given language or lack thereof influences categorization and discrimination of minimal segmental contrasts, i.e., pairs of consonants or vowels that differ by a single critical phonetic feature that is contrastive in a given language. Moreover, that work has focused largely on “first encounters” of non-native contrasts by listeners naïve to the stimulus language and the target contrasts. However, it is complemented by studies of L2 speech perception by late learners, who come to the task with substantial L1 biases.

In this chapter we compare and contrast three current theoretical models with respect to their hypotheses about the nature of similarities and differences between non-native speech contrasts and those of the listener’s native language that shape the perception of non-native speech. We go on to provide findings from a novel study designed to compare those hypotheses, and we discuss the theoretical implications of our findings.

We turn first to the aspects of cross-language perceptual research that are most relevant to those theoretical comparisons and the study we report here. Research and theory on non-native speech perception by naïve adult listeners has from early days focused on their difficulties with categorizing and discriminating minimal phonetic contrasts from unfamiliar languages. The classic proposition that adults possess a native-language *phonological filter* (or sieve) that results in a kind of *phonological deafness* to non-native speech contrasts, as originally

posited in the 1930's (Polivanov, 1931/1974; Trubetzkoy, 1939/1969), has been generally accepted based on evidence of naïve listeners' perceptual difficulties with many non-native phonetic contrasts (e.g., Abramson & Lisker, 1970, 1973; Dupoux & Peperkamp, 2002; Iverson et al., 2003; MacKain, Best, & Strange, 1981; Miyawaki et al., 1975; Polka, 1991, 1992; Strange, 1995; Tsao, Liu, & Kuhl, 2006; Werker, Gilbert, Humphrey, & Tees, 1981), as well as of similar difficulties even in early L2 bilinguals (e.g., Sebastian-Gallés & Soto-Faraco, 1999).

However, the stimulus contrasts used in those studies appear to have all been of one particular type, namely cases in which the phonetic characteristics of the contrasting non-native phones align both of them to a single native phoneme. To naïve listeners, these non-native phones are perceived as equally good or poor exemplars of that one native phoneme. Such a narrow range of target stimuli may have led to only partial understanding of the role of experience in non-native speech perception. Findings published since then support that possibility, indicating that non-native phonemes are not all equally difficult to categorize and non-native contrasts are not all equally difficult to discriminate. Performance on both types of tasks is seen to vary when a wider range of non-native phonemes and contrasts has been used.

In light of that variation, several theoretical models of adults' cross-language speech perception have been proposed, which offer a richer, more nuanced view of language-specific perceptual attunement than is captured by the classic phonological deafness concept. We consider here the two models that are most relevant to the perceptual study reported in this chapter<sup>1</sup>: The Perceptual Assimilation Model (PAM: Best, 1995; e.g.,

<sup>1</sup> Other models of cross-language speech perception, while also influential, do not apply as straightforwardly to our reported study on adults' perception of two types of non-native fricative contrasts. Three such models focus on developmental changes in infant rather than adult speech perception as a result of language experience: WRAPSA (Word Recognition And Phonetic Structure Acquisition: Jusczyk, 1993, 1997), NLM (Native Language Magnet: e.g., Kuhl 1993a, b; NLM<sup>c</sup> = expanded: Kuhl et al., 2008) and PRIMIR (Processing Rich Information from Multidimensional Interactive Representations: Werker & Curtin, 2005; Curtin, Byers-Heinlein & Werker, 2011). Three other models address adult cross-language speech perception more centrally but have focused specifically on vowel perception: NRV (Natural Reference Vowels: Polka & Bohn, 2003, 2011), L2LP (Second Language Linguistic Perception: e.g., Escudero & Boersma, 2004 e.g., Leussen & Escudero, 2015) and ASP (Automatic Selective Perception: Strange, 2011, e.g., Strange & Shafer, 2008).

Best, McRoberts & Sithole, 1988; Best, McRoberts & Goodell, 2001) and the Speech Learning Model (SLM: Flege, 1995, 2003, 2007; e.g., Bohn & Flege, 1990, 1993; Guion, Flege, Akahane-Yamada, & Pruitt, 2000).

PAM was originally created to account for variations in perception across a wider range of types of speech contrasts by listeners of a range of L1s who are completely naïve to the target language and specific contrasts being tested. It has since been extended to address experience-related changes in perception and production of L1, L2 and/or unfamiliar speech contrasts by L2 learners (PAM-L2: Best & Tyler, 2007; e.g., Bundgaard-Nielsen et al., 2011a, b, 2012) and bilinguals (Antoniou et al., 2010, 2011, 2012, 2013; Krebs-Lazendic & Best, 2013). PAM's core principle is *perceptual assimilation*, i.e., the idea that listeners have a strong tendency to perceive unfamiliar non-native phones as exemplars of their L1 phonemes, a tendency grounded in detecting articulatory phonetic and/or phonological similarities to them. If the listener perceives a non-native phone as an acceptable exemplar of a single native phoneme, it is *Categorized*. If a non-native phone is instead perceived to have weaker similarities spread across two or more L1 phonemes, it is an *Uncategorized* consonant or vowel. Very rarely, a non-native phone will fail to be perceived as having any similarity to any native phonemes and will remain *Non-Assimilated*, i.e., be heard as a non-speech sound, as is the case for click-language-naïve English speakers' perception of southern African click consonants (Best, McRoberts & Sithole, 1988; Best, Traill, Harrison, Carter, & Faber, 2003).

When two contrasting non-native phones are each categorized to a different L1 phoneme, this constitutes *Two Category* (TC) assimilation, and discrimination is predicted to be excellent. If instead the members of a non-native contrast are both categorized to the same single L1 phoneme, they may be perceived as equally good or poor exemplars of it (*Single Category* assimilation: SC) or one may be a perceptibly poorer fit than the other (a *Category Goodness* difference in assimilation: CG). Discrimination of CG contrasts is predicted to be very good but significantly lower than for TC contrasts, whereas SC contrasts are predicted to be poorly discriminated. If one or both members of a non-native contrast are uncategorized (UC or UU, respectively), discrimination performance level will depend on the subtype of uncategorized assimilation(s) involved, for example, whether or not the contrasting non-native phones show overlap in the L1 phonemes to which similarities are perceived (see Faris, Best, & Tyler, 2016; Faris, Tyler, & Best, 2018). PAM-L2 (Best & Tyler, 2007) makes the case that L2 learning is most likely to result in improved categorization and discrimination of

L1 contrasts that were initially CG or uncategorized assimilations. Note that discrimination of non-native contrasts is now assumed to be better for non-overlapping than overlapping assimilations, within each of the relevant contrast assimilation types: TC, UC and UU (see Faris, Best, & Tyler, 2016; Fenwick, Best, David, & Tyler, 2017; Tyler, Best, Faber, & Levitt, 2014).

Whereas PAM's central aim is to account for variations in non-native speech perception, SLM instead aims to understand the factors that give rise to foreign accent in L2 speech production. Still, SLM makes strong perceptual assumptions, arguing that the most important source of foreign accent is L1 biases in the speaker's perception of L2 speech. A core SLM premise is that L2 speech production can only be as accurate as L2 speech perception permits. SLM posits that L1 perceptual biases lead to *equivalence classification* of L2 phones as being either *identical*, or *similar*, or *new* with respect to L1 phonemes. *Identical* L2 phones pose no difficulty for perception or production, of course, as they correspond well to L1 phonemes. And although *new* L2 phones may pose some difficulties initially, the model predicts that they will be fairly easily established as new, separate L2 categories in both perception and production. In contrast, SLM predicts that equivalence classification of *similar* L2 phones to L1 phonemes results in a persisting L1 perception bias and L1-accented production.

Thus, PAM and SLM have somewhat different yet overlapping and/or complementary foci and conceptual principles. While PAM's central goal is to account for how L1 experience shapes speech perception, particularly of non-native minimal consonant contrasts by naïve listeners, SLM's is to understand the factors contributing to accented speech production by L2 learners/bilinguals, with particular focus on vowels as individual categories. Nonetheless, although the primary foci of the two models are complementary, their secondary emphases bring them back to common ground. SLM considers L1 influences on non-native speech *perception* to be the most important contributor to accented L2 speech production and has fostered investigations of L2 perception (e.g., Bohn & Flege, 1990). Conversely, PAM has been extended to address speech *production* (e.g., Antoniou, Best, Tyler, & Kroos, 2010, 2011; Bundgaard-Nielsen, Best, Kroos, & Tyler, 2012) as well as perception (e.g., Antoniou et al., 2012, 2013; Bundgaard-Nielsen et al., 2011a, b; Krebs-Lazendic, & Best, 2013). In theoretical terms, both PAM and SLM posit that non-native speech is perceived in relation to native (L1) phonemes. Moreover, it is far from



obvious whether their proposed processes of perceptual assimilation and equivalence classification, respectively, differ conceptually very much if at all.

The models do differ, however, in their assumptions about the nature of speech information that perceivers use in the L2aL1 process. PAM posits that the process relies on perceiving information about the articulatory gestures that produced the phones, whereas SLM assumes that it relies on acoustic-phonetic similarities between L2 and L1 phones. Other points of relative difference are that SLM investigations have focused more on vowel than consonant perception, and on individual phonetic rather than on minimal contrasts, whereas PAM research has specifically addressed contrasts and has examined consonant perception more than vowels. Nonetheless, some SLM studies have examined consonants (e.g., Bohn & Flege, 1993), while some PAM studies have addressed perception of vowel contrasts, both from other languages (Bundgaard-Nielsen et al., 2011a, b; Faris, Best, & Tyler, 2016, 2018; Tyler, Best, Levitt, & Faber, 2014) and from other L1 regional accents (Best et al., 2013, 2015a, b; Shaw et al., 2014, 2018). PAM has also been applied to perception of non-native lexical tone contrasts (Hallé, Chang, & Best, 2004; Reid et al., 2015; So & Best, 2010, 2011, 2014).

Consideration of lexical tone perception by naïve listeners of non-tone L1s raises an important question that has not been directly addressed by either PAM or SLM: How might perceptual assimilation/classification work in cases where the non-native contrast uses articulatory/acoustic-phonetic properties that are not employed for segmental contrasts in the listeners' L1? This is one of the questions addressed in the study we report in this chapter. Whereas tone languages engage laryngeal mechanisms to produce fundamental frequency (and sometime voice quality) differences that serve as sub-lexical phonological contrasts that are analogous to minimal segmental contrasts between consonants or vowels, non-tone languages only use tonal patterns at higher, suprasegmental prosodic levels in the phonological hierarchy (e.g., stress, accent, and phonological and intonational phrase boundaries). This means that non-tone language speakers cannot assimilate non-native lexical tones to L1 segments; they may instead perceive them in relation to higher-level prosodic patterns in their L1. Neither SLM nor PAM were designed to address this type of phonological tier discrepancy between the non-native target items and the most likely L1 referent categories (see Best, 2019). Indeed, the phonological tier mismatch is reflected in the performance of naïve non-

tone L1 listeners on PAM-based perceptual tests with non-native lexical tone contrasts, where their assimilations to L1 prosodic categories have been fairly weak while conversely their discrimination of tone pairs has been better than expected from those assimilation patterns (Hallé, Chang, & Best, 2004; Reid et al., 2015; So & Best, 2010, 2011, 2014).

But what might happen when there is *not* a phonological tier mismatch? How might listeners perceive non-native segmental contrasts, e.g., consonants, that use articulatory/acoustic-phonetic properties not employed in their L1 phonology at *either* the segmental or suprasegmental level? If we extrapolate from SLM principles, it seems likely that such consonants would not be equivalence classified as either *identical* or as *similar* to even the acoustically closest L1 consonants because they would nonetheless be too distant from all native consonants; they would instead be perceived as *new* consonants. While they should therefore be easily distinguished from any L1 consonants, it is not clear from SLM whether contrasting pairs of such non-native consonants would be easily discriminated from *each other*, because its principles focus on individual L2 phones in relation to L1 phonemes, not on discrimination of L2 contrasts. And although PAM directly addresses perception of non-native contrasts, it has not explicitly considered how the assimilation may be affected when the articulators involved in the non-native contrast are not employed in the L1. Will such non-native phones be categorised to the most articulatorily similar L1 consonants, or instead more ambiguously assimilated as uncategorized consonants (or possibly even Non-Assimilated, i.e., heard as nonspeech)? And how would discrimination be expected to be affected by these differing possibilities? These questions are examined by the study we report here.

The Articulatory Organ Hypothesis of infant speech perception (AOH: Goldstein & Fowler, 2003; see also Best, Goldstein, Tyler, & Nam, 2016) could potentially offer some more straightforward predictions, however, if we extend it to adult non-native consonant perception. Originally designed to predict developmental changes in infants' perception of native and non-native phonetic contrasts as a result of experience (Studdert-Kennedy & Goldstein, 2003; see Best & McRoberts, 2003), the AOH posits that between-organ articulatory contrasts are easy to distinguish perceptually, even if they are non-native (do not occur in the infant's environment), whereas within-organ contrasts are more difficult to discriminate and to learn even if they occur in native speech. In between-organ contrasts the contrasting consonants

use different primary articulators, e.g., the ejective stops of Tigrinya, /pʰ/ (lips) versus /tʰ/ (tongue tip), whereas in within-organ contrasts the consonants use the same primary articulator but with contrasting place, manner or voicing, e.g., the Hindi dental versus retroflex coronal stops /ɖ/-/ɖʰ/ (tongue tip for both, but at two contrasting places). A few speech perception studies have tested the AOH with infants, with mixed results (*supported*: Best & McRoberts, 2003; *compatible*: Kuhl et al., 2006; Polka, Colantonio, & Sundara, 2001; *not supported*: Tyler, Best, Goldstein, & Antoniou, 2014). Adults, however, with their much greater L1 experience, might possibly show more clearly differentiated perceptual responses to non-native within- versus between-organ contrasts, especially for articulatory organs not employed distinctively in their L1.

To examine the sets of questions raised above, a listener language and target stimulus languages were needed for which one set of non-native consonant contrasts uses articulatory organs employed in the listeners' native language, while another set uses articulatory organs not employed in their language. Italian meets these requirements with respect to two sets of non-native voiceless fricative place of articulation distinctions. Regarding the native-articulators set, Standard Italian has a series of place contrasts among anterior voiceless fricatives that employ the lips (labiodental /f/) and the tongue tip (lamino-alveolar or dental /s/ [respectively, Mioni, 2001, p. 157; Bertinetto & Loporcaro, 2005], and palato-alveolar /ʃ/, which also has secondary tongue body and lip constriction). English offers a set of voiceless anterior fricatives using the lips and/or tongue tip, /f, θ, s, ʃ/, which adds an interdental place of articulation for tongue tip constrictions that is lacking in Italian (no /θ/). Thus, the English series offers two non-native contrasts for which the primary articulators are nonetheless used in Italian, /f/-/θ/ and /θ/-/s/. Those two pairs also provide the required between-organ (/f/-/θ/: lips vs. tongue tip) versus within-organ contrasts (/θ/-/s/: both tongue tip). The remaining minimal-place contrast (/s/-/ʃ/) we define here as a mixed/overlapping organ contrast, given that both of these fricatives use tongue tip but only /ʃ/ also involves constriction of tongue body and lips.

For the non-native-articulators set, that is, consonants that use articulators not employed in Italian, we chose the Nuu-Chah-Nulth (a First Nations Wakashan language, British Columbia, Canada) four-way series of posterior voiceless fricatives, /x, χ, ɦ, h/ (velar, uvular, pharyngeal, glottal), in which the primary articulatory organs are either not used at all in Italian phonological contrasts or are not used for fricative manner. These posterior

fricatives provide three non-native minimal-place contrasts. Within-organ /x/-/χ/ both use the tongue body, which is only employed in Italian for velar stops, not fricatives; Italian does not employ the uvular place for any consonants. Nuu-Chah-Nulth /h/-/h/ is a between-organ contrast, for which /h/ employs the tongue body plus tongue root (see Carlson & Esling, 2003) while the articulator for /h/ is the glottis (vocal cords). Italian does not use the tongue root (pharyngeal constriction) for consonant contrasts, nor does it employ the glottis as an active articulator for fricatives, for which it does not have voicing distinctions. The final minimal-place contrast, /χ/-/h/, is mixed/overlapping organ (tongue body vs tongue body+root) and involves places of articulation and an organ (tongue root) that are not employed in Italian, as well as the fricative manner that is not used in Italian posterior to the hard palate.

We examined native Italian listeners' assimilation of these English and Nuu-Chah-Nulth fricatives to Italian consonants, using an L1-categorization and goodness rating task. An AXB discrimination task was used to assess their discrimination of the three English and three Nuu-Chah-Nulth minimal-place distinctions. In order to avoid having the L1 categorizations contaminate discrimination performance, for each stimulus language the AXB task was completed first, followed by the categorization and rating task.

## 2. Method

**2.1 Participants.** The listeners were 24 native speakers of Italian who also spoke Veneto Dialect (Venetan); all studied/worked at the University of Padova ( $M_{\text{age}}=27.96$  years; range=19-43; 13 female). Only one participant had ever lived outside of Veneto (one year each in Florence and Stockholm, during his late 30's). All had either acquired both languages from birth ( $n=17$ ) or had acquired Italian first and Venetan as an early second language ( $n=7$ ). They gave high self-ratings for comprehending ( $M=4.83$  on a 5-point scale) and speaking ( $M=4.5$ ) Venetan. Twenty-two learned Central Venetan, which like Italian has no interdental consonants. The other two had learned Northeast Venetan (Treviso), which has an interdental fricative /θ/ (Zamboni, 1974, 1988; see also Avesani, Galatà, Vayra, Best, Di Biase, Tordini, & Tisato, 2016; Avesani, Galatà, Best, Vayra, & Ardolino, 2017). For one of these two participants, Italian was the native language, Venetan was later-learned and weaker; his North Venetan experience did not enhance his detection of the dental feature of English /θ/, which he categorised 80%

of the time as /f/. For the other, Italian and Venetan were learned from birth and were equally strong; nonetheless he categorized English /θ/ similarly to the majority, as a mediocre Italian /t/ (see 3. Results).

Only one participant had not learned any additional languages at school. Twenty-two had studied English ( $M_{\text{onset-age}}=7.5$  years, range=4-11;  $M_{\text{duration}}=11.5$  years, range=7-20). Although this suggests they should be familiar with English /θ/ and /h/, we note that their mean self-ratings for speaking ( $M=3.0$  out of 5) and comprehending English ( $M=3.3$ ) were only fair. Other foreign languages were learned by fewer people and received even lower self-ratings: Spanish ( $n=6$ ,  $M_{\text{speaking}}=2.2$ ;  $M_{\text{comprehending}}=2.2$ ), French ( $n=12$ ,  $M_{\text{speaking}}=1.9$ ;  $M_{\text{comprehending}}=2.1$ ) and German ( $n=5$ ,  $M_{\text{speaking}}=2.6$ ;  $M_{\text{comprehending}}=2.4$ ). These languages do have some posterior fricatives, though none has the full array found in Nuu-Chah-Nulth. Spanish has only /x/, with some uvular variants [χ] in northern and central Spain (Hammond, 2001). Standard French has only [χ] as a positional devoiced allophone of its voiced uvular fricative /r/ ([ʀ]) following voiceless stops (e.g., *lettre*). German has three voiceless guttural fricatives: /x/, /χ/ and /h/ (no pharyngeal /ħ/), with /x/ displaying two vowel-context conditioned allophones, palatal [ç] and velar [ax̣].

## 2.2 Stimuli

**2.2.1 English.** Multiple tokens of the English anterior voiceless fricatives labiodental /f/, interdental /θ/, alveolar /s/ and palato-alveolar /ʃ/ ( $n=12$  each) in /Ca/ syllables were recorded in random order at Western Sydney University, Australia, by an Australian female speaker in her late 50's whose voice quality was similar to that of the Nuu-Chah-Nulth speaker (see 2.2.2). To ensure that discrimination of the English and Nuu-Chah-Nulth fricatives contrasts would not be confounded by non-criterial acoustic differences between the stimuli of the two languages, the English tokens were adjusted in Praat (Boersma & Weenink, 2009), using overlap-add resynthesis, to achieve a similar mean and range of consonant and vowel durations as the Nuu-Chah-Nulth stimuli. To reduce the possibility that /θ/-/s/ would be discriminated solely on intensity differences, /f/ and /θ/ were additionally amplified by 5 dB and /s/ reduced by 6 dB. This still left /s/ with a higher amplitude than /f/ and /θ/ to maintain naturalness. Vowel intensities for all tokens were adjusted to the same level of acoustic intensity.

**2.2.2 Nuu-Chah-Nulth.** Multiple natural tokens of the Nuu-Chah-Nulth posterior voiceless fricatives velar /x/, uvular /χ/, pharyngeal /ħ/ and glottal /h/ ( $n=15+$  each) in /Ca:/ syllable context (e.g., /xa:/) were produced in random order by a female native speaker in her 60's from the traditional tribal area on Vancouver Island, British Columbia, Canada. The recordings were made at the University of British Columbia. We chose a speaker of the elder generation because they maintain the /x/-/χ/ distinction, which younger speakers may have lost, i.e., /x/-/χ/ appears to have undergone merger over recent decades. For the perceptual tests we selected four tokens of each of the four target syllables, matched across fricative categories in duration, amplitude and pitch contour. Vowel intensities of all tokens were adjusted to the same level of acoustic intensity. (Note: the /x, χ, ħ/ and /f, θ, s/ stimuli were used in studies with infants in Tyler, Best, Goldstein, & Antoniou, 2014.)

**2.3 Procedure.** Participants completed a discrimination test followed by a categorization and rating test on the Nuu-Chah-Nulth consonants with respect to an on-screen array of Italian consonant choices, then discrimination followed by categorization and ratings of the English consonants. Discrimination was assessed prior to categorization in order to minimize confounding effects of prior categorizations on discrimination.

**2.3.1 Discrimination.** A categorial AXB discrimination task was used because it has lower memory demands and minimizes response biases relative to other standard discrimination protocols (see Best & Strange, 1992; Pollack & Pisoni, 1971; Strange & Shafer, 2008). On each trial participants received three stimuli separated by 1 s interstimulus intervals (ISIs), of which the first and third (A and B) were contrasting consonants and the middle item, X, was a different token of either the A or B consonant category. They had to indicate as quickly and accurately as possible whether X matched category A or B. Each stimulus triad appeared in four trial configurations: AAB, ABB, BBA, BAA. Inter-trial intervals (ITIs) were 3.5 s. Three contrasts were tested for each language, with 48 trials per contrast (4 trial types x 4 stimulus token triads x 3 repetitions) in separate blocks: Nuu-Chah-Nulth /x/-/χ/, /χ/-/ħ/ and /ħ/-/h/ and English /f/-/θ/, /θ/-/s/, /s/-/ʃ/. Test order of the contrast blocks within each language were randomized across participants. Before the first discrimination block they

received a short set of practice AXB trials that used an unrelated non-native lateral fricative voicing distinction from isiZulu (from Best, McRoberts, & Goodell, 2001).

**2.3.2 Categorization and goodness ratings.** On each trial of the categorization task following discrimination in each language, participants were presented with a single token and had to indicate which Italian consonant the non-native token sounded most similar to, selecting from a set of printed on-screen consonant+/a/ syllables using standard Italian spelling, which transparently conveys to Italians how to pronounce the consonants: FA, SA, SCIA, PA, TA, CA, LA, GLIA, RA, UA, CIA, JA, ZA and HA. We also provided examples of Italian words beginning with the relevant consonant. The fine-grained pronunciations of the initial consonants in the Venetan variety of Italian spoken by our participants are given in narrow IPA transcription as follows: FA [f], SA [s], SCIA [ʃ], are pronounced as in English. The voiceless stops differ from English, however, as they are unaspirated, which in initial position is phonetically more similar to English *voiced* stops: PA [p], TA [t], CA [k]. The glides of Italian also differ in phonetic details from English: LA [l] is lighter than in English (flatter tongue, less velarised), RA [r] is an alveolar tap/trill, GLIA [ʎa] is a palatal lateral, and UA [ua] differs dynamically from English [w]. The Italian affricate CIA is pronounced like English <ch> [tʃ]. The spelling JA was taken from the English loanword <jazz> because Italians pronounce it as [dʒa], whereas the Italian spelling GIA would be pronounced as a bisyllable [dʒia]. ZA [dza] is a dental affricate that does not exist in English. HA is pronounced [\_\_a] in Italian, i.e., with a “silent h” [∅]<sup>2</sup> rather than an aspirated [h] preceding the vowel, which would likely have a glottal stop onset. We asked them to choose the item with the most similar pronunciation of the consonant in Italian. Given that 22 of our 24 participants had studied English, we cannot rule out the possibility that some may have used <H> to indicate the English aspirated glottal fricative [h] despite our instructions to focus on Italian pronunciations. However, as a reminder, they self-rated their proficiency in speaking and understanding spoken English to be mediocre on average.

<sup>2</sup> <H> in Italian spelling is an orthographic convention. If it is inserted between <C, G> and <I, E>, it specifies that <C, G> are pronounced as the stops [k] and [g], rather than as the palatalized affricates [tʃ] and [dʒ] that are indicated by <CI, CE> and <GI, GE>, i.e., with no <H> intervening. Initial <H> also occurs but is silent in the written first, second and third person singular and third person plural forms of the verb <AVERE> ‘to have’ (HO, HAI, HA and HANNO, respectively).

After making their choice, they heard the same token again and had to rate how good a match it was to their selected Italian consonant, using a 1-7 Likert scale (1 = poor match, 7 = excellent match). There were 64 categorization trials per language (4 target consonants x 4 tokens each x 4 repetitions of the set), presented in random order. The first categorization test was preceded by a short practice set of the Zulu voiced and voiceless lateral fricatives.

### 3. Results

**3.1 Categorization and goodness ratings.** Although the categorization test was run *after* the discrimination test for each language, the categorization results will be presented first, as they determine the assimilations of the non-native fricatives to Italian consonants, which in turn provides the PAM predictions for discrimination performance differences among contrasts.

Italian labels [IPA]	NON-NATIVE TARGET FRICATIVES							
	English				Nuu-Chah-Nulth			
	/f/	/θ/	/s/	/ʃ/	/x/	/χ/	/h/	/h/
<C> [k]					<i>24</i> (3.37)			
<F> [f]	<b>96</b> (6.32)	25 (4.62)						
<H> [∅]					<i>57</i> (2.59)	<b>84</b> (3.98)	<b>95</b> (5.38)	<b>93</b> (5.20)
<J> [dʒ]					<i>15</i> (3.69)			
<S> [s]			<b>96</b> (6.20)					
<SCIA> [ʃ]				<b>96</b> (6.27)				
<T> [t]		<b>66</b> (3.08)						

Table 1. Mean percent categorizations and goodness ratings (1-7 scale; in parentheses) of each English and Nuu-Chah-Nulth target fricative to the Italian consonant choices (in Italian orthography and IPA). Boldface indicates significantly above chance. Italicized indicates significantly above chance but chosen significantly less often than the modal choice. Only labels selected significantly above chance (7%) are displayed. Italian labels chosen < 7% of the time for any target: <CIA>, <LA>, <PA>, <RA>, <ZA>, <GLIA>, <UA>).



Table 1 shows the categorization and goodness ratings for each English and Nuu-Chah-Nulth target fricative in relation to the Italian consonant choices. We used statistical criteria, rather than a pre-set threshold as in previous research, to determine whether a target consonant was Categorized to a single native consonant or was instead Uncategorized. The thresholds used in prior studies have not been standardized (varying among 50%, 70%, 90%), and their rationales have been somewhat subjective and arbitrary, which has made cross-study comparisons problematic. To address this, we created a new statistical criterion that can be applied systematically across different types of targets (consonants, vowels, tones) and across studies. Specifically, we designate a non-native target as Categorized if one L1 consonant was chosen significantly more than all other choices, and if it was also chosen significantly above chance.<sup>3</sup> If it did not meet both criteria it was deemed Uncategorized.

By these criteria, all English fricatives were Categorized. Although <T> was chosen for English /θ/ only 66% of the time on average, this was significantly above chance and significantly greater than choices of the next highest Italian category, <F> ( $M=25\%$ ),  $t_{(23)}=2.76$ ,  $p<.015$ . Each English fricative was Categorized to a different Italian label, making all pairwise assimilations Two Category (TC) contrasts. However, the TC assimilation for /f/-/θ/ differed in fine-grained detail from that for /f/-/s/ and /s/-/ʃ/. There was partial overlap in the use of <F> for both English /f/ and /θ/, yielding an overlapping TC assimilation pattern, or TC-O (see Tyler, Best, Faber, & Levitt, 2014), whereas there was no overlap in choices for the other two English contrasts, which were therefore non-overlapping TC assimilations (TC-N). In light of our previous arguments that overlapping assimilations should be more difficult to discriminate than non-overlapping ones within a contrast assimilation type (Faris, Best, & Tyler, 2016; Fenwick, Best, David, & Tyler, 2017; Tyler, Best, Faber, & Levitt, 2014), /f/-/θ/ should show poorer discrimination than /θ/-/s/ and /s/-/ʃ/, which should not differ from each other.

<sup>3</sup> These criteria differ from those proposed in Faris, Best, & Tyler (2016), which apply to a non-native item (vowels in their study) that had first been designated as Uncategorized according to a 50% threshold, i.e., the top choice native category was chosen less than half the time. If a single <50% category was nonetheless significantly above chance and no other categories were significantly above chance it was considered Focalised-Uncategorized. By our new statistical criteria, none of the English or Nuu-Chah-Nulth fricatives were Uncategorized.

In contradistinction to the English fricatives, a single label, <H>, was the most common choice for all four Nuu-Chah-Nulth fricatives; again they all met the Categorized criteria. For /x/, <H> was chosen ( $M = 59\%$ ) significantly more often than the two next-higher, above-chance category choices of <C> ( $M=23\%$ ),  $t_{(23)}=2.764, p<.012$ , and <J> ( $M=15\%$ ),  $t_{(23)}=3.69, p<.002$ . The pairwise assimilation patterns for the Nuu-Chah-Nulth contrasts were constrained both by the categorization of all four fricatives to <H> and by the more dispersed choices for /x/. Contrast /χ/-/ħ/ was assimilated to Italian <H> as a Category Goodness difference (CG) contrast, given that the ratings of goodness of fit to <H> were significantly lower for /χ/ ( $M_{\text{rating}}=4.0$ ) than /ħ/ ( $M_{\text{rating}}=5.4$ ),  $t_{(23)}=4.58, p<.001$ . The /x/-/χ/ contrast was also a CG assimilation, in which /x/ was rated a significantly poorer <H> ( $M_{\text{rating}}=2.6$ ) than /χ/ was ( $M_{\text{rating}}=4.0$ ),  $t_{(23)}=3.35, p<.002$ . However, /ħ/ and /h/ ratings as <H> did not differ significantly, making the assimilation of /ħ/-/h/ a Single Category (SC) contrast. Based on PAM predictions (Best, 1995), then, /ħ/-/h/ should show poorer discrimination than /χ/-/ħ/ and /x/-/χ/, which should not differ from each other, yet should show lower discrimination than the English TC contrasts.

**3.2 Discrimination.** Discrimination was above chance (50% on AXB tasks) for each of the six contrasts. Five contrasts were significantly above chance at  $p<.001$ : /f/-/θ/,  $t_{(23)}=9.418$ ; /θ/-/s/,  $t_{(23)}=11.988$ ; /s/-/ʃ/,  $t_{(23)}=16.522$ ; /x/-/χ/,  $t_{(23)}=8.770$ ; /χ/-/ħ/,  $t_{(23)}=6.367$ . Nuu-Chah-Nulth /ħ/-/h/ performance was also above chance,  $t_{(23)}=2.326, p<.03$ .

We conducted a repeated measures Analysis of Variance (ANOVA) on the accuracy data for four within-subjects factors: Language (English, Nuu-Chah-Nulth), Contrast Type (between-organ, within-organ, mixed/overlapping organs), Consonant (whether X in the AXB trials was the more anterior or more posterior consonant of the contrast) and Match (whether X matched the consonant of the first or third item in the AXB trial). For English, the between-organ contrast was /f/-/θ/ (lips vs. tongue tip constriction), the within-organ contrast was /θ/-/s/ (both tongue tip constrictions) and the mixed/overlapping organ contrast was /s/-/ʃ/ (tongue tip vs. tongue tip+body constrictions). For Nuu-Chah-Nulth the between-organ contrast was /ħ/-/h/ (tongue body+root constriction vs. glottis wide for aspiration), the within-organ contrast /x/-/χ/ (both tongue body

constrictions) and the mixed/overlapping organ contrast /χ/-/h/ (tongue body vs. tongue body+root constriction). The results are displayed in Figure 1.

The main effect of Language was significant,  $F_{(1,23)}=75.31, p<.001, \eta_p^2=.766$ , indicating that mean discrimination accuracy was significantly higher overall for English ( $M=84.9\%$ ) than for Nuu-Chah-Nulth ( $M=63.4\%$ ). Contrast Type was also significant,  $F_{(2,46)}=31.84, p<.001, \eta_p^2=.581$ . Pairwise tests on this effect indicated that, counter to AOH predictions, performance was significantly *lower* rather than higher for the between-organ contrasts ( $M=66.1\%$ ) relative to both the within-organ ( $M=78.6\%$ ),  $p<.001$ , and mixed/overlapping contrasts ( $M=77.7\%$ ),  $p<.001$ , which did not differ significantly. However, these main effects were modulated by three significant interactions: Language x Contrast Type,  $F_{(2,46)}=4.42, p=.018, \eta_p^2=.161$ ; Contrast Type x Consonant,  $F_{(2,46)}=5.34, p=.008, \eta_p^2=.188$ ; and Language x Contrast Type x Consonant,  $F_{(2,46)}=4.38, p=.018, \eta_p^2=.161$ .

To break down the three-way interaction, we ran separate ANOVAs for each language on the within-subjects factors Contrast Type x Consonant x Match. For the English ANOVA, only the main effect of Contrast Type was significant,  $F_{(2,46)}=12.051, p<.001, \eta_p^2=.344$ . Pairwise comparisons indicated that performance on the between-organ contrast, /f/-/θ/, was significantly lower ( $M=78.6\%$ ) than on the within-organ /θ/-/s/ (86.5%),  $p<.007$ , and the mixed/overlapping /s/-/ʃ/ contrast ( $M=89.6\%$ ),  $p<.001$ , but the latter two contrasts did not differ from each other.

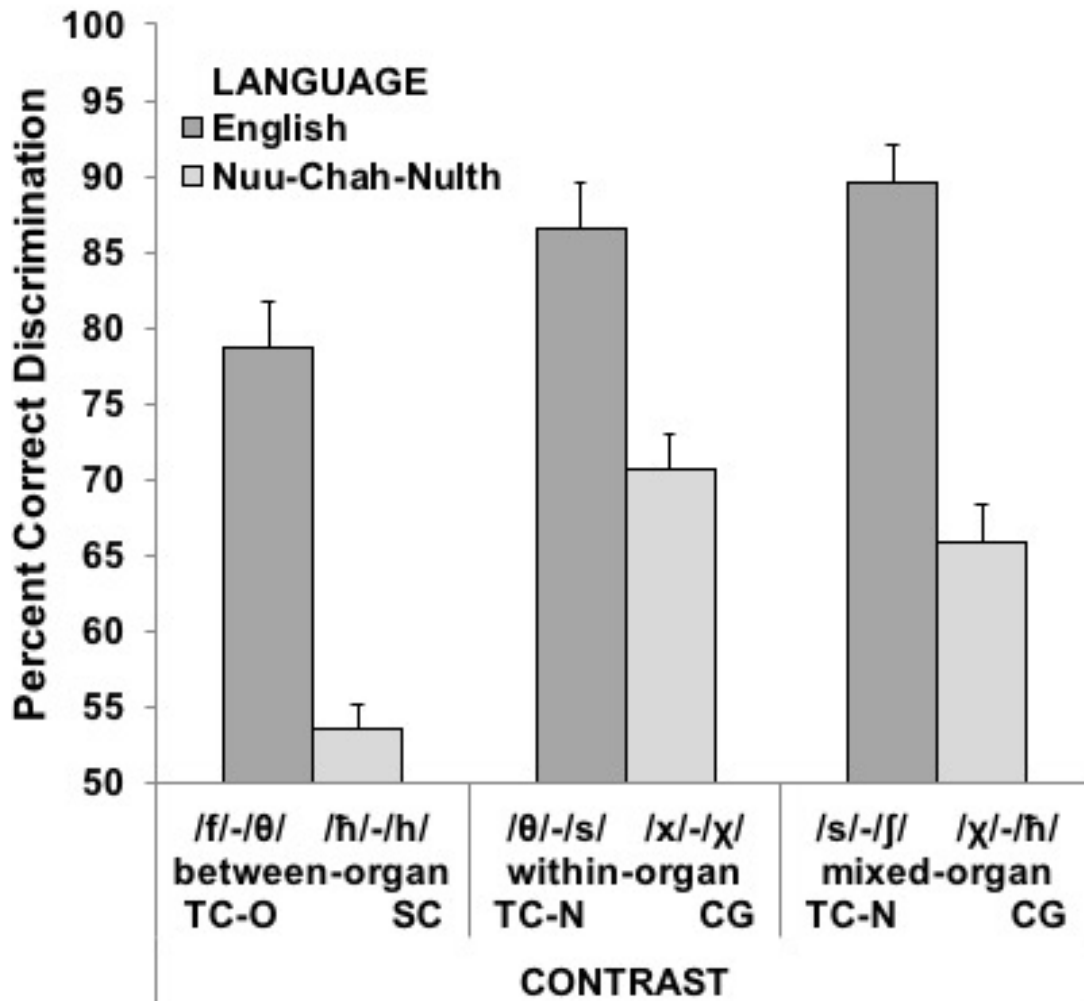


Figure 1. Mean percent correct discrimination of the English and Nuu-Chah-Nulth voiceless fricative contrasts, with Contrast Type and assimilation pattern (from the Categorization/rating results) displayed beneath each discrimination pair on the x-axis.

The significantly lower performance on /f/-/θ/, and the split in categorizations of /θ/ as <T> versus as <F> (see Table 1), are consistent with our predictions regarding TC-O (overlapping) assimilation for this contrast. However, we noted individual variability in the tendency to report <F> for /θ/, which led us to examine individual participants' categorizations of /θ/. In total, 17 of the 24 participants (71%) selected <T> more than 50% of the time. Of the other seven participants (29%), six selected <F> more than 50% of the time. The remaining participant did not select any one label more than 50% of the time, but her highest response was <F> (31%) and she never selected <T>, so we grouped her with the six who had

categorized /θ/ to <F>. Note that the /f/-/θ/ contrast is a non-overlapping Two Category (TC-N) assimilation for listeners who categorize /θ/ to Italian <T>, but a Category Goodness difference (CG) assimilation for those who categorized /θ/ as Italian <F>; they gave a very good rating of English /f/ as <F> ( $M_{\text{rating}}=6.04$ ) as compared to a moderate rating of /θ/ as <F> ( $M_{\text{rating}}=4.91$ ),  $t_{(6)}=.29$ ,  $p<0.032$  (one-tailed, as better ratings as <F> are predicted for /f/ than /θ/ stimuli). Given these subgroup differences, we conducted a new ANOVA on the English discrimination data, with Subgroup (/θ/-as-<T> vs. /θ/-as-<F>) as a between-subjects factor, and Contrast Type (between-organ, within-organ, mixed/overlapping organs) as a within-subjects factor. Only the interaction was significant,  $F(2, 44)=5.60$ ,  $p=.007$ ,  $\eta_p^2=.20$ . Post-hoc  $t$ -tests revealed a significant difference between the Subgroups on discrimination of /f/-/θ/,  $t(22)=2.81$ ,  $p=.01$ , with the TC-N /θ/-as-<T> categorizers showing better discrimination ( $M=83\%$ ) than the CG /θ/-as-<F> categorizers ( $M=67\%$ ). There was no Subgroup difference for the other two contrasts (see Figure 2).

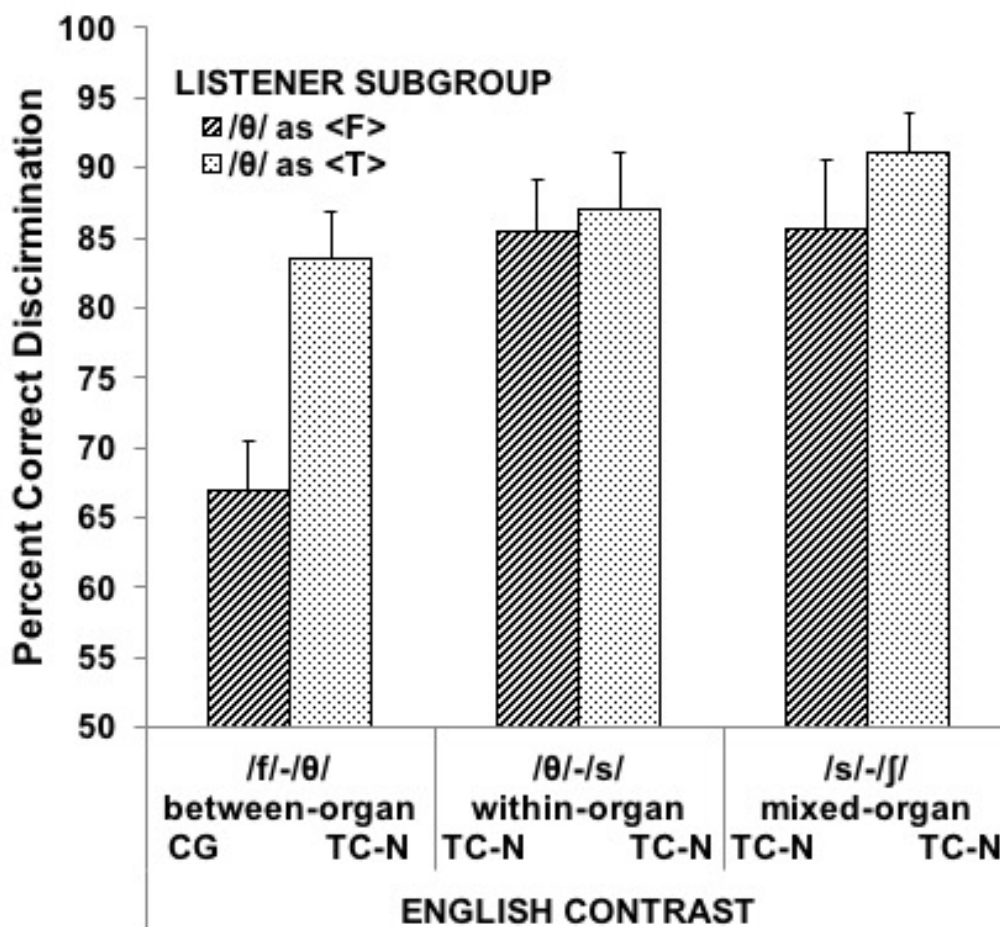


Figure 2. Mean percent correct discrimination of the English voiceless fricative contrasts by the participants who assimilated English /θ/ as <T> versus the participants who assimilated it as <F>.

The Nuu-Chah-Nulth breakdown ANOVA also found a significant main effect of Contrast Type,  $F_{(2,46)}=24.5$ ,  $p<.0001$ ,  $\eta^2=.516$ , for which pairwise tests indicate that discrimination of the between-organ SC contrast, /h/-/h/, was significantly lower ( $M=53.6\%$ ) than the within-organ /x/-/χ/ ( $M=70.7\%$ ),  $p<.001$ , and mixed/overlapping /χ/-/h/ CG contrasts ( $M=65.9\%$ ),  $p<.001$ , which did not differ significantly. The significant Contrast Type x Consonant interaction,  $F_{(2,46)}=6.68$ ,  $p<.003$ ,  $\eta^2=.225$ , revealed that discrimination of the CG within-organ contrast /x/-/χ/ was better when X in the AXB trials was /x/ ( $M=76\%$ ) than when it was /χ/ ( $M=65.3\%$ ), but there was no Consonant effect for the SC between-organ contrast /h/-/h/ ( $M=50.1$  vs.  $56.4\%$ ) or the CG mixed-organ contrast /χ/-/h/ ( $M=67.2\%$  vs.  $64.6\%$ ).

To probe that interaction, we looked for individual differences in categorization of Nuu-Chah-Nulth /x/, as we had for English /θ/. In this case, 23 participants formed three subgroups of responders; the 24<sup>th</sup> split her responses 50/50 between <C> and <H>. The largest subgroup Categorized /x/ above 50% as <H> ( $n=12$ ;  $M_{<H>}=92.75\%$ ), followed by those who Categorized it as Italian <C> above 50% (or in one case as the most frequent choice at 44%) ( $n=7$ ;  $M_{<C>}=67.14\%$ ), and the smallest number Categorized it as <J> ( $n=4$ ;  $M_{<J>}=70.5\%$ ). Thus, the /x/-as-<H> subgroup assimilated Nuu-Chah-Nulth /x/-/χ/ as a CG difference within <H>, but the remaining two subgroups assimilated /x/-/χ/ as a TC-O (overlapping) contrast (<C> or <J> vs. <H>). Therefore, we combined the <C> and <J> categorizers into a single TC-O subgroup ( $n=11$ ) and conducted an ANOVA on the between-subjects factor Subgroup (<H> vs <C/J> categorizers, i.e., CG vs. TC-O, respectively) and within-subject factors Contrast Type x Consonant that had interacted in the Nuu-Chah-Nulth breakdown analysis. Neither the Subgroup main effect nor any interactions with it were significant. Thus, unlike the case with English /θ/, the Nuu-Chah-Nulth /x/ categorization subgroups did not differ in discrimination performance, not even on the /x/-/χ/ contrast despite their CG vs. TC-O assimilation differences. However, we should note that discrimination of a CG assimilation may not necessarily be expected to be much better than discrimination of a TC-O assimilation type (see Faris, Best, & Tyler, 2016; Fenwick, Best, David, & Tyler, 2017; Tyler, Best, Faber, & Levitt, 2014). Consistent with the Nuu-Chah-Nulth breakdown ANOVA, the significant effects of the current analysis were Contrast Type,  $F_{(2,42)}=20.884$ ,  $p<.0001$ ,  $\eta_p^2=.499$ , Consonant (of X in AXB),  $F_{(2,42)}=4.511$ ,  $p<.047$ ,  $\eta_p^2=.177$ , and their interaction,  $F_{(2,42)}=6.272$ ,  $p<.005$ ,  $\eta_p^2=.230$ .

#### 4. Discussion

The listeners showed Categorized assimilations of each of the English and Nuu-Chah-Nulth consonants to native Italian consonants. Whereas each English fricative was categorized as a different Italian consonant (/f/-as-<F>; /θ/-as-<T>; /s/-as-<S>; /ʃ/-as-<SCIA>), the Nuu-Chah-Nulth fricatives were all Categorized as Italian <H>. Given that Italian (and Venetan Dialect) does not have the phoneme /h/, and that <H> in written Italian words is “silent” [Ø], this may mean that the listeners heard *no* consonant at the syllable onsets of the Nuu-Chah-Nulth target stimuli. On the other hand, even though they had been instructed to indicate which *Italian* consonant they perceived, as noted earlier we cannot rule out that they may have chosen <H> to indicate they heard an English [h] for the Nuu-Chah-Nulth consonants given that all but two participants had learned English at school<sup>4</sup>. Those two still chose <H> 94-100% of the time for all Nuu-Chah-Nulth consonants, however, like the L2-English majority. In any case, whether their choices of <H> indicate Italian silent [Ø] or English [h], the listeners heard all Nuu-Chah-Nulth fricatives as most similar to the same single category <H>.

The pairwise assimilation patterns were Two Category (TC) for all three English contrasts, two of them non-overlapping (TC-N: /θ/-/s/ and /s/-/ʃ/) and the other overlapping (TC-O: /f/-/θ/), whereas the Nuu-Chah-Nulth contrasts instead showed either Single Category (SC) assimilation (pharyngeal vs. glottal /ħ/-/h/) or Category Goodness difference (CG) assimilation (uvular vs. pharyngeal /χ/-/ħ/, and velar vs. uvular /x/-/χ/). PAM predictions were that the English contrasts should be discriminated significantly better than the Nuu-Chah-Nulth contrasts, which was supported by a main effect of Language. Moreover, both Nuu-Chah-Nulth CG contrasts were predicted by PAM to be discriminated significantly better than the SC contrast, but not to differ from each other, also upheld by the analyses. Thus, overall the PAM predictions were supported quite well.

The core AOH predictions about discrimination levels for within-versus between-organ contrasts, on the other hand, were not supported. Indeed, the observed patterns actually run counter to AOH predictions of better discrimination for between than within-organ non-native contrasts. Performance was better for contrasts involving natively-used and L2-learned articulatory organs, which follows more from PAM principles than AOH predictions. Whereas English-learning infants fail to discriminate

---

<sup>4</sup> One had only learned French, which also lacks [h]; the other had learned no foreign languages.

English anterior fricatives better than posterior Nuu-Chah-Nulth fricatives (Tyler et al., 2014), our Italian adult participants did discriminate the more familiar L2-English fricatives better than the completely unfamiliar non-native Nuu-Chah-Nulth ones.

We must consider, as well, how the results might relate to other models of non-native speech perception. The SLM prediction that *new* phones from a non-native language should be perceived more accurately than *similar* non-native phones, due to the latter being more readily equivalence classified to native phonemes, is contradicted by our finding of significantly *poorer* discrimination for the Nuu-Chah-Nulth fricatives than the English fricatives. However, the present study does not address SLM's predictions that the Nuu-Chah-Nulth fricatives should be more readily established as new L2 phonemes, including more accurate L2 production, as compared to learning and production of English /θ/. Further research would be needed to evaluate those SLM predictions.

Meanwhile, the two English /θ/-assimilation subgroups and their differences in discrimination of English /θ/-/f/ challenge claims that cross-language speech perception is driven primarily by acoustic similarity/distance (e.g., L2LP: Escudero & Boersma, 2005; Holt & Lotto, 2008), given that the listeners who categorized /θ/ to the acoustically more similar /f/ were in the minority rather than the majority, and despite rating the goodness of /θ/ significantly lower than that of /f/ this subgroup showed substantially poorer discrimination of /θ/-/f/ than the majority of listeners who categorized /θ/ to the acoustically more dissimilar Italian /t/. Assimilating /θ/ to Italian /t/ has potential L2 phonological benefits over a more acoustically-based categorization to /f/: it could help L2 Italian learners of English to maintain both the English /θ/-/f/ distinction as a TC assimilation to Italian /t/-/f/, and the English /θ/-/t/ distinction as a CG assimilation between a poor /t/ (2.8 rating for /θ/ as Italian short-lag /t/) versus a moderately good /t/ (English long-lag /t/ as Italian /t/).

Further studies comparing perception of these English versus Nuu-Chah-nulth contrasts by listeners of varying L1s that differ in contrastive use of the tongue body, tongue root, epiglottis and glottis for voiceless fricatives could further delineate the contributions of perceptual assimilation (PAM) and equivalence classification (SLM) on categorization and discrimination of non-native and L2 consonant contrasts. Studies on the impact of L2 learning or laboratory perceptual training on perception and production of the posterior fricative series by different L1 groups would also be informative.



## 5. References

- Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. *6<sup>th</sup> International Congress of Phonetic Sciences, Prague*, 569-573.
- Abramson, A. S., & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, 1, 1-8.
- Antoniou, M., Best, C. T., & Tyler, M. D. (2013). Greek-English bilinguals' and Greek and English monolinguals' perception of nonnative Ma'di voicing contrasts. *Journal of the Acoustical Society of America*, 133, 2397-2411.
- Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do bilinguals perceive stop voicing differently according to language mode? *Journal of Phonetics*, 40, 582-594.
- Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, 38, 640-653.
- Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics*, 39, 558-570.
- Avesani, C., Galatà, V., Vayra, M., Best, C. T., Di Biase, B., Tordini, O., & Tisato, G. (2016). In M. Vayra, C. Avesani, & F. Tamburini (Eds.), *Italian roots in Australian soil: Coronal obstruents in native dialect speech of Italian-Australians from two areas of Veneto [Language acquisition and language loss: Acquisition, change and disorders of the language sound structure]* (pp. 74-98). Milan: Studi AISV. ISBN: 978-88-97657-11-8.
- Avesani, C., Galatà, V., Best, C. T., Vayra, M., Di Biase, B., & Ardolino, F. (2017). Phonetic details of coronal consonants in the Italian spoken by Italian-Australians from two areas of Veneto. In C. Bertini, C. Celata, G. Lenoci, C. Meluzzi, & I. Ricci (Eds.), *Social and biological factors in speech variation* (pp. 281-306). Milan, Studi AISV. ISBN 978-88-97657-19-4.
- Bertinetto, P. M., & Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetics Association*, 35(2), 131-151.
- Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 167-200). Timonium, MD, York Press.
- Best, C. T. (2019). The diversity of tone languages and the roles of pitch variation in non-tone languages: Considerations for tone perception research. *Frontiers in Psychology*. (online publication).
- Best, C. T., Goldstein, L., Tyler, M. D., & Nam, H. (2016). Articulating what infants attune to in native speech. *Ecological Psychology*, 28, 216-261.

- Best, C. T., & McRoberts, G. W. (2003). Infant perception of nonnative consonant contrasts that adults assimilate in different ways. *Language & Speech*, *46*, 183-216.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology. *Journal of the Acoustical Society of America*, *109*, 775-794.
- Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 45-60.
- Best, C. T., Shaw, J., & Clancy, E. (2013). Recognizing words across regional accents: The role of perceptual assimilation in lexical competition. *Interspeech*, *2013*, 2128-2132.
- Best, C. T., Shaw, J., Mulak, K., Docherty, G., Evans, B., Foulkes, P., Hay, J., Al-Tamimi, J., Mair, K., & Wood, S. (2015a). Perceiving and adapting to regional accent differences among vowel subsystems. *18<sup>th</sup> International Congress of Phonetic Sciences, Glasgow*, 0964. <http://www.icphs2015.info/pdfs/Papers/ICPHS0964.pdf>
- Best, C. T., Shaw, J., Mulak, K., Docherty, G., Evans, B., Foulkes, P., Hay, J., Al-Tamimi, J., Mair, K., & Wood, S. (2015b). From Newcastle MOUTH to Aussie ears: Australians' perceptual assimilation and adaptation for Newcastle UK vowels. *Interspeech*, *2015*, 1932-1936.
- Best, C. T., & Strange, W. (1992). Effects of language-specific phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*, 305-330.
- Best, C. T., Traill, A., Carter, A., Harrison, K. D., & Faber, A. (2003). !Xóǀ click perception by English, Isizulu, and Sesotho listeners. *15<sup>th</sup> International Congress of Phonetic Sciences, Barcelona*, 853-856.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. Munro, & O.-S. Bohn (Eds.), *Second Language Speech Learning* (pp. 13-34). Amsterdam: John Benjamins Publishing.
- Boersma, P., & Weenink, D. (2009). *Praat: Doing phonetics by computer* [Computer program]. Retrieved 2009. <http://www.fon.hum.uva.nl/praat/>
- Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*, 303-328.
- Bohn, O.-S., & Flege, J. (1993). Perceptual switching in Spanish/English bilinguals: Evidence for universal factors in stop voicing judgments. *Journal of Phonetics*, *21*, 267-290.

- Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011a). Vocabulary size is associated with second language vowel perception performance in adult second language learners. *Studies in Second Language Acquisition*, 33, 433-461.
- Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C., & Tyler, M. D. (2011b). Vocabulary size matters: The assimilation of L2 Australian English vowels to L1 Japanese vowel categories. *Applied Psycholinguistics*, 32, 51-67.
- Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C., & Tyler, M. D. (2012). Second language learners' vocabulary expansion is associated with improved Second Language vowel intelligibility. *Applied Psycholinguistics*, 33, 643-664.
- Carlson, B. F., & Esling, J. H. (2003). Phonetics and physiology of the historical shift of uvulars to pharyngeals in Nuuchahnulth (Nootka). *Journal of the International Phonetic Association*, 33, 183-193.
- Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, 39, 492-504.
- Dupoux, E., & Peperkamp, S. (2002). Fossil markers of language development: Phonological 'deafnesses' in adult speech processing. In J. Durand, & B. Laks (Eds.), *Phonetics, phonology, and cognition* (pp. 168-190). Oxford, UK: Oxford University Press.
- Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26, 551-585.
- Faris, M. M., Best, C. T., & Tyler, M. D. (2018). Discrimination of uncategorized non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics*, 70, 1-19.
- Faris, M. M., Best, C. T., & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *JASA-Express Letters*, 139, EL1-5.
- Fenwick, S. E., Best, C. T., Davis, C., & Tyler, M. D. (2017). The influence of auditory-visual speech and clear speech on cross-language perceptual assimilation. *Speech Communication*, 92, 114-124. <http://dx.doi.org/10.1016/j.specom.2017.06.001>
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 233-277). Timonium, MD: York Press.
- Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer, & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319-355). Berlin: Mouton de Gruyter.

- Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole, & J. Hualde (Eds.), *Laboratory Phonology 9* (pp. 353-380). Berlin: Mouton de Gruyter.
- Goldstein, L. M., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller, & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159-207). Berlin: Mouton de Gruyter.
- Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, *107*, 2711-2724.
- Hallé, P. A., Chang Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Chinese versus French listeners. *Journal of Phonetics*, *3*, 395-421.
- Hammond, R. M. (2001). *The sounds of Spanish: Analysis and application*. Somerville, MA: Cascadilla.
- Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science*, *17*, 42-46.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47-B57.
- Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, *21*, 3-28.
- Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Krebs-Lazendic, L., & Best, C. T. (2013). First language suprasegmentally-conditioned syllable length distinctions influence perception and production of second language vowel contrasts. *Laboratory Phonology*, *4*, 435-474.
- Kuhl, P. K. (1993a). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259-274). Dordrecht, the Netherlands: Springer.
- Kuhl, P. K. (1993b). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics*, *21*, 125-139.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *363*, 979-1000.

- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*, F13-F21.
- Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*, 864-886.
- Leussen, V., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*, 1000.
- MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*, 369-390.
- Mioni, A. (2001). *Elementi di fonetica*. Padova: Unipress.
- Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*, 331-340.
- Polivanov, E. (1931/1974). La perception des sons d'une langue étrangère. *Travaux du Cercle linguistique de Prague*, *4*, 79-96. Translated by D. Armstrong (1974). *Polivanov E. Selected Works. Articles on general linguistics*. Paris: Mouton.
- Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, *89*, 2961-2977.
- Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, *52*, 37-52.
- Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*, 221-231.
- Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*, 467-478.
- Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/-/ð/ perception: evidence for a new developmental pattern. *The Journal of the Acoustical Society of America*, *109*, 2190-2201.
- Pollack, I., & Pisoni, D.B. (1971). On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science*, *24*, 299-300.
- Reid, A., Burnham, D., Attina, V., Kasisopa, B., Schwarz, I.-C., & Best, C. T. (2015). Perceptual assimilation of lexical tone: The role of language experience and visual information. *Attention, Perception & Psychophysics*, *77*, 571-591.
- Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*, 111-123.
- Shaw, J. A., Best, C. T., Docherty, G., Evans, B., Foulkes, P., & Hay, J. (2018). Resilience of English vowel perception across regional accent variation. *Laboratory Phonology*, *9*(1), 11, 1-36. <http://doi.org/10.5334/labphon.87>.

- Shaw, J., Best, C. T., Mulak, K., Docherty, G., Evans, B., Foulkes, P. Hay, J., Al-Tamimi, J., Mair, K., Peek, M., & Wood, S. (2014). Effects of short-term exposure to unfamiliar regional accents: Australians' categorization of London and Yorkshire English consonants. *Speech Science and Technology*, 2014, 71-74.
- So, C., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language & Speech*, 53, 273-293.
- So, C., & Best, C. T. (2011). Categorizing Mandarin tones into listeners' native prosodic categories: The role of phonetic properties, *Poznań Studies in Contemporary Linguistics*, 47, 133-145.
- So, C., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, 36, 195-221.
- Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 3-45). Timonium, MD: York Press.
- Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39, 456-466.
- Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. H. Edwards, & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153-192). Amsterdam: John Benjamins.
- Studdert-Kennedy, M., & Goldstein, L. (2003). Launching language: The gestural origin of discrete infinity. In M. H. Christensen, & S. Kirby (Eds.), *Language Evolution* (pp. 235-254). Oxford, UK: Oxford University Press.
- Trubetzkoy, N. S. (1939/1969). Grundzüge der phonologie, *Travaux du Cercle Linguistique de Prague*, 7, 1-271. Translated by C.A.M. Baltaxe (1969). *Principles of phonology*. Berkeley, CA: University of California Press.
- Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *Journal of the Acoustical Society of America*, 120, 2285-2294.
- Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, 71, 4-21.
- Tyler, M. D., Best, C. T., Goldstein, L. M., & Antoniou, M. (2014). Investigating the role of articulatory organs and perceptual assimilation in infants' discrimination of native and non-native fricative place contrasts. *Developmental Psychobiology*, 56, 210-227.
- Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language learning and development*, 1, 197-234.

- Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52, 349-355.
- Zamboni, A. (1974). *Veneto*. Pisa: Pacini Editore.
- Zamboni, A. (1988). Italienisch: Areallinguistik IV a) Venezien. In G. Holtus, M. Metzeltin, & C. Schmitt (Eds.), *Lexicon der Romanistischen Linguistik*, vol. IV (pp. 517-538). Tübingen, Germany: Niemeyer Verlag.

## **Paa Paa Plack Sheep: Discrimination of L2 Stop Voicing Contrasts in the Absence of L1 Stop Voicing Distinctions**

Rikke Louise Bundgaard-Nielsen  
Western Sydney University

Brett Joseph Baker  
University of Melbourne

### **Abstract**

More than 50 years of research has shown that native language experience shapes the perception not only of an individual's first/native language, but also languages subsequently acquired. This pervasive shaping effect of native language acquisition often results in 'accented' second language speech perception and production, when the languages differ in their phonemic inventory or the phonetic realisation of shared phonemes. Little, however, is known about the way in which nonnative and second language contrasts are acquired when they involve linguistic dimensions that are un-exploited and non-contrastive in an individual's native language (as opposed to a different organisation of a shared linguistic dimension). We examine this scenario in a study of VOT-based stop contrast discrimination by participants without native VOT-based stop experience, and participants whose native language exploits VOT-differences as a secondary cue. The results suggest that even extensive second language experience is insufficient for second language learners without native voicing experience to acquire such a distinction.

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 41-63). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.



## **1. Introduction**

All of the world's languages make use of speech sounds – phonemes – that are commonly referred to as ‘stop consonants’. Stop consonants, such as English /p t k/ and /b d g/ are made by the forming of and the releasing of a constriction somewhere in the oral cavity, at the lips for /p b/; the alveolar ridge for /t d/; and the velum for /k g/. In many languages, including English, Spanish and Mandarin, stop consonants form pairs which share their place of articulation – /t d/, /p b/, and /k g/ – but differ in the timing of vocal fold vibration relative to the release of the constriction. In each pair, the consonants /p t k/ are ‘voiceless’, as vocal fold vibration (voicing) generally does not occur until after the release of the oral constriction. The other three stops /b d g/ are ‘voiced’, as vocal fold vibration begins before release of the constriction, or at the time of release (see Lisker & Abramson, 1964; Abramson & Lisker, 1970; Maddieson, 1984; Henton, Ladefoged & Maddieson, 1992; Cho & Ladefoged, 1999; and see a recent review by Abramson & Whalen, 2017). In a smaller set of languages, such as Thai, speakers produce and perceive three distinct stops /p<sup>h</sup> p b/, differing primarily in VOT and aspiration (the ‘puff’ of air associated with release of the oral constriction), at each place of articulation (Tingsabadh & Abramson, 1993). For illustration of the distribution of VOT in two languages with two series of stops (English; Spanish) and one language with three series of stops (Thai), see Figure 1 below. A yet smaller number of languages even have four categories at each place of articulation, including Hindi (Gopal, 1993).

Many Australian Indigenous languages famously lack both voicing distinctions and fricatives altogether, while others employ consonant contrasts characterized by duration differences rather than VOT (see Fletcher & Butcher, 2014). In this chapter, we report a two-part study examining whether speakers of such a language can discriminate nonnative (English) stop and fricative consonants contrasts which differ just in voice or in voice as well as duration.

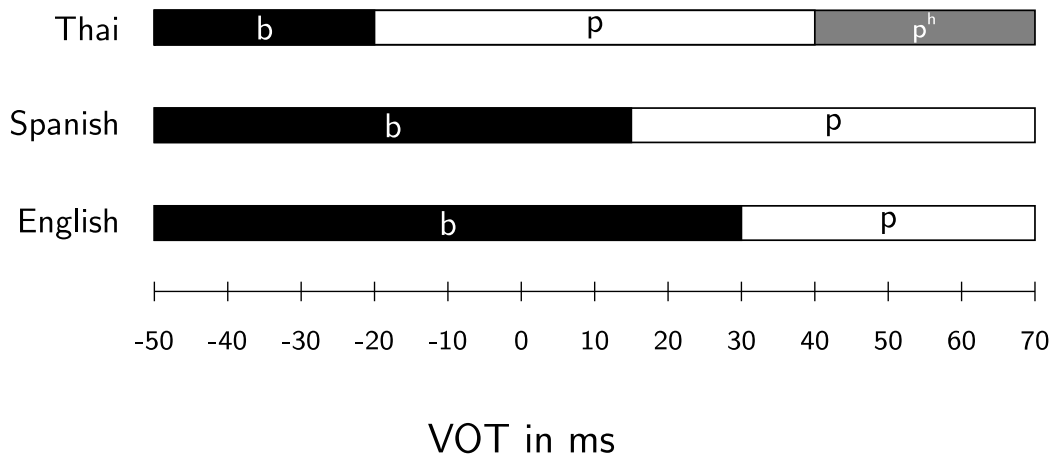


Figure 1. VOT boundaries across three languages. Adapted from Abramson & Lisker, 1970.

### 1.1 Background

More than 50 years of research has demonstrated that native/first language (L1) stop voicing perception is highly automatic and categorical, with language specific and relatively sharp perceptual boundaries marking the shift from one phoneme category to another (Abramson & Lisker, 1970). Indeed, a native listener will generally perceive differences in VOT between two native phones *only* when those two phones fall on either side of the category boundary. If the two phones fall within the same category, even relatively large differences in VOT are ignored. This is crucial for efficient first language processing, but also has important implications for second language (L2) and cross-language speech perception. Indeed, decades of segmental perception research focusing on voicing-based stop distinctions have resulted in three important observations in this regard.

Firstly, it is clear that nonnative listeners systematically use their L1 VOT contrast boundary in perceiving phones in an unfamiliar language or L2. This is the case even when the phonetic realisation of the contrast in the L2 differs from that of the listeners' L1, such as is the case of Spanish learners of English, who will perceive some English /b/s as Spanish /p/s (Abramson & Lisker, 1973; Flege, 1987), and English learners of Thai, who will perceive only a two-way stop distinction in Thai, despite the fact that Thai has a three-way distinction (Abramson & Tingsabadh, 1999). Such application of L1 categorical boundaries to L2 speech is a key contributor

to what we might refer to as an ‘accent on the ears’, as well as the perhaps more commonly noted ‘accent’ in nonnative speech production.

Secondly, we know that the number of native versus non-native/L2 VOT-based phonemic contrasts (two as in English; three as in Thai; four as in Hindi) is important to a non-native/L2 listener. Another important aspect is the magnitude of acoustic/articulatory difference between the native and nonnative/L2 phones, even when the L1 and L2 are matched in terms of their phonological inventories. Just as is the case for child L1 learners (see for instance Davis, 1995), it is easier for L2 users to perceive non-native contrasts with large acoustic differentiation. For example, adult English speakers are better at discriminating Thai stops /p/ vs. /p<sup>h</sup>/, which differ in aspiration *in addition* to VOT, than Thai stops /p/ vs /b/, which differ *only* in VOT (Beach, Burnham & Kitamura, 2001; Pater, 2003; Tsukada, 2004).

Thirdly, research has shown that this phenomenon of ‘accented perception’ often persists also for even highly proficient L2 language users. Indeed, this has been shown on the *phonological* level for English learners of Thai who struggle to discriminate the three Thai stop VOT categories (Abramson & Lisker, 1970; Strange, 1972; Pisoni, Aslin, Perey, & Hennessey, 1982) and on a *phonetic* level in, for instance, the difficulty experienced by Spanish-English bilinguals whose languages differ in the VOT setting for voiced and voiceless stops (short-lag/long-lag versus pre-voiced/short-lag stop realisation) (Abramson & Lisker, 1973; Flege, 1987).

The findings listed above are but a sliver of a rich field of research into L2 segmental perception which has been interpreted from a number of theoretical perspectives, including the Perceptual Assimilation Model (PAM: Best, 1994; Best 1995; PAM-L2: Best & Tyler, 2007), and the Speech Learning Model (Flege, 1995). According to PAM/PAM-L2, L1 phonological learning shapes the way in which L2 phones are perceived, and L1 phonological and phonetic knowledge subsequently imposes structure on the perception of non-native/L2 material. PAM predicts that L2 phones are discriminated on the basis of their mapping into L1 phoneme categories in a number of different patterns (Best, 1994; Best 1995; Best & Tyler, 2007), including: (1) Single Category (SC) contrasts in which two L2 phones are perceived as equally good instances of the same L1 phonemic category, and discrimination is expected to be poor; (2) Category Goodness (CG) contrasts in which two L2 phones are instances of the same L1 phonemic category, but one L2 phone is perceived as a ‘better’ fit (phonetically) than the other, and discrimination is predicted to be moderate to good and; (3) Two-Category (TC) contrasts where two L2

phones are assimilated into separate L1 phoneme categories. Discrimination is predicted to be excellent. According to SLM, which historically has had a greater focus on second language segmental production than perception, what matters for second language learning is also the relationship between the segments of the native and nonnative language(s), though the focus is on equivalence classification predominantly on the acoustic-phonetic level, rather than the abstract phonological level.

The predictions of PAM/PAM-L2 and SLM with respect to the processing of non-native/L2 segmental information (and even supra-segmental information such as tone: So & Best, 2010, 2011, 2014; Wu, Bundgaard-Nielsen, Baker, Best, & Fletcher, 2015; Wu, Fletcher, Baker, & Bundgaard-Nielsen, 2016; Wu, Fletcher, Bundgaard-Nielsen, & Baker, 2016) have been tested using a range of language combinations, differing in the phonetic realization of the same number of voicing-categories (such as English and Spanish) as well as languages differing in the number of phonemic distinctions (such as English and Thai). Significant differences in the theoretical underpinnings of the models, as well as in the role of abstract phonological knowledge in second language learning aside, many of these results are relatively consistent with the key assumptions of the two models that the specific native language experience of any second language learner is important to non-native and second language segmental perception. Neither theory, however, makes any explicit predictions for the scenario explored in the present chapter (but see Best, Avesani, Tyler & Vayra in the present volume for detailed discussion of PAM-L2, as well as implications for the Articulatory Organ Hypothesis). What happens when a learner must add a novel dimension to their linguistic repertoire in order to successfully acquire another language?

No work has yet examined nonnative VOT-based stop contrast discrimination by L1 speakers of languages without voicing-based contrasts altogether, such as the Indigenous Australian language Wubuy (see below). This means that we have very limited knowledge of what happens when speakers are introduced to a novel language which makes use of systematic differences on the linguistic dimension of voicing to which speakers have not had to attend in their L1. And while such languages are typologically rare, experimental examination of the way in which speakers *without* a voicing-based distinction acquire new languages which *do* make use of voicing-based distinctions might provide crucial insights into the question of how flexible the speech perception system is when it comes to a dimension of speech not exploited in the native language.

Indeed, in such a scenario, successful L2 perception is not achieved by the shifting perceptual boundaries through re-attunement of native phonetic knowledge (as in the Spanish-English pairing where the voiced-voiceless category boundary must shift) or via a re-phonologisation of the acoustic/articulatory space (as in the English-Thai pairing where two categories must become three, or vice versa). Rather, this case presents the task of attuning to systematic distributional differences in a perceptual dimension (presence/absence of vocal fold vibration) that has hitherto not afforded the listener any systematic information relevant to categorical perception in his or her L1.

The following presents a two-part study testing the perception of voicing contrasts in stops and fricatives, by participants who differ systematically in their native language experience with voicing distinctions. Study 1 tested the discrimination of stop consonants while Study 2 tested the discrimination of English fricatives as well as a fricative-stop contrast by Wubuy, Roper Kriol, and Australian English listeners.

Wubuy (also known as ‘Nunggubuyu’; Heath, 1984) is a highly endangered Indigenous Australian language spoken in south-eastern Arnhem Land, on the coast of the Gulf of Carpentaria around the southern part of Blue Mud Bay in the Northern Territory of Australia. It is the first language for adults over the age of around 55 in the community of Numbulwar, as well as a first or second language for some adults on the neighbouring island Groote Eylandt in the Gulf of Carpentaria. The children growing up in Numbulwar are no longer acquiring Wubuy as a first language, though most children are exposed to Wubuy through interactions with older family members, and through the language revitalisation efforts at Numbulwar school. There is some degree of receptive Wubuy skills in some younger adults as well. There are perhaps 60 fluent L1 Wubuy speakers in Numbulwar and neighbouring communities.

The phonology of Wubuy resembles the neighbouring Yolngu languages in having the rare four-way coronal place distinction among the stops /t̪, t, ɬ, ɬ̪/, in addition to stops with labial and velar place of articulation (Bundgaard-Nielsen, Baker, Kroos, Harvey, & Best, 2015; Bundgaard-Nielsen, Kroos, Baker, Best, & Harvey, 2016). Wubuy does not have a voicing distinction in stops, nor a fortis-lenis (long-short) stop contrast found in other languages in the area (see Fletcher & Butcher, 2014 for discussion). Like most other Australian languages, Wubuy is also unusual cross-linguistically in that it has no fricatives, though the fricative /s/ occurs in one lexical item /sa/ – an exclamation in frequent use to shoo

away the many dogs that roam relatively freely in the community. The obstruent inventory of Wubuy is presented in Table 1.

<b>Lab.</b>	<b>Lam.- dent.</b>	<b>Apic.- Alv.</b>	<b>Apic- postalv.</b>	<b>Lam.- postalv.</b>	<b>Vel.</b>
p	t̪	t	t̺	ɸ	k

Table 1. The obstruent inventory of Wubuy.

Roper Kriol is an English-lexified creole which developed in the drainage basin of the Roper River in the late 19th and early 20th century as a result of contact between English speakers and speakers of traditional Indigenous languages (Harris, 1986; Sandefur, 1986; Munro, 2011). It is a lingua franca throughout South-Eastern Arnhem Land and adjacent regions, and a major variety of the largest Indigenous language in Australia, apart from English. There are an estimated 20,000 L1 speakers of Kriol (AIATSIS, 2005), including speakers of closely related varieties such as Roper Kriol (Baker, Bundgaard-Nielsen, & Graetzer, 2014; Bundgaard-Nielsen & Baker, 2016) and Fitzroy Crossing Kriol (Hudson, 1983), across Northern Australia.

According to recent work (Baker, Bundgaard-Nielsen & Graetzer, 2014; Bundgaard-Nielsen & Baker, 2016), the obstruent inventory of Roper Kriol is English-like in its stop voicing distinction: Roper Kriol stop contrasts are based on a short-lag versus long-lag VOT distinction, similar to that in English (the Indigenous substrate languages of Kriol, including Wubuy, do not have such a distinction). Notably, however, the VOT of voiceless stops in Kriol appears to be more extreme than what is typically found in English, despite the origins of this VOT based distinction. This is perhaps a result of target overshoot, as the English stop distinction was incorporated and grammaticalised in Kriol by speakers of Indigenous languages that did not previously use VOT contrastively – an interpretation in line with the above observations that greater acoustic differentiation of nonnative phones is helpful to the nonnative listener (Beach, Burnham, & Kitamura, 2001; Pater, 2003; Tsukada, 2004). Also similarly to English, Roper Kriol relies on a vowel duration difference to distinguish voiced from voiceless stops in syllable-final positions, such that vowels preceding a voiced syllable-final stops are longer than those preceding voiceless syllable final stops.

Despite these clear segmental affinities with English, Roper Kriol also unquestionably exhibits traits from some of the substrate Indigenous Australian languages of the region in terms of the constriction durations of stops (Fletcher & Butcher, 2014). Indeed, Kriol voiced and voiceless stops differ not only in terms of VOT, but also, in a decidedly un-Australian English-like fashion, in terms of their constriction duration, with voiceless stops having much longer duration than voiced stops. It is possible that this durational contrast is the primary cue to phoneme identity in stop contrasts in word-medial position (see Bundgaard-Nielsen & Baker, 2016 for evidence that the word medial realisation of a VOT contrast may be less robust than the realisation of a constriction duration difference, at least in child speakers of Kriol).

Kriol fricatives also differ from English in the absence of voicing-based contrasts. Indeed, all Kriol fricatives are voiceless in every position in the word, though [v] frequently occurs as a lenited realization of Kriol /b/, a process also characteristic of the Kriol substrate languages, including Wubuy. The obstruent inventory of Roper Kriol (Baker, Bundgaard-Nielsen, & Graetzer, 2014) is presented in Table 2.

Lab.	Dent.	Alv.	Retrofl.	Alv.- pal.	Vel.	Glott.
p b	t̪ d̪	t d	ɖ		k g	
				tʃ dʒ		
f		s		ʃ		h

Table 2. The obstruent inventory of Roper Kriol.

## 2. Method

### 2.1 Stimuli

We recorded three female speakers of Australian English in a recording studio at Melbourne University. All speakers were from the Greater Melbourne area in Victoria, Australia, and all had native English-speaking parents. All had substantial phonetics training. None reported having fluency in any language other than English, though all had studied other languages in a foreign language program in a high school or university setting.

The speakers produced five repetitions of the target consonants /p b k/ in an /aCa/ (i.e. intervocalic) context for Study 1, as well as five repetitions of the target consonants /b v s z ʃ/ in a /##Ca/ (i.e. utterance-initial) context for Study 2. The speakers were encouraged to familiarise themselves with the nonsense word list prior to the recording, to ensure a natural and highly fluent delivery. During the recording, the women were instructed to speak in a clear, comfortable voice as though they were speaking to a friend. All dysfluent and mispronounced tokens were re-recorded. All recordings had a 16-bit sampling depth with a sampling rate of 44.1 KHz.

All recorded tokens were segmented by hand, and measures of the preceding vowel duration and F0 (Experiment 1), the following vowel duration and F0 (Studies 1 and 2), as well as VOT and constriction duration were extracted using a custom-made *praat* script (Boersma & Weenink, 2010). Three tokens per target consonant per speaker (9 unique tokens) were selected as stimuli for the perception studies on the basis of the greatest possible similarity in terms of speaking rate, vowel duration, F0, and intonation pattern. Finally, each excised token was enveloped with a 20 ms ramp-in and a 10 ms ramp-out.

The selected /apa/, /aba/, and /aka/ tokens recorded and selected for use in Study 1 allowed the creation of a control contrast involving the discrimination of English /p/ and /k/, which differ in place of articulation rather than voicing; an English /p b/ contrast testing the participants' ability to discriminate bilabial stops that differ (primarily) in terms of VOT; and a Kriol-like /p b/ contrast which differs not only in terms of VOT, but also in terms of constriction duration.

In order to test discrimination of a Kriol-like /p b/ voicing distinction, i.e., one which is maintained by both a VOT as in English and by constriction duration, the duration of the silent constriction phase of the English /p/ tokens was manipulated to create a 'Kriol-like' /p/ (henceforth /p+/). The average constriction duration difference between Kriol /p/ and /b/ is approximately 60 ms, in clear lab-like speech, commensurate with the speech used in the present study (see Baker et al., 2014), while the average /p b/ stop constriction (CD) difference in the English targets recorded for this study is 10 ms. Consequently, 50 ms of silence was generated, mid closure, for each intervocalic English /p/ token from Study 1, in order to create plausible Kriol-like /p+/ tokens, which maintain their natural variation in VOT.



The recorded /ba/, /va/, /sa/, /za/, and /ʃa/ tokens selected for use in Study 2 allowed the creation of three contrasts testing the participants' ability to discriminate English fricatives /s ʃ/ and /s z/, and syllable initial /b v/. Similarly to Study 1, Study 2 includes a control contrast, /s ʃ/, based on a difference in place of articulation for speakers of both Australian English and Kriol, as well as the voicing based test contrast /s z/ and the constriction-based contrast /b v/.

## **2.2 Experimental design**

The study consists of two randomized, cross-speaker, categorical XAB discrimination tasks with speakers of Wubuy, Kriol and Australian English (control group). Study 1 tested discrimination of English intervocalic stops /p k/, /p b/, and the Kriol-like manipulated contrast of /p+ b/. Study 2 tested discrimination of syllable-initial English /s ʃ/ and /s z/, and /b v/. Each of the six contrasts (/p k/, /p b/, /p+ b/ in Study 1, and /b v/, /s ʃ/ and /s z/ in Study 2) were presented to the listeners in 6 unique triads, with 12 repetitions per triad, equaling 72 triads/contrast per listener. The task was explained to the participants as one in which a 'teacher' (first voice) was being imitated by a 'good student' and a 'bad student' (voices two and three). The participants then had to indicate (with a key press on the keyboard) which was the 'good student' who copied the teacher correctly. While this type of contextualization is not generally provided in speech research of this type, often conducted with university students, this approach was adopted as it has previously proved very helpful to participants from an Indigenous Australian background, and with limited computer literacy (see Bundgaard-Nielsen et al., 2015).

The discrimination tasks were programmed in Psyscope (Cohen, MacWhinney, Flatt, & Provost, 1993), with the stimuli presented over headphones from a MacBook computer. For both studies, the inter stimulus interval (ISI) was 500 ms, while the response window was presented for three seconds. The inter-trial interval was one second. All missed trials were replayed, at a random time, during the remainder of the test. The duration of the experiment ranged from approximately 45 minutes to an hour.

Despite widely accepted best-practice recommendations of counterbalancing the order in which the participants complete multiple tasks, all participants completed Study 1 first, and the order of presentation of the blocks comprising each of the studies was kept constant (in Study 1: /p k/, /p b/, /p+ b/; in Study 2: /b v/, /s ʃ/, /s z/). This decision reflects previous

observations of high rates of participant loss when blocks of ‘difficult’ nonnative contrasts are presented before participants are confident with the testing procedure. This decision also reflects prohibitive rates of participant loss when the study was initially piloted with a design that presented participants with blocks of randomized trials involving all six contrasts /p k/, /p b/, /p+ b/, /b v/, /s ʃ/, and /s z/, rather than trials blocked by contrast type.

## **2.3 Participants**

### **2.3.1 Wubuy**

11 native speakers of Wubuy (approximate age range 25-65 years) participated in the present study. One of these participants did not complete the /p b/ discrimination task, and another failed to complete the /s ʃ/ task due to technical problems, but the data collected from all 11 participants were included in the analyses. Some of the participants were literate and some semi-literate in Wubuy, as well as English (the medium of instruction at school). The Wubuy-speaking participants also spoke community language Roper Kriol to varying levels of proficiency. Another four Wubuy speakers were tested but excluded from the analyses for the following reasons: three failed to understand the task, and one was decided to withdraw due to fatigue. All testing took place in a quiet home in Numbulwar, in the Northern Territory of Australia. All procedures were explained in English as well as in Wubuy by a native speaker, assisting with translation when needed. Each participant was compensated for their time and effort by a \$100 payment.

### **2.3.2 Roper Kriol**

11 native speakers of Kriol (approximate age range 18-50 years) participated in the study. One of these participants failed to complete the /b v/ and the /s z/ tasks, while another failed to complete the /s z/ task, again due to technical problems. Data from all participants were included in the analyses. All participants were literate (to some extent) in English and had some competence in reading and writing Wubuy and Kriol. Kriol is not formally taught at school in Numbulwar, and while some participants had some Kriol literacy instruction through church activities, others were autodidact, mainly through the use of social media (texting on mobile phones, facebook, etc.). The testing conditions and compensation were identical to those of the Wubuy speakers. A Kriol translator was available when needed.

### 2.3.3 Australian English

13 native speakers of Australian English (*Mean* age 20 years; range 18-33 years) participated in the study. Data from all participants were included in the analyses. All participants were University of Melbourne undergraduates and recruited by word of mouth. Most had some competence in at least one other language acquired through formal instruction in a primary or secondary school setting. One participant was excluded due to a history of learning disorders, another due to having Italian-speaking background: Italian VOT distinctions differ systematically from those found in English, and moreover, Italian features long and short consonants (one of the parameters tested in the present study). All testing took place at University of Melbourne. Each participant was compensated for their time and effort by a \$30 payment.

## 2.4 Predictions

On the basis of PAM/PAM-L2 (PAM: Best, 1994; Best 1995; PAM-L2: Best & Tyler, 2007), it is possible to make one general prediction, as well as a number of language-specific predictions, outlined below.

Firstly, **all** listeners will discriminate the (TC) control contrast /p k/ successfully, though it is likely that the three participant groups may appear to achieve different levels of discrimination accuracy, and what constitutes ‘success’ may differ between the groups. In the case of a native/native-like control contrast such as /p k/, this is unlikely to reflect differences in perceptual acuity or ease and much more likely to reflect quite substantial differences in task familiarity, confidence and other non-linguistic and task-specific competences (differences in literacy achievement included). For a discussion of such differences in discrimination accuracy between Wubuy and Australian English participants, see Bundgaard-Nielsen et al., 2015. We argue that this particular point deserves careful attention as this has considerable bearing on the meaningfulness of conducting statistical comparisons between the three participant populations: Meaningful comparison requires that the participants differ only in terms of the variable of interest (here, native language), and this cannot be assumed in the present study.

Secondly, **Wubuy** listeners will perceive /p p+ b/ as instances of Wubuy /p/ and fail to discriminate them (SC contrast). Discrimination of /b v/ will be moderate as listeners will perceive /b/ as a good and /v/ as a ‘less good’ instance of Wubuy /b/ (CG discrimination). Discrimination of /s ʃ/

will be moderate (rather than poor) due to 1) experience with multiple place of articulation contrasts in the alveolar region and, 2) the occurrence of /s/ in the single, highly frequent, word /sa/! (an exclamation used exclusively to shoo away camp dogs), resulting in listeners perceiving /s/ as a ‘good’ and /ʃ/ as a ‘less good’ instance of marginal Wubuy /s/ (CG contrast). Discrimination of /s z/ will be poor as both are instances of /s/ and listeners have no L1 experience with fricative voicing contrasts (SG contrast). **Kriol** listeners will perceive /p p+/ as instances of Kriol /p/, and /b/ as Kriol /b/ and discriminate them though /p b/ will be discriminated less successfully than /p+ b/ due to the lack of native Kriol-like duration differentiation (TC contrasts). Discrimination of /b v/ will be moderate as Kriol listeners will perceive /b/ as a good and /v/ as a ‘less good’ instance of Kriol /b/ (CG discrimination). Discrimination of the place-of-articulation contrast /s ʃ/ will be excellent as this is a native TC contrast. Finally, discrimination of /s z/ will be poor as both are instances of /s/ and Kriol speakers have no experience with fricative voicing (SG contrast). **Australian English** listeners will successfully discriminate all native contrasts, including the enhanced Kriol-like /p+ b/ contrast.

### 3. Results

The discrimination accuracy for each of the three participant groups (Wubuy; Kriol; Australian English) is presented in Figure 2 (Study 1) and Figure 3 (Study 2) below. The average discrimination accuracy of the Wubuy participants was 59% (Study 1) and 64% (Study 2), with an *M* accuracy of the control condition /p k/ of 68%, while the average discrimination accuracy for the Kriol participants was 66% (Study 1) and 65% (Study 2), with an *M* accuracy of the control condition /p k/ of 73%. The average discrimination accuracy for the Australian English-speaking participants was 95% (Study 1) and 96% (Study 2), with an *M* accuracy of the control condition /p k/ of 94%.

In the following sections, we present statistical analyses of the results from the three participant groups separately. We do not formally compare the discrimination accuracy of the three groups, as the averages reported above, as well as the group discrimination accuracy means for the control contrast /p k/ clearly indicate systematic differences in performance, most likely unrelated to the variable of interest of native language background.

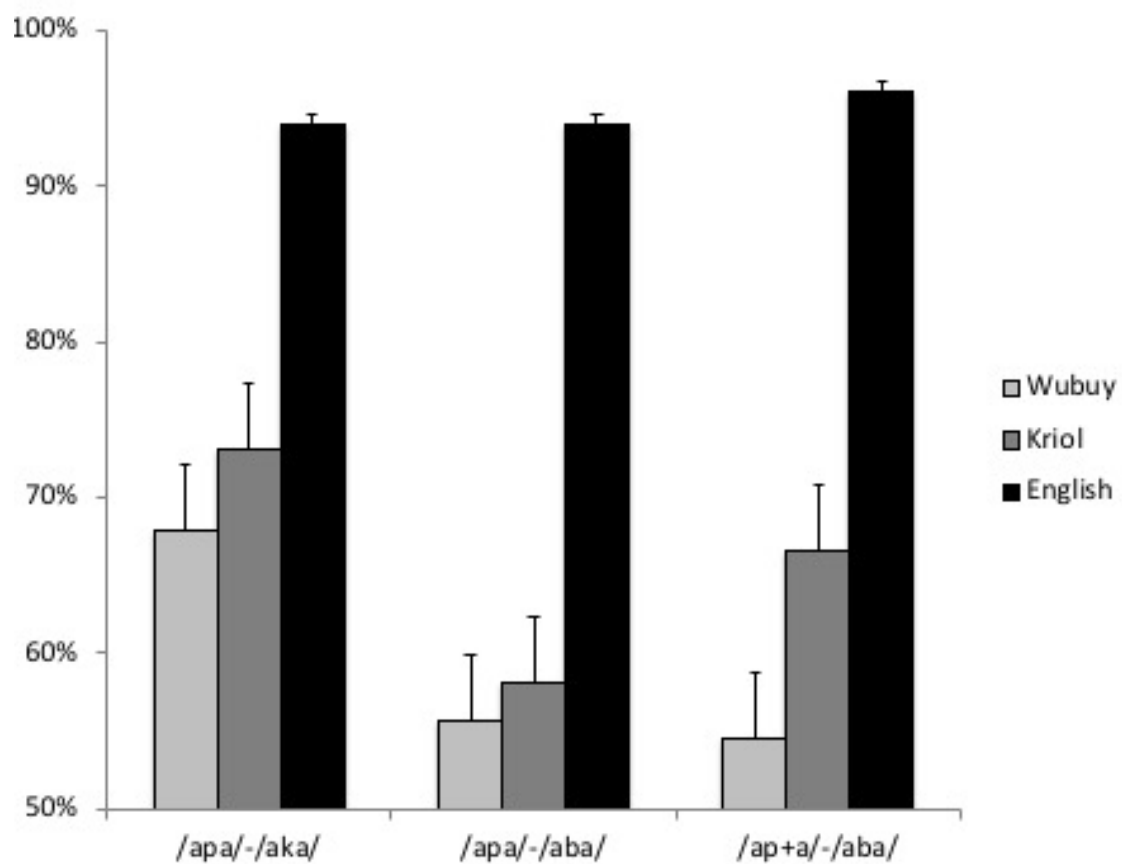


Figure 2. Mean discrimination accuracy for Wubuy, Kriol and English speakers in Study 1. 50% indicates chance performance. Error bars indicate S.E.

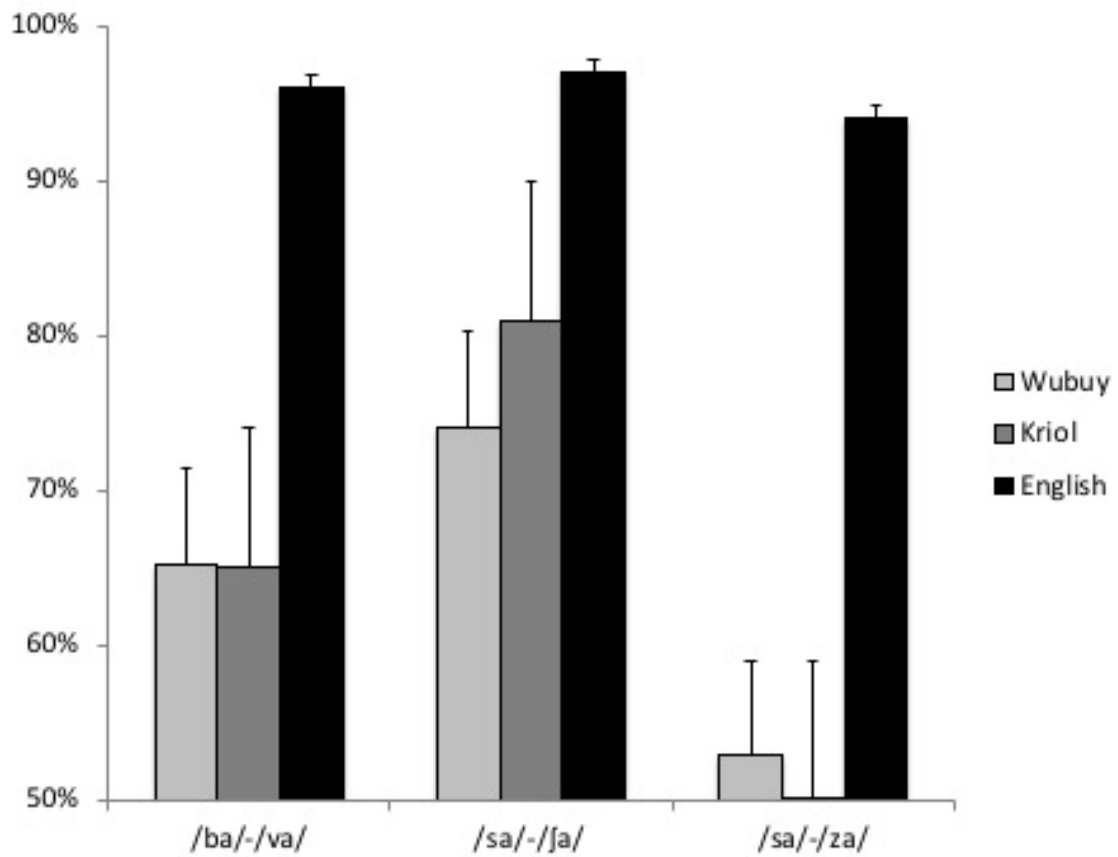


Figure 3. Mean discrimination accuracy for Wubuy, Kriol and English speakers in Study 2. 50% indicates chance performance. Error bars indicate S.E.

### 3.1 Wubuy results

To assess whether the Wubuy listeners were able to discriminate the target contrasts, including the control contrast /p k/ in Studies 1 and 2, we first conducted a series of one-sample *t*-tests against chance performance. The results indicate that the Wubuy speakers generally are able to discriminate four of the six contrasts above chance, including, importantly, the native-like control contrast /p k/ ( $p=.01$ ), /p b/ ( $p=.025$ ), /s ʃ/ ( $p<.001$ ) and /b v/ ( $p=.013$ ). The Wubuy speakers' discrimination accuracy for the Kriol-like /p+ b/ ( $p=.079$ ) and /s z/ ( $p=.204$ ) did not differ significantly from chance performance. Two separate One-Way ANOVAs revealed a significant main effect of contrast for each of the two studies (Study 1:  $F(2,28)=6.275$ ,  $p=.006$ ; Study 2:  $F(2,28)=12.535$ ,  $p<.001$ ). Post-hoc Bonferroni-corrected comparisons confirmed that the main effect of contrast in Study 1 was due to a significant discrimination accuracy difference between /p k/ and the other two contrasts (/p b/ and /p+ b/:  $p=.015$  for both). Post-hoc Bonferroni comparisons of the difference in contrast discrimination in Study 2 likewise confirmed that the main effect was due to the poor discrimination accuracy for the voicing based distinction /s z/ relative to the place of articulation-contrast /s ʃ/ and the manner of articulation contrast /b v/ ( $p=.001$  for both).

The results from Study 1 are fully consistent with the PAM-based predictions above and suggest that L2 acquisition of voicing-based contrasts is extremely difficult when the learner's L1 has led him/her to consistently ignore the feature 'voicing'. The Wubuy speakers find English VOT-based labial stop contrasts very difficult to discriminate. This is also true of the Kriol-like labial stop contrast based on VOT and duration differences: unlike other languages of the area, such as nearby, related Ngandi (Heath, 1978), Wubuy does not implement a stop contrast based on duration or any other correlate. The results from Study 2 are also consistent with the predictions: Wubuy speakers are unable to discriminate the voicing-based fricative distinction /s z/, though they can discriminate the CG contrasts /s ʃ/ and /b v/.

### 3.2 Kriol results

To assess the Kriol discrimination performance, we first conducted a series of one-sample *t*-tests against chance performance, which indicate that the Kriol speakers are able to discriminate all contrasts above chance level ( $p<.01$  for /p k/, /p b/, /s ʃ/;  $p=.05$  for /p+ b/). The contrast /b v/ approached significance ( $p=.06$ ); but /s z/ was clearly not significantly different from

chance ( $p=.956$ ). Two separate One-Way ANOVAs revealed a significant main effect of contrast for each of the two studies (Study 1:  $F(2,30)=4.386$ ,  $p=.021$ ; Study 2:  $F(2,27)=16.017$ ,  $p<.001$ ). Subsequent Bonferroni post-hoc comparisons revealed that the main effect in Study 1 was due to English /p b/ being less accurately discriminated than the control contrast /p k/ ( $p=.018$ ). There was no significant difference in discrimination accuracy for /p+ b/ and /p b/ ( $p=.316$ ), nor in the discrimination accuracy of /p k/ and /p+ b/ ( $p=.627$ ). In Study 2, Bonferroni post-hoc comparisons revealed that discrimination accuracy of /s ʃ/ was higher than the discrimination accuracy for /b v/ ( $p=.018$ ) and /s z/ ( $p<.001$ ). The discrimination accuracy of /b v/ was also greater than the discrimination accuracy of /s z/ ( $p=.037$ ).

The results from Study 1 suggest that Kriol speakers rely on duration as a means of distinguishing the voicing contrast, although the difference in performance with the lengthened contrast /p+ b/ versus /p b/ was not significant. However, the fact that /p+ b/ was not significantly different from /p k/, but /p b/ was, also suggests a difference not reflected in the statistical inference: that detecting voicing without a concomitant duration difference is harder for Kriol speakers than detecting a simple place difference. The performance of the Kriol listeners in Study 2 supports the conclusion drawn on the basis of the Wubuy participants' results: lack of native language experience with a voicing contrast (for Kriol listeners: with fricatives only) leads to an inability to discriminate that contrast, even for L2 learners with extensive L2 exposure. Interestingly, however, in the case of the Kriol listeners, their experience *with* voicing contrasts in stops does not translate to an ability to perceive this characteristic in fricatives. We return to this point in the discussion.

### 3.3 Australian English results

Finally, a series of one-sample *t*-tests against chance performance indicated that – as is apparent from Figures 2 and 3 – the English listeners' discrimination of all six contrasts was significantly better than chance ( $p<.001$ , in all cases). A final set of One-Way ANOVAs revealed there was no significant effect of contrast for either Study 1 ( $F(2,36)=2.153$ ,  $p=.131$ ) or Study 2 ( $F(2,36)=1.003$ ,  $p=.377$ ).

These results, unsurprisingly, provide evidence that the English listeners are well able to discriminate all the English obstruent contrasts included in Study 1 (stops) and 2 (fricatives), as these of course straight-forwardly map onto their native English obstruent categories. The fact that the discrimination accuracy for the Kriol-like /p+ b/ contrast is on par with the



discrimination accuracy for the original English /p b/ contrast suggests that the additional CD difference did not disturb or disrupt the listeners' ability to discriminate, either because they continued to pay attention to the VOT difference alone, or because the CD co-varied with the VOT difference and thus was consistent with the VOT-based discrimination. The very high accuracy in all tasks, including /p b/ and /p+ b/, makes it difficult to assess whether the added durational cue resulted in increased discrimination accuracy.

#### 4. Discussion

The present studies offer a first systematic assessment of the perception of non-native voicing distinctions (in stops and fricatives) by speakers whose native language does not make use of such distinction (here, Wubuy). It also examines the perception of non-native stop and fricative voicing distinctions by speakers whose native language (Northern Australian Kriol) uses VOT and stop duration to maintain stop voicing distinctions, but does not make a voicing-based distinction in fricatives.

The results of the present studies show – unsurprisingly – that L1 background systematically shapes perception of L2 phonological contrasts that (1) do not align with L1 phoneme boundaries, or (2) differ drastically from the L1 phonemes in their phonetic realisation. Indeed, they show that native speakers of Wubuy, which is characterized by a single series of stops and the absence of fricatives altogether, find the discrimination of nonnative (English) voicing-based English stop and fricative voicing distinctions extremely difficult, even after years of exposure to, and use of, English as a second language. The results also show that the addition of a second acoustic cue to the distinction of voiced and voiceless stops (stop duration) does not lead to improved performance in stop voicing discrimination for these participants, despite extensive exposure to Kriol (as a community language spoken widely in the area where the Wubuy speakers live). The results also show that speakers of Kriol who rely on stop duration in addition to VOT to differentiate voiced and voiceless stops are less accurate in discriminating (English) stop contrasts that lack the durational cue but are consistent in VOT differentiation. The Kriol speakers also demonstrate that they find the application of voicing as a contrastive feature difficult in the case of the discrimination of the non-native English voiceless-voiced fricative contrast /s z/. Finally, the results suggest that adding to the number of phonetic cues available, here by creating Kriol-like stop-contrasts that differ in both VOT and constriction duration, does

not impair the performance of participants who either successfully continue to rely on their native voicing cue (VOT) exclusively, or successfully incorporate a co-varying secondary cue (constriction duration) into their perception.

The results from Studies 1 and 2 presented here are consistent with results of previous studies (see the Introduction above). These well-established findings suggest that difficulties in non-native obstruent perception can arise from differences in the L1 and L2/non-native phonetic realization of shared/overlapping phonological categories (as is the case with the perception of English stop-voicing distinctions by speakers of Spanish, and here speakers of Kriol). The results are also consistent with findings which suggest that difficulties can arise due to differences in the phonological inventories of the L1 and the L2/non-native language (as is the case with the perception of the Thai three-way stop voicing distinction by speakers of English, and the perception of English and Kriol-like stop voicing distinctions by speakers of Wubuy in the present study). The present study however tests this second point in the novel context of testing discrimination of obstruent voicing distinctions by participants who do not have native language experience with this parameter.

This particular aspect is of importance to theories of both first and second language acquisition, as it indicates that L2 phonemic learning may be near-impossible if a learner's L1 has not provided him/her with some familiarity with voicing used as a contrastive feature (or with constriction duration-based). This observation is consistent with PAM/PAM-L2 predictions that SC contrasts (as opposed to CG contrasts) can pose persistent difficulties for learners as both non-native phones may represent phonetically perfectly good instances of a given native phone. Indeed, this is likely to be the case for speakers of Wubuy tasked with discriminating voiced and voiceless English stops as the primary distinguishing feature of such stops are to be found in a linguistic dimension to which the listener is not attending. In other words, the creation of a perceptual boundary within what a learner perceives to be a singular – and importantly, linguistically irrelevant – dimension through re-attunement and rephonologisation is unlikely (see discussion in Best et al., this volume). They also indicate that familiarity with a particular linguistic dimension, here voicing, in one phonemic domain (stops), does not necessarily translate to discrimination ability in another (fricatives), despite both of these categories being phonologically classified as obstruents, and thus belong to the category where (if anywhere) we expect voicing to be implemented phonemically.

This result leads us to question the extent to which a phonological feature such as [ $\pm$ voice] can be said to be activated by native language input – a question of central importance to any consideration of the Articulatory Organ Hypothesis (for a discussion see Best et al., this volume). We find this question particularly important for theories of segmental acquisition and organization, given that the Wubuy and Kriol listeners are regular users and likely end-state second language learners of English, and appear to behave very differently from native speakers of English, and from each other with respect to their ability to perceive voicing-based obstruent contrasts.

### Acknowledgements

We wish to thank the Wubuy, Kriol and English listeners and speakers who participated in the present studies. We would also like to thank the Wubuy and Kriol translators who assisted with data collection, as well as Josh Clothier for assisting with the testing of the English control group. We also thank Dr. Thomas Britz for an elegant Fig. 1. We gratefully acknowledge the Australian Research Council for generously supporting this research through DP130102624 ‘Learning to talk Whitefella Way’.

### Author Note

This research was supported in part by the Australian Research Council Discovery Project DP130102624 ‘Learning to talk Whitefella Way’.

Correspondence concerning this article should be addressed to Rikke Louise Bundgaard-Nielsen, 14 High Street, Coburg VIC 3058, Australia

Contact: rikkelou@gmail.com

### References

- Abramson, A. S., & L. Lisker L. (1970). Discriminability along the voicing continuum: Cross-language tests. In B. Hala, M. Romportl, & P. Janota (Eds.), *Proceedings of the Sixth International Congress of Phonetic Sciences* (pp. 569-573). Prague: Academia.
- Abramson, A. S., & L. Lisker L. (1973). Voice-timing perception in Spanish word-initial stops, *Journal of Phonetics*, 1, 1-8.
- Abramson, A. S., & Tingsabadh, K. (1999). Thai final stops: Cross-language perception. *Phonetica*, 56, 111-122.
- Abramson, A. S., & Whalen D. (2017). Voice Onset Time (VOT) at 50: Theoretic-

- cal and practical issues in measuring voicing distinctions, *Journal of Phonetics*, 63, 75-86.
- Australian Institute of Aboriginal and Torres Strait Islander Studies/Commonwealth of Australia (2005). *The National Indigenous Languages Survey Report*. Canberra: Department of Communications, Information Technology and the Arts.
- Baker, B., Bundgaard-Nielsen, R. L., & Graetzer, S. (2014). The obstruent inventory of Roper Kriol, *Australian Journal of Linguistics*, 34(3), 307-344.
- Beach, E. F., Burnham, D., & Kitamura, C. (2001). Bilingualism and the relationship between perception and production: Greek/English bilinguals and Thai bilabial stops. *The International Journal of Bilingualism*, 5(2), 221-235.
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman, & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, Massachusetts: MIT Press.
- Best, C. T. (1995). A direct-realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In J. Munro, & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34). Amsterdam: John Benjamins.
- Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program], Version 5.1.44.
- Bundgaard-Nielsen, R. L., & Baker, B. J. (2016). Fact or Furphy? The Continuum in Kriol. In F. Meakins, & C. O'Shannessy (Eds.), *Loss and renewal: Australian languages since contact* (pp. 177-216). Berlin: Mouton de Gruyter.
- Bundgaard-Nielsen, R. L., Kroos, C., Baker, B., Best, C. T., & Harvey, M. (2016). Consonantal timing and release burst acoustics distinguish multiple coronal stop place distinctions in Wubuy (Australia). *Journal of the Acoustical Society of America*, 140(4), 2794-2809.
- Bundgaard-Nielsen, R. L., Baker, B. J., Kroos, C., Harvey, M., & Best, C. T. (2015). Discrimination of multiple coronal stop contrasts in Wubuy (Australia): A Natural Referent Consonant account. *PloS ONE*, 10(12), e0142054. doi:10.1371/journal.pone.0142054
- Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27, 207-229.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments, *Behavioral Research Methods, Instruments, and Computers*, 25(2), 257-271.
- Davis, K. (1995). Phonetic and phonological contrasts in the acquisition of voic-

- ing: Voice onset time production in Hindi and English. *Journal of Child Language*, 22, 275-305.
- Flege, J. E. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, 15, 67-83.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 233-277). Timonium, MD: York Press.
- Fletcher, J., & Butcher, A. (2014). Sound patterns of Australian languages. In H. Koch, & R. Nordlinger (Eds.), *The languages and linguistics of Australia: A comprehensive guide* (pp. 91-138). Berlin: Walter de Gruyter.
- Gopal, H. S. (1993). VOT values of voiceless and voiced stop contrasts in Hindi and Kannada. *Journal of the Acoustical Society of America*, 93, 2298.
- Harris, J. W. (1986). *Northern Territory Pidgins and the origin of Kriol*, Series C, vol 89. Canberra: Pacific Linguistics.
- Heath, J. (1978). *Ngandi grammar, texts, and dictionary*. New Jersey: Australian Institute of Aboriginal Studies.
- Heath, J. (1984). *Functional Grammar of Nunggubuyu*. Canberra: Canberra: Australian Institute of Aboriginal Studies.
- Henton, C. G., Ladefoged, P., & Maddieson, I. (1992). Stops in the world's languages. *Phonetica*, 49, 65-101.
- Hudson, J. (1983). *Grammatical and semantic aspects of Fitzroy Valley Kriol (Work Papers of SIL-AAB A8)*. Darwin: Summer Institute of Linguistics.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge studies in speech science and communication. Cambridge: Cambridge University Press.
- Munro, J. (2011). Roper River Aboriginal language features in Australian Kriol: Considering semantic features. In C. Lefebvre (Ed.), *Creoles, their substrates, and language typology* (pp. 461-487). Amsterdam: John Benjamins.
- Pater, J. (2003). The perceptual acquisition of Thai phonology by English speakers: task and stimulus effects. *Second Language Research*, 19(3), 209-223.
- Pisoni, D. R., Aslin, A., Perey, & Hennessy, B. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 297-314.
- Sandefur, J. (1986). *Kriol of North Australia: A language coming of age*. Darwin: Summer Institute of Linguistics.
- So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language & Speech*, 53(2), 273-293.
- So, C. K., & Best, C. T. (2011). Categorizing Mandarin tones into listeners' native

- prosodic categories: The role of phonetic properties. *Poznań Studies in Contemporary Linguistics*, 47, 133.
- So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, 36(2), 195-221.
- Strange, W. (1972). *The effects of training on the perception of synthetic speech sounds: Voice onset time* (Ph.D. dissertation). Minnesota: University of Minnesota.
- Tingsabadh, K., & Abramson, A. (1993). Thai. *Journal of the International Phonetic Association*, 23(1), 24-28.
- Tsukada, K. (2004). Cross-language perception of final stops in Thai and English: A comparison of native and non-native listeners. In *Proceedings of the 10th Australian International Conference on Speech Science & Technology, Macquarie University, Sydney, December 8 to 10*. Australian Speech Science & Technology Association Inc.
- Wu, M., Fletcher, J., Baker, B., & Bundgaard-Nielsen, R. (2016). How pitch moves: production of Cantonese tones by speakers with different tonal experience. *5th International Symposium on Tonal Aspects of Language, Buffalo*, 133-137.
- Wu, M., Fletcher, J., Bundgaard-Nielsen, R., & Baker, B. (2016b). Production of Cantonese tones by speakers with different tonal experiences. *Speech Prosody 2016, Boston*, 133-137.
- Wu, M., Bundgaard-Nielsen, R., Baker, B., Best, C. T., & Fletcher, J. (2015). Perception of Cantonese tones by Mandarin speakers. *18th International Congress of Phonetic Sciences, Glasgow*.



## Acoustic Comparison of Mandarin and Danish Postalveolars

Sidsel Rasmussen  
Aarhus University

Mengzhu Yan  
Victoria University of Wellington

### Abstract

This paper examines similarities between the series of Mandarin “palatal” and “retroflex” affricates and fricatives. The distinct sets of Mandarin phones are known to be perceived as similar by nonnative listeners whose first language does not deploy the same postalveolar place contrast. Danish, for example, has phonological onset clusters articulated as alveolo-palatal sibilants, and previous studies indicate a tendency for Danish L1 speakers to confuse the two sets of Mandarin sounds. We obtain production data from native speakers of Danish and Mandarin and compare acoustic measurements to gain a better understanding of the cross-linguistic similarities and explore some of the perceptual problems indicated from perceptual assimilation data.

### 1. General introduction

It is well known that the sound inventories of the world’s languages differ in how phonetic features distinguish contrastive categories, and a perceptual “retuning” may be one of the many tasks required of second language learners. Much of the existing literature on L2 acquisition research targets English sounds, but the body of literature targeting Mandarin Chinese is currently growing too, as more and more students around the world become interested in learning Chinese as a foreign language. The four lexical

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 65-80). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.



tones of Mandarin are notoriously difficult for learners of non-tonal languages to acquire, but what about acquisition at the segmental level? Mandarin Chinese has a rare three-way distinction of coronal sibilants: /ts, ts<sup>h</sup>, s/, /tɕ, tɕ<sup>h</sup>, ç/ and /tʂ, tʂ<sup>h</sup>, ʂ/, in the literature typically referred to as the series of “dentals”, “palatals” and “retroflexes”, respectively. These are used as cover terms rather than precise denotations of place of articulation, and we follow the standard terminology throughout this paper. The “dental” sibilants have the most fronted constriction with the tongue against the incisors. The “palatals” are produced in the postalveolar region with the blade of the tongue, and they are often referred to as alveolopalatal consonants to specify a more fronted constriction than the term otherwise suggests. The “retroflexes” are apical post-alveolars (W.-S. Lee & Zee, 2003). In the following, we briefly introduce some of the research on Mandarin sibilant similarity before addressing the consequences of this crowded phonetic inventory for nonnative speech perception.

### **1.1. Introduction to Mandarin sibilant similarity**

The Mandarin system of affricates and fricatives has been researched quite extensively. The palatal series is particularly interesting due to its seemingly predictable environment. By many analyses the palatals are said to appear only before the high front vowels [i] and [y] or the corresponding glides [j] and [ɥ], causing phonologists to treat them as allophones of other consonants (e.g. Hartman, 1944). A popular synchronic analysis posits them as allophones of the dentals (Duanmu, 2007), and previous studies have found the series of palatals to be acoustically most similar to the dental sibilants (S. Li & Gu, 2015; C.-Y. Lee, Zhang, & Li, 2014). However, the similarity (and perceptual confusability (Zhang, Lü, & Qi, 1982)) of this dental-palatal place contrast is presumably minimized by the nearly complementary distribution (Svantesson, 1986). Li and Zhang (2017) examined this claim experimentally and found that discrimination of the dental-palatal sibilant place contrast was less prone to errors when followed by [-high] nuclear vowels as opposed to an allophone of the high front vowel. Furthermore, they showed that the alternation of /i/ decreased the discrimination errors for both L1 and L2 listeners as well as increased listeners’ response time as opposed to the condition in which both palatal and dental sibilants preceded [i]. The contrast between easily confusable Mandarin consonants seems to be enhanced by the alternation in the vowel quality as shown for the dental-palatal contrast, and the same is likely true for the retroflex series, which, just like the dentals, also combines with

a homorganic apical vowel instead of [i]. But how does discrimination fare in vowel contexts that do not offer an enhanced distinction which can facilitate consonant discrimination? For the non-high main vowels the distributional patterns between palatals and the other sibilant series are less distinct, if indeed present at all. For example, Ladefoged and Maddieson (1996) argue that the so-called “palatal” fricative followed by /a/ as the nuclear vowel does not have an intermediate glide linking the consonant and the low vowel: “from a phonetic point of view there is nothing other than a normal transition between the initial consonant and the following vowel” (p. 150). The consequence of this statement is a contrast neutralization which implies that a syllable like 下 <xia> ([ɕa]) forms perfect minimal pairs with e.g. [sa] and [ʂa], and this view resonates with recent phonological analyses that object to the notion of a medial glide in the traditional CGVX model<sup>1</sup> of the Mandarin syllable. Instead, a more complex consonantal onset inventory and a simpler CVX syllable structure has been proposed (e.g. Ao, 1992; Duanmu, 2017). The debate about Chinese medial glides is long and complicated, but if we look to these recent analyses and suspend the prescribed notion of the palatals always preceding high front segments, then the proposed complementary distribution of the traditional phonological analyses should not deter us from investigating consonants that are at least by some accounts, and in some environments, indeed minimally contrastive. This also means that the three-way place distinction in Mandarin sibilants may indeed be phonemic. Acoustic comparisons of these similar, yet contrastive sets of phones yield further insights into the exact cues listeners must attend to for successful discrimination.

## **1.2. The palatal-retroflex contrast in non-native perception**

Norman (1988), among others, notes how L1 English learners of Mandarin tend to associate the palatals more closely with the retroflex series. This is quite interesting considering how the “palatal direction of confusability” is forward (towards the dentals) for native speakers, but backward (towards the retroflexes) for L1 English speaking learners. The L1 sibilant categories are very likely a source of influence here, given that English only has one place for articulating postalveolar sibilants, which might cause an “equivalence classification” of the distinct L2 phones. In this vein Chang

---

<sup>1</sup> In the traditional view the Mandarin syllable consists of maximally four segments of which only the nuclear vowel is obligatory: C (consonant) – G (glide) – V (vowel) – X (offglide or nasal coda).

et al. (2011) investigated the production of the Mandarin sibilants /s, ç, ʃ/ and the English /s, ʃ/ by L1 Mandarin speakers, native English speaking L2 learners of Mandarin, and heritage speakers of Mandarin who had grown up in English dominant communities. They found that the phones most likely to be merged in production across all groups were the Mandarin /ʃ/ and the English /ʃ/, but with heritage speakers distinguishing the five sibilants better than speakers in the two groups of late learners. Their findings point to an assimilation of the English and Mandarin postalveolars for both of the nonnative groups, which makes sense considering the similar articulatory descriptions of both fricatives. Danish, like English, has a contrast between alveolar and postalveolar sibilants: Specifically, Danish has a palatalized postalveolar sibilant [ç] produced either apically or laminally, typically denoted as an “alveolo-palatal” voiceless fricative, which is the phonetic realization of the onset sequence /s+j/ (Grønnum, 2009). Unlike English, however, Danish has no phonological affricates, but phonetically the clusters of /dj/ and /tj/ result in [tç] and [tç<sup>h</sup>], respectively.<sup>2</sup> Just like Mandarin consonants, these Danish sibilant clusters are phonotactically restricted to onset position, and their phonetic similarity to the Mandarin alveolo-palatal sibilants is evident from the identical IPA symbols alone, but this type of comparison between transcriptions should only serve as an “initial heuristics in attempts to establish similarity” (Bohn, 2002: 198). Rasmussen and Bohn (2017) therefore examined the cross-linguistic mapping of Mandarin initial consonants as perceived by naive native Danish speaking listeners to test the perceived similarity between Mandarin and Danish consonants. 24 Danish L1 listeners with no prior knowledge of Chinese languages listened to Mandarin CV syllables over headphones and assimilated the initial consonant to a native onset represented in standard Danish orthography corresponding to unambiguous phonetic categories. The forced identification task was supplemented by a 7-point Likert scale on which, for each trial, participants indicated the perceived similarity between stimulus and the selected L1 response category. Results indicated that palatals and retroflexes were assimilated to the same native category of onset clusters with varying degrees of “goodness” of the perceived match. The theoretical implications of a two-to-one L2-L1 mapping is posited in the framework of the Perceptual Assimilation Model (Best, 1995).

---

<sup>2</sup> We refer to the Danish sibilants by their phonemic representations throughout this paper in order to avoid confusion with the Mandarin alveolopalatals transcribed by the same symbols. We find it useful to deploy the impressionistic labels “lenis” and “fortis” when referring to the collective groups of unaspirated and aspirated affricates, respectively.

If two nonnative sounds are heard as equally good variants of a single native phoneme, discrimination difficulties are predicted. This is known as a Single Category assimilation type (SC). If the nonnative phones are assimilated to the same native category with *differing* ratings of the match, a Category-Goodness (CG) difference is observed, and discrimination between the L2 segments should be less problematic. Findings from the perceptual assimilation study revealed a strong effect of the following vowel on the cross-linguistic matching of Mandarin palatals. Considering the phonotactics of Mandarin as well as the restrictions imposed by the alternating vowel quality on Mandarin syllables, the data included here only lists results for the vowel context for /a/, which is preceded by both retroflexes and palatals.

DA Response	MA Stimuli					
	/tɕ/	/tɕ <sup>h</sup> /	/ç/	/tʂ/	/tʂ <sup>h</sup> /	/ʂ/
/dj/	<b>47 (5.53)</b>			<b>60 (5.14)</b>		
/tj/	35 (4.08)	<b>78 (5.16)</b>		22 (5.25)	<b>71 (3.98)</b>	
/sj/			<b>74 (4.11)</b>			<b>85 (4.85)</b>

Table 1. Confusion matrix of Mandarin (MA) stimuli presented before /a/ (horizontally) and selected Danish (DA) response (vertically). First number indicates percentage of match, and number in parentheses is average “goodness”. Bolded cells show most frequent match.

As Table 1 shows, the palatal-retroflex place contrast is not attended to by naïve Danish listeners, who identify both series of Mandarin tokens as above average exemplars of their articulatorily closest native counterpart. Interestingly, there does not seem to be a clear preference for Mandarin palatals which as the most similar to the Danish sibilants if we are to trust phonetic descriptions: only the palatal aspirated affricate fares better than its retroflex counterpart in both number of matches and similarity rating. The unaspirated Mandarin affricates are most frequently perceived as fairly good exemplars of Danish /dj/, but surprisingly, the fortis /tj/ response is also selected as the preferred match for 22-35% of trials. The results from this small part of a perceptual assimilation study yield two main questions of interest: Firstly, if Danish sibilants are produced as the Mandarin

palatals, why is there for unaspirated Mandarin sibilants seemingly a (small) preference for the retroflexes in a cross-linguistic mapping task? Secondly, why would the Mandarin unaspirated affricates sometimes be perceived as good exemplars of Danish /tj/? To shed light on the differences and similarities between the nine sibilants discussed here it will be useful to examine spectral properties of the contrasts. Only one study has so far examined the Danish-Mandarin sibilant contrast acoustically. Mikkelsen (2016) studied Danish L2 Mandarin learners' production of the L1 [ç], the L2 English [ʃ] and L3 Mandarin [ç] and [ʃ]. She measured COG<sup>3</sup> values for the four fricatives produced in different vowel contexts and found that the Mandarin postalveolars were produced more similarly preceding non-high vowels. This can probably be explained as an effect of the differences in vowel quality on the preceding consonant. In the /a/ condition, the retroflex and the Danish fricative were not produced significantly differently, while the other sibilants were statistically distinguished by the spectral means obtained. Mikkelsen's data, although limited to few tokens per target consonant, indicate that Danish learners of Chinese produce the retroflex fricative similar to their native category /sj/. It also indicates that there is an effect of following vowel, a finding in agreement with observations from other acoustic studies of native Mandarin production as well as the perceptual assimilation study mentioned above. Additional acoustic measures, a larger data set and inclusion of the homorganic affricates will likely provide additional insights to these previous cross-linguistic studies.

### **1.3. The current study**

In this paper we revisit the two-to-one mapping of Mandarin palatals and retroflex consonants to Danish onset clusters /dj, tj, sj/. As Bohn (2002) notes, direct comparison of cross-linguistic similarity is often not possible due to methodological differences between the studies that report acoustic measurements. The most informative and best comparable production data will require attention to factors such as elicitation method and phonetic environment of target sounds. In this study we present data from both languages, elicited specifically for the purpose of allowing for comparisons between relevant acoustic measures of the two languages in order to explore an additional aspect of the phonetic similarity indicated in a previous study. We compare our Mandarin data to existing literature on the acoustics of Mandarin sibilants. We are not aware of any acoustic studies examining the

---

<sup>3</sup> Center of gravity is the spectral mean, often used to classify aperiodic speech sounds such as fricatives.

Danish target categories under investigation, so this study can hopefully both serve the purpose of detailed acoustic comparisons between Mandarin and Danish postalveolars, probing the questions raised from a perceptual assimilation study, as well as provide baseline acoustic measures for three Danish consonant clusters.

## **2. Methods**

### **2.1. Participants**

Five female native speakers of each language were recorded. The five native Mandarin speakers (mean age: 27.6, SD: 3.0) were recorded in New Zealand, but all reported frequent use of L1 Mandarin Chinese in their everyday communication. The Danish speakers had a mean age of 26.6 (SD: 3.1) and were recorded in Aarhus, Denmark. A basic language background questionnaire ensured that they had no prior experience with any Chinese languages. All ten speakers participated as volunteers in this study, and none of them reported any speech or hearing problems.

### **2.2. Recordings**

Four Mandarin affricates /tʂ, tʂʰ, tɕ, tɕʰ/ and two fricatives /ʂ, ʃ/ were included in the Mandarin stimuli list. The target consonants were combined with /a/ and the falling tone (T4). Each target phoneme was repeated ten times, resulting in a total of 300 tokens (6 target consonants \* 1 vowel \* 1 tone \* 10 repetitions \* 5 speakers). The syllables were embedded in the middle position of a carrier sentence 我把\_\_读出来 (“I take \_\_ and read out loud.”) and the corresponding Pinyin Romanization and tone number were also written after the target syllable. The Danish stimuli were created with the intent to provide a context as similar to the Mandarin syllables as possible. Unlike the Chinese stimuli, the Danish target syllables do not correspond to morphemes, so elicitation relied on speakers’ production of nonsense syllables for which rhyming real words were provided prior to recording. To achieve similarity in vowel context we created a list of open syllable non-words that combined the target Danish onsets with a low back vowel [ɑ]. The back vowel allophone surfaces due to a fusion between /a/ and /r/ (Basbøll, 2005: 149), so to ensure the anticipated vowel quality in the targets, the orthographic representations listed were <djɑr>, <tjɑr>, <sjɑr>. High frequency words (e.g. ‘har’ [hɑʔ]<sup>4</sup> (present tense of “to

---

<sup>4</sup> We anticipated glottal stop (stød) on the open monosyllables in focus position. Mandarin T4 with a falling contour has the shortest duration (Ho, 1976) and T4 syllables are perceptually not unlike syllables with stød.

have”)) were provided as rhymes to targets. All Danish targets and fillers were produced in the carrier sentence *Jeg siger \_\_ til dig*. (“I say \_\_ to you.”). Both sets of stimuli were pseudo randomized so that no more than two identical sentences occurred together. Non-target fillers were included as the last sentence in the written material to avoid end of list intonation. For the longer set of stimuli (the Mandarin list), breaks were provided. Each session lasted approximately 1-2 minutes.

### 2.3. Measurements

The segmentation and measurements of the target fricatives and affricates were handled separately by the authors in Praat (Boersma & David Weenink, 2018). The task was divided between the authors so the stimuli matched the L1 of the author in order to also perceptually verify the onset labels, which resulted in the discarding of 1 Mandarin and 4 Danish tokens. Speech segmentation was conducted on the basis of waveform and wideband spectrogram. For fricatives, the entire noise portion was classified as the target consonant. For affricates, frication noise was defined between the beginning of the burst and the onset of the vowel. A number of studies have provided acoustic analyses for Mandarin, and different acoustic measurements have been found to differentiate the Mandarin sibilants. For the place distinction, spectral moments (notably the spectral mean, i.e. COG) and F2 frequency at onset of following vowel have been good discriminators (C.-Y. Lee et al., 2014; S.-I. Lee, 2011; S. Li & Gu, 2015). Manner differences, such as state of aspiration in the affricates has been shown to be distinguished well by duration and amplitude (S. Li & Gu, 2015). For this study, we extracted the following five acoustic parameters using Praat: normalized duration of frication and the four spectral moments: center of gravity (COG), dispersion, skewness and kurtosis. ProsodyPro (Xu, 2013) was used to generate the actual durations of segment proportions, and following S. Li & Gu (2015) we calculated the normalized durations (i.e. the ratio of the consonant duration to the duration of the entire syllable) to avoid the influence of speaking rate. A script (Mayer, 2011) was used for obtaining measures for the four spectral moments calculated over the middle 40 ms of the frication portion.

### 3. Results

A set of linear regression models were used to analyze the data (i.e. normalized duration, four spectral moments) as the dependent variable and interaction between place of articulation and manner as independent

variables using R (R Core Team, 2018). In order to see which groups differed from each other, ‘esmeans’ (Lenth, Singmann, Love, Buerkner, & Herve, 2018) was used for pairwise comparison.

### 3.1. Normalized duration of frication

Table 2 lists the values for the nine target consonants, the entire target syllable as well the duration proportion of the consonant relative to the syllable. All values are averaged across speakers and repetitions. The Danish alveolo-palatals are termed “Palatal” for short in the following presentations of data. Data is arranged by the three sets of sibilant (Danish, Mandarin palatals and Mandarin retroflexes, i.e. a presumed place contrast), and by manner (fricative = blue, unaspirated affricate = green, aspirated affricate = red).

Place	Onset consonant	Consonant duration (ms)	Syllable duration (ms)	Normalized duration*
Danish Palatal	/dj/	57.01	347.20	0.16
	/tj/	130.96	390.17	0.34
	/sj/	181.30	420.50	0.43
Mandarin Palatal	/tɕ/	72.26	289.25	0.25
	/tɕ <sup>h</sup> /	130.57	344.22	0.38
	/ç/	153.04	432.17	0.35
Mandarin Retroflex	/tʂ/	50.57	274.60	0.18
	/tʂ <sup>h</sup> /	117.85	324.43	0.36
	/ʂ/	153.38	400.41	0.38

Table 2. Consonant duration, syllable duration and normalized duration. \*Normalized duration is calculated as a mean of all the individual token’s normalized duration rather than as a ratio of the two means given in the previous columns in the chart.

Table 2 and Figure 1 show that the manner of articulation (i.e. aspirated affricates, unaspirated affricates and fricatives) played an important role in the duration of the target consonants, such that fricatives were the longest followed by aspirated affricates and unaspirated affricates. The role of the



aspiration was in line with measurements presented by S. Li & Gu (2015), who found the same hierarchy of frication duration between the three manner distinctions. As Figure 1 shows, the Danish short lag affricate /dj/ did not differ from /tʃ/ in terms of normalized duration ( $t=-1.061, p=0.9793$ ), but /dj/ was shorter than /tʃ/ ( $t=4.104, p=0.0016$ ). Duration of the Danish long lag (and aspirated) affricate /tj/ was similar to /tʃ<sup>h</sup>/ and /tʃ<sup>h</sup>/ (/tj/ vs. /tʃ<sup>h</sup>/:  $t=-1.317, p=0.9260$ ; /tj/ vs. /tʃ<sup>h</sup>/:  $t=-2.570, p=0.2022$ ). Duration of /sj/ was similar to both /ʃ/ and /ʃ/ (/sj/ vs. /ʃ/:  $t=-0.720, p=0.9985$ ; /sj/ vs. /ʃ/:  $t=-0.648, p=0.9993$ ). Duration proportions indicate that the Danish fricative and aspirated affricate are similar to both of their Mandarin counterparts while Danish unaspirated /dj/ is produced most similar to retroflex /tʃ/, and is clearly produced with a shorter frication proportion than Mandarin /tʃ/.

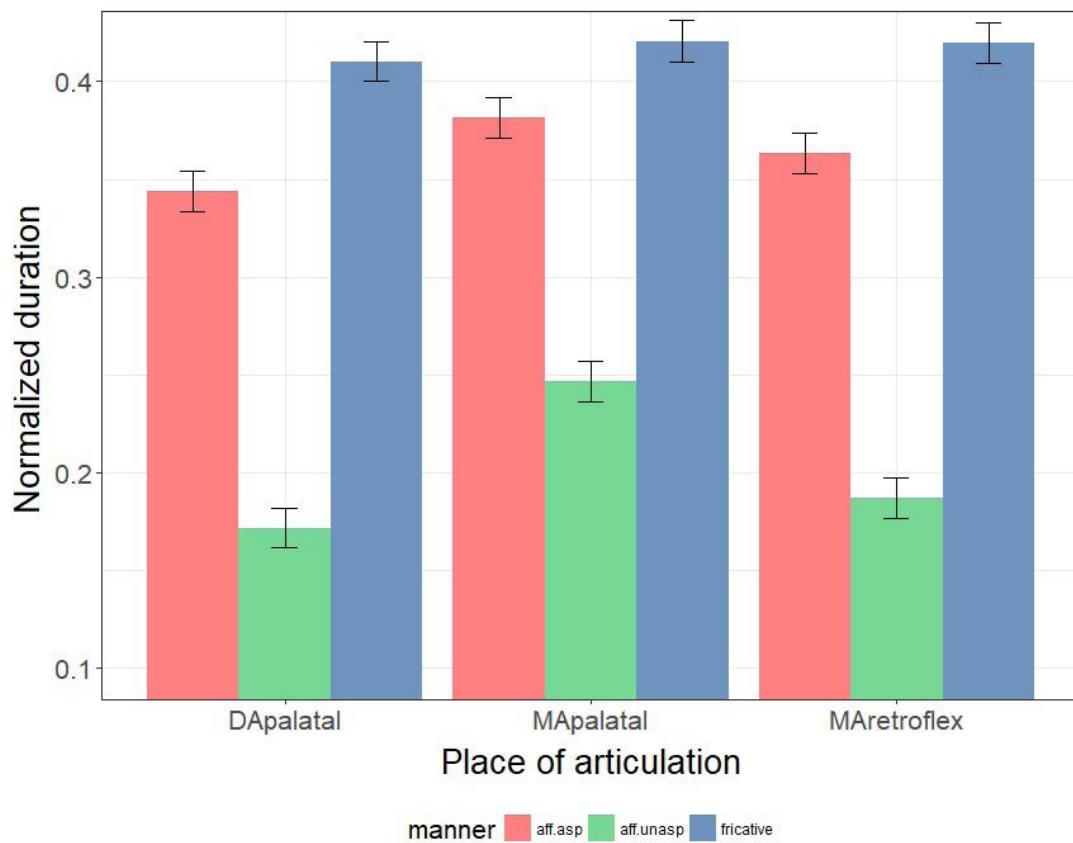


Figure 1. Normalized durations (proportional duration of C relative to the entire syllable). Error bars indicate standard errors.

### 3.2. Spectral moments

Table 3 displays values for four spectral moments averaged across speakers and repetitions.

Place	Onset consonants	COG (Hz)	Dispersion (Hz)	Skewness	Kurtosis
Danish Palatal	/dj/	5638	2042	2.28	8.56
	/tj/	5714	2123	2.08	6.53
	/sj/	5571	2112	2.23	7.90
Mandarin Palatal	/tɕ/	8284	1864	1.16	3.65
	/tɕ <sup>h</sup> /	7714	2006	1.30	3.17
	/ɕ/	8103	1916	1.30	3.37
Mandarin Retroflex	/tʂ/	5737	2074	1.27	2.70
	/tʂ <sup>h</sup> /	5496	2368	1.23	2.39
	/ʂ/	5652	2278	1.26	2.35

Table 3. Mean spectral moments.

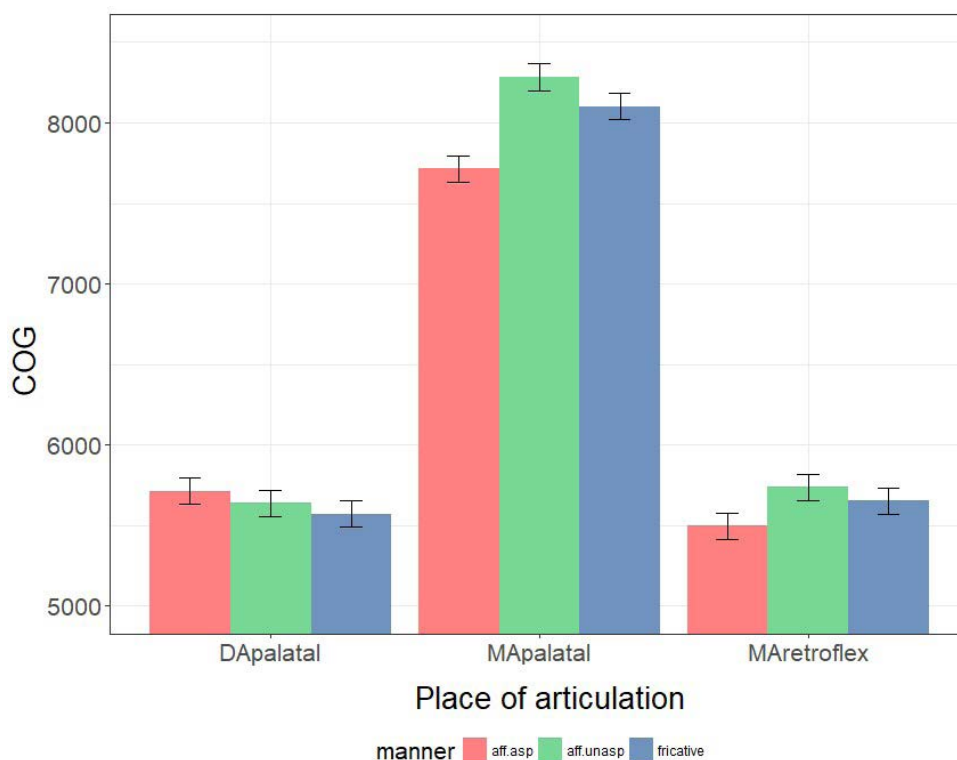


Figure 2. Center of gravity.

As shown in Table 3 and Figure 2 above, place of articulation was the main effect on COG. Post-hoc pairwise comparisons revealed that for Mandarin onsets, palatals had higher COG than retroflexes ( $p < 0.05$ ), in line with findings from previous studies (e.g. S.-I. Lee, 2011) and consistent with the fact that higher COG values are found for fricatives produced anteriorly (e.g. Jongman, Wayland, & Wong, 2000). The Danish alveolo-palatals were not significantly different from Mandarin retroflexes ( $p > 0.05$ ) which is in line with what Mikkelsen (2016) found in her cross-linguistic comparison of fricatives. In terms of the second spectral moment, dispersion, Danish palatals were not significantly different from their corresponding Mandarin palatals ( $p > 0.05$ ). However, Danish aspirated palatal /tj/ had a smaller standard deviation compared to the Mandarin palatal /tɕ<sup>h</sup>/ ( $t = -3.515$ ,  $p = 0.05$ ) resulting from less variability among the Danish aspirated affricate tokens than their Mandarin palatal counterparts. Danish fricative /sj/ had similar standard deviation as Mandarin /ç/ ( $t = 2.877$ ,  $p = 0.0972$ ). Regarding the third and the fourth spectral moments, Danish sibilants differed significantly from all Mandarin onsets ( $p < 0.05$ ). Svantesson (1986) plots the two first spectral moments and displays how the fricative tokens from his four male speakers form clusters that are not clearly separate from one another. We similarly graph dispersion as a function of COG in Figure 3 and Figure 4 below to visually represent the overlaps in acoustic space for three sibilants *within* each series, as well as *between* the Danish “palatal” and the Mandarin retroflex series (compare Figures 3 and 4).

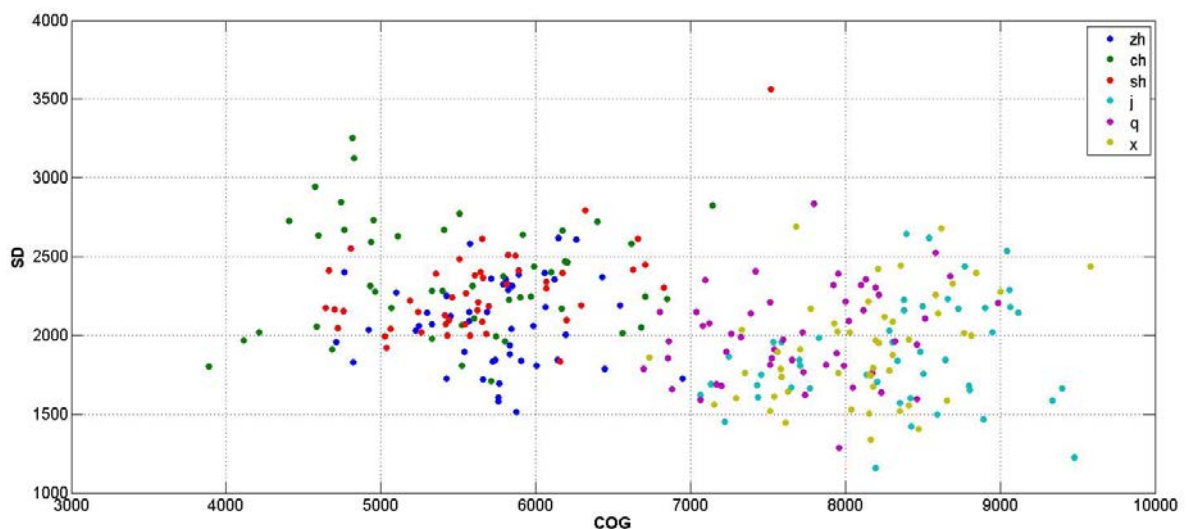


Figure 3. Mandarin sibilants. Items are labelled in pinyin in the legend: The retroflex series /tʂ, tʂ<sup>h</sup>, ʂ/ corresponds to <zh, ch, sh> and the palatal series /tɕ, tɕ<sup>h</sup>, ç/ corresponds to <j, q, x>.

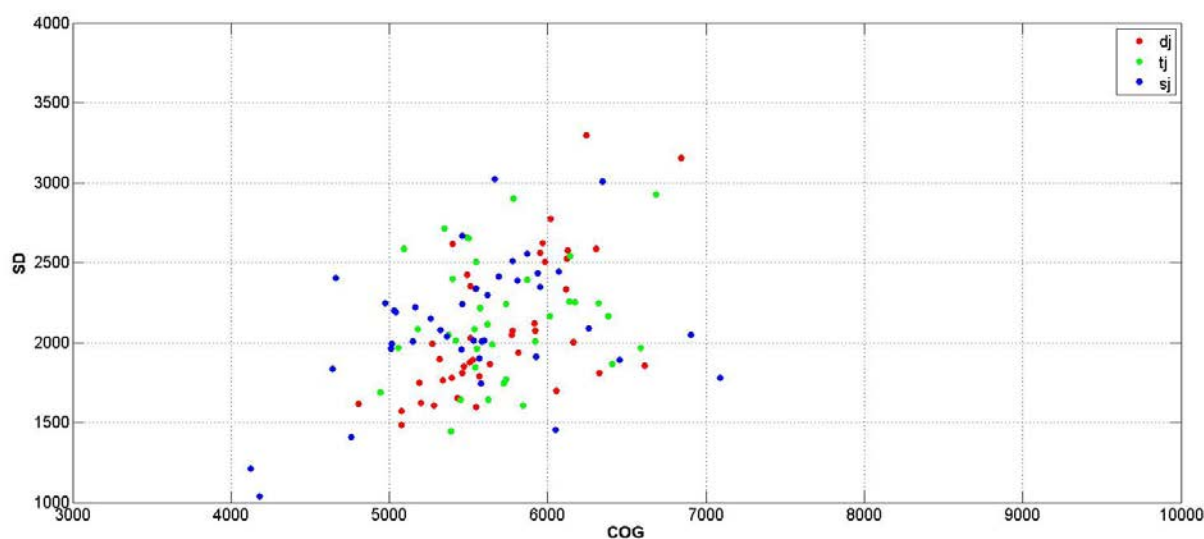


Figure 4. Danish sibilants.

#### 4. Discussion

This study addressed some of the questions arisen from a cross-linguistic assimilation study, which investigated the mapping of Mandarin consonants by naïve Danish listeners. The previous study indicated that, preceding low vowels, two sets of Mandarin coronal sibilants /tʃ, tʃʰ, ʃ/ and /tʂ, tʂʰ, ʂ/ were assimilated to only one set of Danish sibilants /dj, tj, sj/, with Mandarin retroflexes typically faring slightly better than the palatals in terms of assimilation percentage. This modest preference for retroflexes as matches of the native sibilants was surprising considering the respective articulatory descriptions of categories, from which a closer similarity with the Mandarin so-called “palatal” series would be expected. Our results from the current study, however, reveal that the Danish sibilants are acoustically more similar to the Mandarin retroflexes than the palatals, most clearly shown in the comparison of spectral moments. COG values for the series of Danish sibilants were significantly different from values obtained for the Mandarin palatal consonants, while measures for the Danish onsets and the Mandarin retroflexes were not significantly different. Since Danish only has one series of comparable postalveolar sibilants, the two-to-one mapping is nevertheless quite expected. The acoustic data presented here would suggest that the assimilation of Mandarin postalveolars preceding /a/ should yield a Category-Goodness distinction, with the retroflex series now identified as the most similar in terms of the acoustic signal. As COG measures do not reliably distinguish Danish sibilant onsets from Mandarin retroflexes in the production of two groups of native speakers, Mikkelsen’s

(2016) production data from Danish learners of Mandarin does not necessarily indicate the merging of a non-native and a native category.

The second question examined in this study concerned the unexpected assimilation of Chinese unaspirated affricates to the Danish aspirated /tj/. While /dj/ was still the preferred response, the native aspirated affricate /tj/ was sometimes selected as the closest native match to both /tʂ/ (22%) and /tʃ/ (35%). Since duration is one of the cues known to distinguish a [+/-aspiration] contrast we compared the normalized durations of the three sets of unaspirated and aspirated affricates. Our results show that the acoustic properties of Danish /dj/ differ significantly from all Mandarin affricates except /tʂ/, and /tj/ as significantly different from both /tʂ/ and /tʃ/. The normalized durations therefore do not offer any direct explanation for the unexpected mapping, and additional measurements might be needed to be able to fully account for the unexpected perception results. We speculate, however, if the duration differences between the three lenis onsets might hint at an explanation: The fact that the Mandarin affricate /tʂ/ is minimally longer and /tʃ/ is significantly longer than /dj/ might mean that duration for both of the Mandarin unaspirated affricates actually exceeds the perceptual boundary for the Danish /dj/ category, causing native listeners to classify them as acceptable variants of their native /tj/ instead.

Our data also brings into question the articulatory descriptions of these Danish consonant clusters, indicating a more retracted point of constriction in the oral cavity than what is typically assumed. It is, however, still possible that a combination of Danish sibilants + high vowel would yield a more fronted articulation of the consonant clusters, thereby increasing the spectral mean, resulting in an acoustic signal resembling of that for Mandarin palatals. Future studies should investigate how different sibilant + vowel combinations might account for differences in the spectral cues that discriminate sibilant categories. The discrepancy between articulatory descriptions and this newly obtained acoustic data of Danish sibilants is also well worth exploring in further detail.

### **Acknowledgements**

We are grateful to our ten speakers and to Jonas Villumsen for obtaining the Danish recordings while neither of us were in Denmark and to Allard Jongman for his helpful comments on an earlier version of the paper. We are likewise very grateful to the reviewer for helpful comments. Any errors are our own. And finally, we are incredibly grateful to Ocke for encouraging us to start working in the field of phonetics, for being an inspiring teacher and a supportive advisor.

## References

- Ao, B. (1992). Non-uniqueness condition and the segmentation of the Chinese syllable. In E. Hume (Ed.), *Ohio State University: Working papers in linguistics* (pp. 1-15). Columbus: Ohio State University.
- Basbøll, H. (2005). *The phonology of Danish*. Oxford / New York: Oxford University Press.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Baltimore, MD: York Press.
- Boersma, P., & David Weenink. (2018). *Praat: Doing phonetics by computer (Version 6.0.39)*. Retrieved from [www.praat.org](http://www.praat.org)
- Bohn, O.-S. (2002). On phonetic similarity. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 191-216). Trier, Germany: Wissenschaftlicher Verlag.
- Chang, C. B., Haynes, E. F., Yao, Y., & Rhodes, R. (2009). A Tale of Five Fricatives: Consonantal Contrast in Heritage Speakers of Mandarin. *University of Pennsylvania Working Papers in Linguistics*, 15(1), 36-43.
- Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed). Oxford / New York: Oxford University Press.
- Duanmu, S. (2017). From non-uniqueness to the best solution in phonemic analysis: Evidence from Chengdu Chinese. *Lingua Sinica*, 3(1). <https://doi.org/10.1186/s40655-017-0030-7>
- Grønnum, N. (2009). *Fonetik og fonologi: Almen og dansk* (3rd ed.). København: Akademisk.
- Hartman, L. M. (1944). The segmental phonemes of the Peiping dialect. *Language*, 20, 28-42.
- Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica*, 33(5), 353-367. <https://doi.org/10.1159/000259792>
- Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3), 1252. <https://doi.org/10.1121/1.1288413>
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, UK / Cambridge, Mass., USA: Blackwell Publishers.
- Lee, C.-Y., Zhang, Y., & Li, X. (2014). Acoustic characteristics of voiceless fricatives in Mandarin Chinese. *Journal of Chinese Linguistics*, 42(1), 150-171.
- Lee, S.-I. (2011). Spectral analysis of Mandarin Chinese sibilant fricatives. *18th ICPhS, Glasgow*, 1178-1181.
- Lee, W.-S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, 33(1), 109-112. <https://doi.org/10.1017/S0025100303001208>

- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Estimated Marginal Means (Version 1.2.3) [R]. Retrieved from <https://github.com/rvlenth/emmeans>
- Li, M., & Zhang, J. (2017). Perceptual distinctiveness between dental and palatal sibilants in different vowel contexts and its implications for phonological contrasts. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1), 1-27. <https://doi.org/10.5334/labphon.27>
- Li, S., & Gu, W. (2015). Acoustic analysis of Mandarin affricates. *Sixteenth Annual Conference of the International Speech Communication Association / INTERSPEECH 2015, Dresden*.
- Mayer, J. (2011). *Calculate spectral moments*. Retrieved from [http://praatpfanne.lingphon.net/downloads/spectral\\_moments.txt](http://praatpfanne.lingphon.net/downloads/spectral_moments.txt)
- Mikkelsen, S. (2016). *Chanish? A study of a Danish accent in the production of two Chinese sibilants* (Unpublished master's thesis). Aarhus University, Aarhus.
- Norman, J. (1988). Chinese. Cambridge / New York: Cambridge University Press.
- R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org>
- Rasmussen, S., & Bohn, O.-S. (2017). Perceptual assimilation of Mandarin Chinese consonants by native Danish listeners. *The Journal of the Acoustical Society of America*, 141(5), 3518-3518. <https://doi.org/10.1121/1.4987399>
- Svantesson, J.-O. (1986). Acoustic analysis of Chinese fricatives and affricates. *Working Papers in Linguistics*, 25(1), 195-211.
- Xu, Y. (2013). ProsodyPro – A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody*. Aix-en-Provence, France.
- Zhang, J., Lü, S., & Qi, S. (1982). A cluster analysis of the perceptual features of Chinese speech sounds. *Journal of Chinese Linguistics*, 10(2), 190-206.

## Normalization of the Natural Referent Vowels

D. H. Whalen

City University of New York; Haskins Laboratories; Yale University

### Abstract

The Natural Referent Vowel framework makes strong, testable predictions that have already provided fruitful directions for new research. Largely undiscussed, however, is the role that vocal tract normalization plays in the perception of vowels by infants. Two issues arise: First, when presented with a single vowel, how does the infant know whether it is truly a referent vowel or not? Second, if, unlike in all previous studies, vowels attributable to different vocal tracts are perceived, does the infant normalize or not? The first might be answerable with neural imaging. The second can be tested behaviorally, though the design is difficult both mechanically and theoretically.

### 1. Introduction

The Natural Referent Vowel framework (Polka & Bohn, 2011) treats the articulatorily and acoustically extreme vowels as natural reference points that are especially useful to infants learning language. These NRVs are /i a u/. Being on the edge of the vowel space, they can give an infant anchor points for developing a vowel space of their own. Infants can more easily tell that a vowel has changed when the change is toward the periphery (i.e., toward the NRVs) than in the opposite direction. A variety of experimental results are consistent with NRVs being treated differently from other vowels, as summarized in the 2011 paper.

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 81-89). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.



A previous framework, the Native Language Magnet (NLM) effect (e.g., Kuhl & Iverson, 1995), makes some compatible predictions while leaving open the issue of the universality of the NRVs. NLM predicts that vowels that occur in the ambient language will attract nearby vowel tokens into their perceptual category, while vowel categories from other languages will not. Given that most languages have the NRVs in their inventories, this will lead to compatible predictions between the two accounts in many cases, though Polka and Bohn have results that do not depend on the NRV's existence in the language (e.g., Polka & Bohn, 1996). (It would be interesting to see what happens with languages, such as most of the Algonquian languages, which lack /u/.) There is some evidence that the NRVs have a greater perceptual effect than other native vowels (Polka & Bohn, 2011: 476), with /i/ eliciting more reaction (sucking) than /y/ even for infants in a Swedish environment where both vowels are native. As with most issues concerning acquisition, there is much more work to be done.

Vowel identification is not straightforward for listeners, however. Different vocal tract lengths produce different formant patterns for the same vowel. This is clear both on theoretical grounds (Fant, 1960) and in measurements of men, women and children (Peterson & Barney, 1952). Human listeners compensate for such effects, and they seem to do so both with signal-internal ("intrinsic") and ancillary ("extrinsic") information (Ainsworth, 1975). Intrinsic information is entirely within a single vowel. Extrinsic information relates the token to a speaker's vowel space, or at least the immediate environment. That infants are capable of such normalization is indicated by their success at imitating adult productions with their tiny little vocal tracts, even though their formant values were necessarily different from the adult models (Kuhl & Meltzoff, 1996). Indeed, imitation based on reinterpretation of the input signal, including sensitivity to its visual aspects, into something the infant can produce is a prerequisite for speech acquisition (Studdert-Kennedy, 1986).

Many acoustic normalization procedures have been proposed (for a review, see, e.g., Flynn, 2011). To date, they all perform more poorly than human listeners. Humans, of course, have an advantage in having a couple of million years of evolution helping them out, but the algorithms are also hampered by limitations on the input given to them. Typically, the input includes fundamental frequency (F0) and formant values for the midpoint of a vowel, augmented in some cases by duration information. We have known for decades that this is not the information that human listeners depend most on (e.g., Strange et al., 1976), but the levels of performance

obtained are sufficient that the approach continues to be used. However, our formant measurements are not terribly accurate (Klatt, 1986; Shadle et al., 2016), leading to an initial degradation of performance by the normalization algorithms. For one sizable dataset, a model that included F0 and formant measurements at one time point still performed well below human perceptual levels, while using three time points along with duration led to fairly equivalent performances to those of humans (Hillenbrand et al., 1995). It would seem that infants have their work cut out for them.

The research discussed in the previous paragraph included all the vowels of English, but the NRVs are not always the best identified ones. In the Hillenbrand et al. (1995, p. 3108) study, the vowel /i/ was identified correctly by human listeners the most often (99.6%). The vowel /u/ was also highly identifiable (97.2%), but not as much as /o/ (99.2%). The vowel /ɑ/ was noticeably less accurately identified (92.3%). If we ignore the vowel /ɔ/, which is not distinctive in all American English dialects, the next worst rate of identification was 90.8% for /ʌ/. These are, of course, adult perceptions of an established inventory, and they are identification scores, which may be less revealing than discrimination scores. However, they do not immediately indicate that NRVs have a special status.

This paper will explore two predictions that can be drawn from the NRV position. First, experimental studies that present individual tokens of vowels to infants would seem to require that they recognize each token as being an example of an NRV or not. Is this possible? How can we tell? Second, does the need to normalize for more than one vocal tract reduce the effectiveness of the NRVs? Adult listeners show reduced accuracy with multiple speakers (e.g., Assmann et al., 1982); do infants also have trouble adjusting their categories? Possible ways of addressing those questions will be presented.

## **2. Normalization of a single vowel**

Each time a stimulus is presented to a listener (in the current case, an infant), some phonetic information is extracted. Just what kind of information that is remains underspecified. Is it a true representation of the resonances of the vocal tract that produced it? Is it classified into the NRV category it belongs to? Or is it placed into some vaguely specified acoustic space that is only tangentially related to vowel categories? The ultimate perceptual treatment of ambiguous vowels (e.g., Kuhl et al., 1992) is then somewhat unclear: Are these vowels also normalized on intrinsic grounds, or are they (extrinsically) put into the context of the current speaker?

The nature of the low NRV is itself rather ambiguous. There is a great deal of variation in the low vowel used by any specific language, even though some form of low vowel is, perhaps, universal. Should we expect that the NRV status of [ɑ] (or [ɒ] or [ɐ] or [a]) can be determined on a language-specific basis? Or does that violate the principles of the NRV proposal? Certainly, the lowest vowel that an infant hears from a particular speaker might serve as a reference point, but this implies that extrinsic normalization is at work (which would seem to reduce the “referent” component of NRV) and that the infant can associate utterances consistently with a single speaker. It does seem that infants can recognize new voices, at least those that are speaking the ambient language (Johnson et al., 2011), but such recognition might, of course, depend on source characteristics rather than filter characteristics. Would an infant be startled to hear a newly-familiar voice produce a vowel that seemed outside its range? Or would that signal a new speaker? Our current experimental techniques may be inadequate for answering the question directly, but considerations of how we might approach the question allow further insight into the nature of the NRVs.

The statistical distributions of vowel formants might also influence infants’ perception in these experiments, but many of the same issues arise. Are the instances of a formant pattern mapped onto an individual speakers’ vowel space? Can the infant keep multiple maps and update them appropriately? Indeed, how do they know to put them into a vowel space at all, rather than just interesting formant patterns? Their babbling (at one year of age) does reflect the ambient language in the trends of formant values (Boysson-Bardies et al., 1989), an effect those authors attribute to the onset of category formation. Statistical explanations were proposed to avoid nativist explanations of acquisition, by allowing the language patterns themselves to provide the information needed for acquisition. Infants have been shown to be sensitive to a number of statistical properties in speech experiments (Kuhl et al., 1992; Maye et al., 2002; Saffran et al., 1996), although it is less clear that short-term sensitivity predicts long-term retention (Gómez, 2017). Just where and how those statistics are stored is also unspecified. Vowel formants would seem to have to be normalized into some universal space or stored along with speaker identity. Further, if there are no categories, what are the statistics applying to? The earliest stages of acquisition would seem to resist an entirely input-based approach. Beyond that, the way in which speaker-specific storage would then affect the infant’s own production is not obvious.

The nativist positions that were challenged by statistical approaches were largely reacting to theories that assumed some kind of segment as innate, but non-segmental nativist proposals may also be viable. The Articulatory Organ Hypothesis (the AOH; Studdert-Kennedy & Goldstein, 2003) assumes that certain broad gestures (tongue tip, lips, etc.), “organs,” are available to infant perceivers, so that distinctions across organs are more easily perceived than those within. The evidence for the AOH has been primarily examined for consonants, and the results have been largely positive with many problematic cases (Best et al., 2016). For vowels, constrictions of the pharynx, tongue root and tongue tip could provide organs for /a/, /u/ and /i/ respectively. This is, of course, entirely consistent with the NRVs, but seen primarily from the articulatory viewpoint. The essential breadth of the organs in the AOH also allows for the wide variety of low vowels mentioned above to be included in one organ category without removing /a/ (and its neighbors) from primary status. Any success the infant has in recognizing organs in adult speech is as dependent on normalization as in other frameworks, but it may be that this kind of global gesture is more easily computed from the acoustic signal than previously thought, once a fuller (and more realistic) depiction of the acoustic signal is used (Iskarous, 2010).

Whichever framework ultimately provides the best explanation for speech acquisition, it is clear that a great deal of work remains before we will have a satisfactory understanding. If we find a neuroimaging technique that allows us to see a distinct signature for a specific vowel category, perhaps we will be able to see when and where the normalization takes place, how successful it is, and what happens to individual tokens in an experiment. To date, we do not have such signatures, but they may yet be discoverable. If so, exploring the development of categorization will become even more fascinating.

### **3. Multiple vocal tracts in one experiment**

The procedure for testing infant perception of vowels typically involves a single speaker (e.g., Polka & Bohn, 1996) or one speaker per language (e.g., Polka & Werker, 1994). Although this makes the experimental design manageable – infants do not tolerate a huge number of stimuli – it does ensure that the vowels will be maximally informative about a single vocal tract. The imitation studies cited above suggest that infants do, in fact,

perform some kind of normalization, so that they can relate what they hear to what they would produce. How well does this normalization proceed on a token to token basis?

Two possibilities seem likely. The first is that infants normalize primarily on intrinsic grounds, and thus they should be able to identify tokens from different speakers as belonging to the same category. The other, completely opposite possibility is that infants rely greatly on extrinsic grounds (sampling of the total vowel space, for example) and would fail to identify any vowels, including the NRVs, when speakers vary. There are probably other intermediate possibilities, for example, that there is something strongly coherent in the NRV acoustic pattern that is treated as a “magnet” on psychophysical grounds, such as a close proximity of two intense formants (F1 and F2 for [u] and [ɑ], F2 and F3 for [i]). Such an account would be consistent (I think) with the intrinsic normalization variant. In any event, these two starkly contrasting ones suggest a direct test.

We could test these two different predictions by seeing whether having multiple speakers eliminates or reduces the attraction that NRVs have in infant perception. (Multiple talkers were used in Bundgaard-Nielsen et al. (2015), showing that the technique is possible.) In the extreme case, every token presented to an infant could come from a different talker. But even having 10 or 20 talkers would presumably be sufficient to test whether speaker consistency in the input is strictly necessary. In a fantasy world, where infants scheduled themselves in large numbers, we could then imagine doing each of those talkers separately to ensure that the different voices were equally effective. Even in our present world, we could conceivably test 4, or at the very least 2 of the speakers to see if the NRVs had a larger effect with a single speaker than they did with multiple speakers. However, the main issue would not require such an extension: If NRVs are effective with many talkers, it would seem that intrinsic normalization was operative. If they were not effective, then it could be that extrinsic normalization is necessary, but it could also be that infants are not happy with a situation with too many adults and/or a lack of social connection with one or two adults. No doubt there are other possible outcomes and explanations, but in our current state, we don't know how NRVs are normalized.

#### **4. Summary**

The NRV proposal is one that makes an admirable number of predictions possible. Because so much of what happens in speech perception, especially at the very beginning of a speaker/listener's life, is currently unknown,

these predictions must be rather broad. Further, the extensive work that is involved in any study of infant perception guarantees that progress will be slow. Ocke Bohn, and his many collaborators, are to be commended for persevering with this and other questions fundamental to our understanding of speech. While he may not reach the ultimate answers before hanging up his headphones, we can hope that Ocke will continue to explore this endlessly fascinating topic.

## 5. References

- Ainsworth, W. A. (1975). Intrinsic and extrinsic factors in vowel judgements. In G. Fant, & M. A. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 103-113). San Francisco: Academic Press.
- Assmann, P. F., Nearey, T. M., & Hogan, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America*, *71*, 975-989.
- Best, C. T., Goldstein, L. M., Nam, H., & Tyler, M. D. (2016). Articulating what infants attune to in native speech. *Ecological Psychology*, *28*, 216-261.
- Boysson-Bardies, B. de, Hallé, P. A., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, *16*, 1-17.
- Bundgaard-Nielsen, R. L., Baker, B. J., Kroos, C. H., Harvey, M., & Best, C. T. (2015). Discrimination of multiple coronal stop contrasts in Wubuy (Australia): A Natural Referent Consonant account. *PLoS ONE*, *10*(12), e0142054.
- Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.
- Flynn, N. (2011). Comparing vowel formant normalisation procedures. *York Papers in Linguistics Series*, *2*, 1-28.
- Gómez, R. L. (2017). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*.
- Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*, 3099-3111.
- Iskarous, K. (2010). Vowel constrictions are recoverable from formants. *Journal of Phonetics*, *38*, 375-387.
- Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, *14*, 1002-1011.
- Klatt, D. H. (1986). Representation of the first formant in speech recognition and in models of the auditory periphery. In P. Mermelstein (Ed.) *Proceedings of the Montreal satellite symposium on speech recognition, 12th international congress on acoustics* (pp. 5-7). Montreal: Canadian Acoustical Society.

- Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the “perceptual magnet effect”. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121-154). Timonium, MD: York Press.
- Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, *100*, 2425-2438.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. E. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606-608.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101-B111.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175-184.
- Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*, 467-478.
- Polka, L., & Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America*, *100*, 577-592.
- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 421-435.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926-1928.
- Shadle, C. H., Nam, H., & Whalen, D. H. (2016). Comparing measurement errors for formants in synthetic and natural vowels. *Journal of the Acoustical Society of America*, *139*, 713-727.
- Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America*, *60*, 213-224.
- Studdert-Kennedy, M. (1986). In B. Lindblom, & R. Zetterstrom (Eds.), *Development of the speech perceptuomotor system. Precursors of early speech* (pp. 205-217). New York: M. Stockton Press.
- Studdert-Kennedy, M., & Goldstein, L. M. (2003). In M. Christiansen, & S. Kirby (Eds.), *Launching language: The gestural origin of discrete infinity. Language evolution* (pp. 235-254). Oxford: Oxford University Press.

# **PERCEPTION OF ACCENT**

Handling editor: Mette Hjortshøj Sørensen





# Assessing the Effect of Perceptual Training on L2 Vowel Identification, Generalization and Long-term Effects

Angélica Carlet  
Universitat Internacional de Catalunya

Juli Cebrian  
Universitat Autònoma de Barcelona

## Abstract

This paper assessed two high variability phonetic training methods aimed at improving the perception and production of English vowels by Spanish/Catalan speakers. Fifty-four L2 learners of English were assigned to one of three groups: forced-choice identification (ID) training, AX categorical discrimination (DIS) training, and control group (CG). Participants' identification and production of English vowels was assessed before training, after training and two months later. Both trained groups outperformed the CG at posttest and showed evidence of generalization and retention of learning. However, the ID trainees showed greater improvement in perception and significant gain in production, pointing to a potential superiority of this method for vowel learning. These results have implications for future research on phonetic training and practical applications for the teaching of pronunciation.

## 1. Introduction

The acquisition of target second language (L2) sounds can be challenging for the L2 learner due to the interplay of many factors including onset age of learning, length of residence in the target-language country, amount of L2 exposure, amount of L1 and L2 use, learner motivation and aptitude, and linguistic factors like typological relatedness or the role of orthography

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 91-119). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.

(Piske, MacKay & Flege, 2001; Bohn & Munro, 2007). This difficulty is clearly related to the effect of existing L1 phonetic categories<sup>1</sup> and the L2 learners' failure to perceive target L2 sounds accurately, as proposed by L2 speech models such as Flege's (1995a, 2003) Speech Learning Model (SLM), Kuhl and Iverson's (1995) Native Language Model (NLM), and Best and Tyler's (2007) Perceptual Assimilation Model (PAM-L2), among others. According to these models, given enough input and experience, learners may succeed in establishing long-term memory representations for target L2 sounds, separate from pre-existing L1 categories.

The present study is set in an instructional context, that is, learning English as a foreign language in the learners' home country. This setting is characterized by limited exposure to the target language outside the classroom (Muñoz, 2008; Saito, 2015). This scenario may be problematic for accurate second language learning, since extensive exposure to the target language is crucial to develop the ability to distinguish native from non-native sounds (Flege, 1991; Ingram & Park, 1997), a pre-requisite for accurate L2 category formation (e.g., Flege, Bohn & Jang, 1997; Flege, 1995a). Against this background, a possible source of specialized target language input can be found in phonetic training, which aims at directing L2 learners' attention to challenging features or contrasts present in the target language by means of specialized perceptual or pronunciation tasks that generally include corrective feedback (Cebrian & Carlet, 2014). There is evidence that a short training regime may have the same outcome as a prolonged period of instruction, and that training is effective even for learners at different levels of proficiency. Pereira (2014) reported that a group of Chilean learners of English who completed a six-week perceptual training regime were able to improve their perception of English vowels to a similar extent as another group of Chilean learners who had undergone three years of formal instruction. Iverson, Pinet and Evans (2012) explored whether training was equally effective for different settings and levels of proficiency. Beginner and intermediate French learners underwent a vowel identification training regime and were tested on the identification, discrimination and production of 14 English vowels and diphthongs. Both groups showed a slight effect of training on discrimination ability, as well as significantly improved their identification and production as a result of training.

---

<sup>1</sup> Phonetic categories are defined as "the distribution of acoustic tokens which together are perceived as mapping to a phoneme in the listener's inventory" (Earle & Myers, 2014, p. 1192).

There is thus evidence from a considerable amount of studies that phonetic training can be beneficial for different L1-L2 language combinations and different target structures, particularly L2 consonants and vowels (Cebrian & Carlet, 2014; Iverson & Evans, 2007, 2009; Lacabex, García-Lecumberri & Cooke, 2008; Lengeris, 2008; Nishi & Kewley-Port, 2007; Nobre-Oliveira, 2007; Rato, 2014; Thomson, 2012). A number of laboratory training studies have adopted successfully what is known as a high variability phonetic training approach (HPVT), which incorporates multiple stimuli involving a variety of speakers, tokens, phonetic contexts, etc., in an attempt to replicate the variability that characterizes L2 input in a natural environment (Logan, Lively & Pisoni, 1991; Lively, Pisoni & Logan, 1993; see section 1.1). It has been argued that training is truly effective if its effect goes beyond improvement on the trained structures from pretest to posttest, that is, if improvement generalizes to untrained stimuli such as new voices, new items or new modalities (Logan & Pruitt, 1995; Flege, 1995b; Bradlow, 2008). In addition, the efficacy of phonetic training is demonstrated further when the observed improvement is still found some time after training has ended, that is, if learning is retained beyond the training period. According to Logan and Pruitt (1995), generalization and retention provides evidence that robust learning has occurred. This study examines the effect of high variability perceptual training on L2 vowel perception and production and compares the effectiveness of two types of perceptual tasks, identification and discrimination, on the ability to identify and produce L2 sounds. In addition, the study also compares the two perceptual methods on the extent to which the potential improvement generalizes to untrained structures, and is retained after a two-month interval.

### **1.1. Perceptual training studies on vowels**

The learnability of vowels through laboratory training has been investigated extensively in the last few decades (Aliaga-García & Mora, 2009; Cebrian & Carlet, 2014; Iverson & Evans, 2007, 2009; Lacabex et al., 2008; Lambacher, Martens, Kakehi, Marasinghe & Molholt, 2005; Lengeris, 2008; Nishi & Kewley-Port, 2007; Nobre-Oliveira, 2007; Rato, 2014; Rato & Rauber, 2015; Thomson, 2012; Wang & Munro, 2004; among others). For instance, in an HVPT study, 26 Mandarin Chinese speakers were trained to improve the perception of 10 English vowels produced in a post labial stop context (Thomson, 2012). After eight short identification training sessions, learners' ability to identify the English vowels improved

significantly and also generalized to a velar stop context. Moreover, the improvement obtained after training was retained one month after training completion. In fact, several studies have also reported successful retention of learning after periods of time ranging from one to twelve months (Rato, 2014; Wang & Munro, 2004; Nishi & Kewley-Port, 2007), which confirms the robustness of the training procedure and the relevance of phonetic training as an L2 learning tool (Logan & Pruitt, 1995).

Aliaga-García and Mora (2009) investigated the effect of HVPT in a study involving Spanish/Catalan learners of English and found a positive effect of HVPT on the identification and, to a lesser extent, production of English initial stops. Training also improved vowel perception; however, no positive effect of training on vowel production was observed. In a later study, Aliaga-García, Mora and Cerviño-Povedano (2011) found that improvement in L2 vowel perception varied as a function of phonological short-term memory capacity. Further, in a short-term perceptual training study involving Spanish/Catalan speakers, Cebrian and Carlet (2014) assessed the effect of a three-week HVPT regime (four 45-minute sessions) consisting of a variety of perceptual tasks on the perception of two vowel pairs (/i:/-/ɪ/ and /æ/-/ʌ/, as well as two consonant contrasts) by advanced learners of English. They found a positive effect of training for a subset of the target vowels, namely /i:/ and /ʌ/, and partial generalization effects. Finally, Rato (2014) and Rato & Rauber (2015) reported both generalization and retention of learning after a training regime that combined identification and discrimination tasks. These studies, however, combined different perceptual training tasks in the same training regime, so it is not possible to evaluate what the relative contribution of the different tasks may have been. The present study tries to contrast and evaluate the effectiveness of each type of task.

## **1.2 Perceptual training tasks**

Perceptual training studies often make use of discrimination or identification tasks. Even though early findings with stop consonants revealed the efficacy of discrimination (DIS) tasks in modifying learners' categorical perception of these sounds (Pisoni, Aslin, Perey & Hennessy, 1982; McClaskey, Pisoni & Carrell, 1983), the efficacy of identification (ID) training has been said to be superior to discrimination training as an L2 training tool (Jamieson & Morosan, 1986; Logan & Pruitt, 1995, among others). Strange and Dittmann (1984) found that Japanese learners of English improved their identification and discrimination of English /r/-

/l/ after undergoing auditory discrimination training involving synthetic stimuli. However, this improvement did not generalize to novel and natural stimuli. By contrast, identification tasks have been found to promote generalization of learning (Jamieson & Morosan, 1986; Logan et al., 1991). It is possible that DIS tasks promote within-category sensitivity and tap into lower levels of phonological encoding that are not greatly affected by language experience, while ID tasks may enhance between-category sensitivity and involve higher levels of phonological encoding more relevant for L2 categorization (Jamieson & Morosan, 1986; Logan & Pruitt, 1995; Iverson et al., 2012). Still it has been proposed that both ID and categorical DIS may affect similar levels of processing (Flege, 2003; Højen & Flege, 2006) and hence equally promote categorization of L2 sounds (Polka, 1992).

Few prior studies have compared the efficacy of ID and categorical DIS tasks incorporating highly variable stimuli in the same study (Flege, 1995b; Wayland & Li, 2008; Nozawa, 2015; Shinohara & Iverson, 2018). Flege (1995b) assessed the efficacy of both types of task (two-alternative forced-choice identification task and categorical same/different discrimination task) in a single HVPT study aimed at training Mandarin learners of English to identify final /d/ and /t/. Identification scores after seven training sessions showed that the two trained groups outperformed the controls at post-test and showed generalization of knowledge and long term effects. These findings pointed to the efficacy and robustness of both training methods and challenged the general claim that ID training is superior to discrimination training. Wayland and Li (2008) trained Chinese and English listeners to discriminate Thai tone contrasts by means of ID and DIS tasks. The findings revealed that both ID and DIS training procedures were similarly effective in enhancing listeners' discrimination of Thai tone contrasts and that the Chinese group outperformed the English group. Thus, the authors concluded that both methods were equally effective in improving tone perception and that the prior experience with a tone language explained the Chinese participants' advantage.

On the other hand, Nozawa (2015) compared the effect of ID and categorical ABX DIS training on Japanese learners' identification of English coda nasals and vowels in a small scale study involving two training sessions. While Nozawa found that both methods promoted significant gains regarding the final nasals, the ID method was found to be superior to ABX DIS for training vowels. A recent study compared the efficacy of the DIS and ID tasks further by evaluating their effect on the perception

and production of the /r-/l/ contrast by Japanese adult learners of English (Shinohara & Iverson, 2018). L2 learners were assessed on identification, auditory discrimination, category discrimination, and /r-l/ production at three times (pretest/midtest/post-test). Experimental groups were trained with both tasks in a different order. Their results after a 10 session regime showed that both training methods improved Japanese speakers' perception and production of English /r-l/ to a similar extent. In summary, more recent studies comparing ID and categorical DIS tasks have provided comparable results for both methods, particularly for training consonants. To our knowledge, only the study by Nozawa (2015) investigated vowels, showing a greater effect of ID in this case. This study explores the effects of these two methods further by contrasting their effect on L2 vowel perception and production. The main questions the present study aims to address are:

- Which type of training (ID or DIS) is more efficient in promoting improvement on the perception of L2 vowel sounds by Spanish/Catalan bilinguals?
- Which type of training (ID or DIS) is more efficient in promoting improvement on the production of L2 vowel sounds by Spanish/Catalan bilinguals?
- Which type of training (ID or DIS) is more efficient in promoting generalization and long-term effects?

Assuming that both training methods (ID and categorical DIS) tap into similar levels of processing (Flege, 2003; Højen & Flege, 2006) and also promote L2 categorization (Polka, 1992), it is hypothesized that both methods will be equally effective in improving learners' perception after training as well as promoting generalization of learning and retention effects, in accordance with Flege (1995b). Moreover, perceptual training with no focus on production may lead to production gains, even if to a lesser extent than the perceptual gains (Rato & Rauber, 2015; Rochet, 1995; Bradlow, 2008; Hardison, 2004; Iverson & Evans, 2009; Thomson, 2012; Pereira, 2014).

## **2. Methods**

### **2.1. Participants**

Fifty-four learners of English as an L2 took part in a 10-week-long regime and were assigned to one of three groups: 1) forced-choice identification

training (ID,  $N=20$ ), b) AX categorical discrimination training (DIS,  $N=18$ ), or c) control group with no perceptual training (CG,  $N=16$ ).<sup>2</sup> The L2 learners were Catalan/Spanish bilinguals, with a mean age of 19.7, and an initial age of EFL learning of 5.75 years. All subjects were second-year undergraduate students in English Studies at the *Universitat Autònoma de Barcelona (UAB)* enrolled in an introductory phonetics course. The learners' level of English ranged from a B2 to a C1 level on the *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (CEFR) (Council of Europe, 2001), with limited experience in an English-speaking country (average: two weeks) and no self-reported hearing impairments. Participants received course credit for their participation.

## 2.2. Target sounds and stimuli

The target sounds were the five standard Southern British English (SBE) vowels /i: ɪ æ ʌ ɜ:/, which are challenging for native speakers of Spanish/Catalan (Cebrian, 2006; Cebrian, Mora & Aliaga-García, 2011). The stimuli consisted of unmodified CVC nonsense words and real words elicited from ten native speakers of standard Southern British English (SBE) (five females and five males, mean age 27.8, range 23-39). The target vowels were always preceded and followed by obstruent consonants. The words were elicited by means of the carrier sentence: *I say "word", I say "word" now, I say "word" again*. In order to ensure the desired pronunciation of the nonsense words, the phrase *It rhymes with "real word"*, was added at the beginning (e.g., *It rhymes with give, I say "tiv" ....*). Recordings took place in a soundproof booth at the speech laboratory at University College London, UK, and each word was recorded three times. The recordings were carried out using *Cool edit 2000* software, a *Rode NT-1AX* microphone, *Edirol UA25* audio interface and were digitized at a 44.1 kHz sampling rate and 16 bit quantification.

## 2.3. Training stimuli

Training words consisted of nonsense words, so as to eliminate a potential word familiarity effect, given that the use of real words has been found to affect the accuracy and speed of word processing (Grosjean, 1980). These words were obtained from four of the SBE native speakers (two males and two females) with the objective to provide variability, as is characteristic

<sup>2</sup> Originally there were 20 participants in each group, e.g. at pretest, but a few learners did not complete all the training sessions and were discarded.



of HVPT. There were twelve words per target vowel (/i: ɪ æ ʌ ɜ:/), plus six words for two additional vowels (/e/ and /ɑ:/). The latter two were included to be contrasted with /ɜ:/ . Thus there were a total of 288 training stimuli (72 nonsense words x 4 talkers). The same stimuli were used in the identification and discrimination training tasks, as explained below. A list of the perceptual training stimuli can be seen in Appendix 1.

#### 2.4. Testing stimuli

Testing stimuli consisted of a subset of the non-words used at the training phase and involved 30 words (i.e. 5 target vowels x 6 words) of CVC nonsense words produced by 2 novel talkers (one male and one female), that is, different from training talkers, resulting in 60 testing stimuli. Since stimuli from these talkers were not used in the training corpus, testing already examined generalization to new talkers. In addition, 7 non-words were included as practice tokens in order to guarantee that the task procedure was understood and eight non-words involving the vowels /e/ and /ɑ:/ were included as testing fillers. Additionally, 20 CVC real word stimuli and 20 novel non-word stimuli produced by two familiar talkers (i.e. two of the four training talkers) tested generalization to real words and to novel untrained non-words, respectively (5 vowels x 2 words x 2 talkers).

#### 2.5. Procedure

Participants were assessed at three testing times (pre-test, post-test and delayed post-test) by means of the same perception and production tests. The perceptual tests consisted of two 7-alternative forced-choice vowel identification tasks (nonsense and real words) involving stimuli produced by different talkers from those used in the training phase. After training, generalization to new talkers and new words was also assessed by means of the same type of identification tasks. The response alternatives consisted of a phonetic symbol together with two common words representing each sound, specifically: /æ/ *ash, mass*; /ʌ/ *sun, thus*; /ɪ/ *fish, his*; /i:/ *cheese, leaf*; /ɜ:/ *earth, first*; /e/ *less, west*; /ɑ:/ *arm, palm*. Learners' L2 production was elicited by means of a picture naming task before and after training (pre-test and post-test). Participants were asked to name 27 different pictures and repeat the word twice. The 27 test words included the 10 real words containing the target vowel sounds examined between obstruent consonants.<sup>3</sup> A list of the production words can be seen in Appendix 1.

<sup>3</sup> This study is part of a larger scale study, which investigated the effect of HVPT on both consonants and vowel sounds.

Training for the experimental groups consisted of five 30-minute sessions over a 10 week-period and it was administered using TP software (Rauber, Rato, Kluge & Santos, 2011). An approximate study timeline is shown in Table 1. The DIS group was trained by means of AX discrimination tasks with immediate feedback. Participants responded by clicking on “same” or “different”. “Different” trials involved the two high-front vowels (/i:-I/), the two low vowels (/æ-Λ/) or the central vowel /ɜ:/ combined with either /e/ or /ɑ:/. Each pair was presented in the two possible orders in the same session (/æ-Λ/, /Λ-æ/), and in six different talker combinations over the course of the five sessions. There were 288 trials per training session. The ID group was trained by means of a 7-alternative forced-choice identification task with immediate feedback. The training tasks were specifically designed so as to ensure that both groups were exposed to the exact same set of stimuli through training. Thus, the ID tasks consisted of 576 trials per training session, involving the same stimuli presented in a discrimination session (that is, 288 trials involving a pair of stimuli each). Training for the control group was designed to provide the same amount of L2 instruction as the other groups without specific training. Thus, after the pretest, the controls performed five transcription practice sessions using an online platform, *The web transcription tool* (Cooke, García-Lecumberri, Maidment & Ericsson, 2005). Testing and training took place at the Speech Laboratory at UAB.

WEEK 1	Production pre-test (real words)
WEEK 2	Identification pre-tests – non-words / real words
WEEK 3	Training session 1 (ID / DIS) – non-words
WEEK 4	Training session 2 (ID / DIS) – non-words
WEEK 5	Training session 3 (ID / DIS) – non-words
WEEK 6	Training session 4 (ID / DIS) – non-words
WEEK 7	Training session 5 (ID / DIS) – non-words
WEEK 8	Production post-test + Identification post-test
WEEK 9	Generalization test (new non-words)
WEEK 10 (2 months later)	Retention test: Identification tests

Table 1. Study design and approximate timeline

## 2.6. Analysis

The percent correct identification for each sound by participant and group were calculated for each testing phase (pre-test, post-test, generalization and retention test). The L2 production data was analyzed by means of native English speaker judgments. Four Southern British English speakers were asked first to identify the sound they heard and then to rate it on a 9-point Likert scale, where 1 meant “hard to identify as the selected sound” and 9 “easy to identify as the selected sound”.

## 3. Results

### 3.1. L2 vowel perception

Correct identification scores at pre-test and at post-test were calculated for the two groups trained on vowels (ID, DIS) and the control group, and are shown in Table 2 below. Importantly, the groups did not differ statistically at pre-test ( $F(2,51)=.416, p>.05$ ). Therefore, a measure of gain (understood as the difference between posttest and pretest) was calculated (see Figure 1) and will be used for further analyses. Since testing stimuli and training stimuli were from different talkers, the comparison between pretest and posttest scores already examines generalization to new talkers.

	CONTROL		DIS		ID	
Non-words	%	SD	%	SD	%	SD
PRE	54.1	9.9	55.5	6.5	52.9	9.5
POST	57.8	10.2	65.3	9.7	79.1	13.3

Table 2. Percent correct identification at pretest and posttest per group (non-words).

As shown in Table 2, the three groups had similarly low scores at pretest and performed numerically better at post-test. This is particularly evident in the case of the ID group, whose results rose about 26 percentage points from 52.9% to 79.1% correct identification. Improvement was also observed with the DIS group (9.8 percentage points increase). The numerical improvement obtained by the control group is smaller (3.7 increase) and may reflect the influence of the English phonetics course participants were enrolled in, or simply the result of general exposure to English in this and other courses between the pre-test and the post-test

phases. The gain scores were submitted to a generalized linear mixed-effects model (GLMM), with group (ID, DIS and CG) as the fixed effect and participants as a random effect. The analysis revealed a significant main effect of group  $F(2,51)=61.288, p<.001$ . The group effect is related to the fact that the control group performed differently from the experimental groups. In fact, sequential Bonferroni pairwise comparisons confirmed that the two experimental groups outperformed the controls on the overall identification of L2 vowels ( $p<.01$  for the ID group and  $p<.05$  for the DIS group). Moreover, the ID group outperformed the DIS group ( $p<.01$ ).

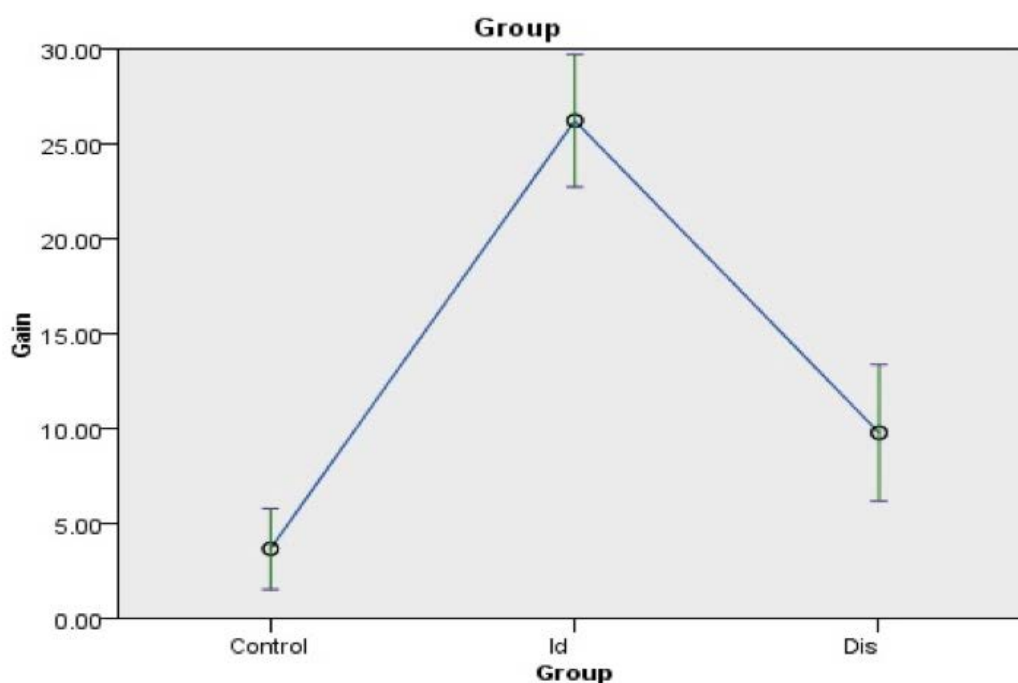


Figure 1. Identification gain (increase in correct identification percentage points) from pre to post-test per group for non-words.

Table 3 shows the mean identification scores at pre and post-test for each individual vowel for each group. It is interesting to note that some vowels seemed to improve more than others. At pre-test, on the whole all groups had the greatest difficulty identifying /æ/ and /ɜ:/, followed by /ʌ/, while /i:/ and /ɪ/ were more accurately identified. The ID group is the group that improved the most, and also the one that obtained more comparable results across vowels at post-test (76-83%, compared to 51-81% for DIS). Misidentification errors generally involved /i:/ - /ɪ/ and /æ/ - /ʌ/ confusions, while /ɜ:/ was most often misheard as /ʌ/ or /e/. The improvement

seen with the control group, mostly for the sounds /ɪ/ and /ʌ/, may be the result of the formal phonetics instruction and the consequent phonological awareness about English sounds. Generally, the ID trainees obtained numerically higher gain scores than the DIS trainees, who in turn also seemed to outperform the CG, in line with the global results across vowels previously described.

SOUND	CONTROL		DIS		ID	
	PRE	POST	PRE	POST	PRE	POST
/æ/	39.8 (17.6)	40.8 (19.9)	42.8 (18.7)	50.6 (25.0)	31.0 (21.4)	77.7 (20.1)
/ʌ/	55.2 (22.7)	62.0 (21.0)	53.4 (12.2)	66.2 (18.6)	53.5 (20.6)	75.6 (18.3)
/i:/	67.4 (15.6)	64.5 (19.2)	61.1 (13.3)	60.4 (16.1)	65.8 (12.9)	77.9 (13.5)
/ɪ/	69.0 (15.7)	80.2 (14.2)	72.0 (15.7)	81.0 (14.4)	75.0 (14.3)	82.9 (13.6)
/ɜ:/	39.1 (21.0)	41.4 (19.6)	47.9 (23.7)	68.1 (20.8)	39.0 (17.8)	81.5 (17.0)

Table 3. Percent correct identification at pretest and posttest for each individual vowel per group (non-words; standard deviations are given in parentheses).

### 3.2 L2 vowel production

Production was assessed at pre-test and at post-test and was analysed by means of native speaker judgments, following Munro (2008), among others, who advocate for the use of listeners' ratings as the most appropriate method of assessment of L2 speech: "From the standpoint of communication, there is no useful way to assess accentedness [...] except through listener responses of some sort" (p. 200). In the present study, twelve native English speakers performed a series of rating tasks that included a subset of all the stimuli so that each stimulus was evaluated by four different native English listeners. Seven identification tests with category goodness ratings were created. The rating scale ranged from 1 (difficult to recognize as the selected sound) to 9 (easy to identify as the selected sound). A reliability analysis using an intra-class correlation coefficient (ICC) with a level of "absolute agreement" was conducted on the rating scores. The results revealed a robust inter-rater agreement in all

cases, as Cronbach’s alpha values ranged from  $\alpha = .741$  to  $\alpha = .905$ . Thus, the median rating score for each participant and group at pre-test and post-test was calculated (see Table 4). The production gain scores were obtained by subtracting the pre-test scores from the post-test scores (Figure 2) and were further submitted to statistical analysis.

L2 Production	CONTROL		DIS		ID	
	Median	SD	Median	SD	Median	SD
PRE	4.5	1.4	4.7	0.8	4.6	1.5
POST	4.2	0.9	5.1	1.1	5.2	1.1

Table 4. Median rating for vowel production at pre-test and post-test per group.

As shown in Table 4, the ratings obtained by the control group showed no improvement from pre-test to post-test. The training groups, on the other hand, were given higher ratings after training. More specifically, the DIS group’s median scores improved by 0.4 and the ID improved by 0.6. The gain scores for L2 vowel production were submitted to a GLMM, with group (ID, DIS, CG) as fixed effect and participants as random effect.

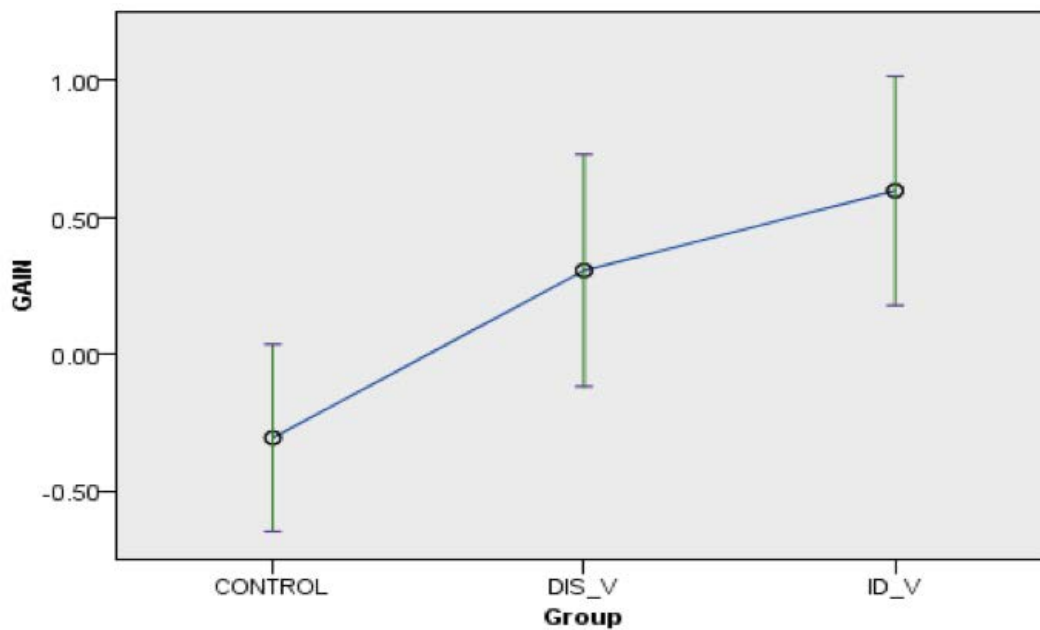


Figure 2. Production improvement from pre to post-test per group (difference between the ratings obtained at posttest and at pretest).

The results yielded a significant main effect of group ( $F(2, 50)=6.13, p<.01$ ), and pairwise comparisons with a sequential Bonferroni correction revealed

that only the ID group significantly outperformed the controls ( $p < .01$ ). The DIS group was marginally significantly better than CG ( $p = .057$ ). Moreover, the two experimental groups didn't differ in performance ( $p > .05$ ), showing a tendency towards a better performance at posttest for the DIS group too. These results suggest that the perceptual training was not only efficient in improving the learners' perception of vowel sounds, but also appeared to modify learners' production as perceived by native English speakers, particularly in the case of the ID group.

With respect to the results obtained per vowel, as was observed for perception, some vowels seemed to yield better results than others and improved to different degrees (see Table 5). No improvement was observed in the production of any of the vowels by the control group. The DIS group improved mostly in the production of vowel /æ/, while the ID group showed some improvement with all the vowels. The vowel that obtained the lowest ratings at the outset was /ʌ/, followed by the vowels /ɪ/ and /æ/. The two highest rated vowels were /i:/ and /ɜ:/. These results differ from the perception results mostly regarding two sounds: /ɜ:/, which was comparatively less successfully identified than other vowels, and /ɪ/, which was better perceived than other vowels. Both /ʌ/ and /æ/ seemed to pose difficulties to learners both in perception and production and the tense sound /i:/ was the least challenging, particularly in production.

SOUND	CONTROL		DIS		ID	
	PRE	POST	PRE	POST	PRE	POST
/æ/	4.7 (2.7)	4.1 (1.8)	4.6 (1.8)	5.8 (2.8)	4.0 (2.5)	4.5 (2.1)
/ʌ/	2.0 (2.2)	1.8 (2.8)	3.3 (2.6)	3.3 (2.2)	2.3 (1.7)	3.4 (3.2)
/i:/	5.1 (1.7)	4.9 (3.2)	5.8 (2.1)	6.1 (2.6)	6.3 (1.7)	6.6 (3.2)
/ɪ/	4.8 (2.5)	4.4 (2.9)	3.9 (2.5)	3.9 (2.9)	4.7 (2.7)	5.1 (2.2)
/ɜ:/	6.1 (1.8)	5.9 (2.8)	6.3 (2.5)	6.4 (2.1)	5.9 (2.8)	6.5 (1.6)

Table 5. Median ratings obtained for each vowel per group (standard deviations are given in parentheses).

### 3.3 Generalization effects

As previously mentioned, the main results provide evidence of generalization to novel talkers, since testing and training talkers differed. Another type of generalization investigated in this study was generalization to new items (i.e. novel non-words and real words).

#### 3.3.1 Generalization to novel non-words

In order to assess the degree to which the effects of training generalized to novel items (i.e. novel CVC non-words) produced by familiar talkers (talkers heard at the training phase), a further test was administered a week after the post-test took place. The scores of the generalization test are contrasted with both the pre-test scores and post-test scores. Generalization is considered to take place if the generalization results are as high as, or higher than, the post-test scores, and differ from pre-test results. The percentage correct identification at pre-test, post-test and generalization test by the two experimental groups (ID, DIS) and the CG are shown in Table 6. It can be observed that the three groups maintained, or even increased, their vowel identification scores from post-test to generalization test. The fact that the CG group's identification scores in the novel word generalization test were higher than at pre and post-test (68% vs. 54% and 58%, respectively) may be related to the formal instruction received. Alternatively, it is possible that these words or talkers posed fewer problems to the learners. Still, the CG's scores were lower than those obtained by the trainees.

	CONTROL		DIS		ID	
	%	<i>SD</i>	%	<i>SD</i>	%	<i>SD</i>
PRE	54.1	9.9	55.5	6.5	52.9	9.5
POST	57.8	10.2	65.3	9.7	79.1	13.3
GEN WORDS	68.4	12.4	75.9	8.3	80.4	9.8

Table 6. Percent correct identification at pre-test, post-test and generalization to novel non-words per group.

The results of the GLMM in this case showed a significant effect of group,  $F(2,153)=13.977$ ,  $p<.001$ , and Bonferroni pairwise comparisons confirmed that both experimental groups outperformed the controls in the perception of target vowels in novel non-words ( $p<.001$  for the ID group and  $p<.01$  for the DIS group). In order to further explore the results for each experimental



group, GLMM analyses were conducted on the percentage scores obtained by each trained group at the three different tests. Regarding the ID group, the results yielded a significant effect of test ( $F(2, 57)=50.42, p<.001$ ), confirming that the ID group performed significantly better after the training than at pre-test. Furthermore, pairwise comparisons with a Bonferroni adjustment confirmed that the ID's pre-test results differed from both the post-test and the generalization test results ( $p<.001$ ). Conversely, post-test and generalization results did not differ significantly. Thus, generalization to novel words was observed for the ID trainees. Regarding the DIS group, the results also revealed a significant test effect ( $F(2, 51)=33.693, p<.001$ ) and sequential Bonferroni pairwise comparisons confirmed that the pre-test scores significantly differed from both the post-test and the generalization scores ( $p<.01$ ). Interestingly, the generalization scores were significantly higher than the post-test results ( $p<.01$ ) for the DIS trainees, suggesting that these tokens (either because of the familiarity with the training talkers or the nature of the word stimuli) may have posed less of a difficulty for these learners, in line with the results observed for the CG.

### 3.3.2 Generalization to real words

Since training made use of non-words only, perception of real words was assessed at pre-test and at post-test in order to have a measure for real word identification comparable to that of nonsense word identification. Correct identification percentages for L2 vowels embedded in real words at pre-test and post-test, and the corresponding gain scores, were calculated for each group. Statistical analyses were carried out on the increase in percentage points from pretest to posttest obtained by each group, as previously done for the nonsense words. The results are given in Table 7.

Real words	CONTROL		DIS		ID	
	%	<i>SD</i>	%	<i>SD</i>	%	<i>SD</i>
PRE	72.2	11	78.2	9.7	73.1	11.2
POST	79.5	10.3	79.7	11.1	88.5	9.5
GAIN (increase in percentage points)	7.3	9.2	1.5	11.7	15.4	8.8

Table 7. Percent correct identification in real words at pre-test and post-test per group (generalization to real words).

Interestingly, vowel identification scores were higher in real words than in nonsense words already at pretest (72-78% vs. 54-56% for non-words), indicating a close relationship between lexical and phonetic categories, as discussed in the last section. Despite the high scores at pretest, improvement from pre-test to post-test was still observed. The ID, the group that improved the most with non-words (26 percentage points increase), was also the group that obtained the greatest gains with real words (15 percentage points). The DIS training regime, on the other hand, did not seem to enhance the ability to identify sounds in real words, as DIS trainees only improved by 1.5 percentage points with the training received. This slight improvement is possibly connected to the fact that their scores were higher at pre-test (78.2%), indicating that there was less room for improvement. In the case of the controls, the learners seemed to improve more in real word identification than when identifying non-words (7.3% vs. 3.7%). The GLMM analysis on gain scores yielded a significant effect of group ( $F(2,51)=8.953, p<.01$ ). Sequential Bonferroni pairwise comparisons confirmed that only the identification group outperformed the control group,  $p<.05$ . Moreover, the ID group outperformed the DIS, indicating that generalization to real words for the trained sounds only occurred after receiving identification training ( $p<.01$ ).

The identification scores for each individual vowel in the real word condition by each group are presented in Table 8. The results show that the control group appeared to improve by more than the DIS group for three out of the five sounds, namely /æ/, /ɜ:/ and /ʌ/. This is probably explained by the higher scores obtained by the DIS at the onset of the study. At post-test, however, the results for these vowels do not seem to differ much across the two groups. The ID, however, obtained numerically higher identification scores than the controls and the DIS group in the identification of /æ/, /ʌ/ and /ɜ:/ after training, and both experimental groups improved numerically more than the controls for the sound /i:/, although all three groups reached similar identification scores with both /i:/ and /ɪ/ at post-test. The pattern of difficulty on the whole matches the one found for non-word identification previously. The vowel /æ/ obtained the lowest scores, while /i:/, /ʌ/ and particularly /ɪ/ were more accurately perceived. Moreover, overall scores were higher with real word identification than with non-words, in particular regarding the sound /ɜ:/.

SOUND	CONTROL		DIS		ID	
	PRE	POST	PRE	POST	PRE	POST
/æ/	50.8 (30.8)	57.0 (34.4)	63.2 (24.9)	63.9 (27.7)	47.5 (30.2)	83.1 (25.1)
/ʌ/	71.1 (30.8)	81.2 (20.9)	86.8 (24.0)	79.8 (23.9)	83.1 (9.6)	92.5 (9.5)
/i:/	82.0 (16.4)	79.7 (17.6)	72.9 (18.3)	78.4 (14.1)	73.1 (17.8)	77.5 (17.5)
/ɪ/	89.0 (22.5)	93.7 (22.1)	88.9 (13.5)	95.1 (8.7)	90 (9.6)	94.3 (9.5)
/ɜ:/	67.9 (27)	85.9 (21.8)	79.2 (23.5)	81.2 (27.9)	71.9 (25.9)	95 (10.2)

Table 8. Percent correct identification at pretest and posttest for each individual vowel per group (real words; standard deviations are given in parentheses).

### 3.4 Retention effects

Two months after the post-test, a delayed post-test (or retention test) was administered. The aim of this test was to assess the long-term effects of training. Given that fewer participants took part in this last phase of the study, the analyses only include the results of the trainees that completed all three tests (pretest, posttest, delayed test). This explains the difference in absolute values between the results reported here and in previous sections. The total number of participants at this phase was less homogeneous among groups, as there were 9 controls, 17 ID trainees and 12 DIS trainees. In the same fashion as in the analysis of generalization results, it was considered that retention had taken place when the delayed test results were greater than the pre-test results and did not differ from (or were greater than) the post-test results. All three groups obtained numerically similar scores at post-test and retention test (see Table 9). GLMM analyses with time as the fixed effect (pre-test, post-test and delayed post-test) for each group showed that there was no significant effect of time for CG ( $F(2, 72)=1.84, p>.05$ ), confirming that this group performed similarly across all three testing times. Regarding the trained groups, the models in each case yielded a significant effect of time (ID:  $F(2, 48)=51.35, p<.001$ ; DIS:  $F(2, 33)=7.62, p<.01$ ) and Bonferroni adjusted pairwise comparisons confirmed that the performance at pre-test significantly differed from the performance at post-

test and delayed post-test ( $p < .001$  in both cases). Importantly, the delayed post-test results did not differ from the post-test results, confirming that learning was retained for a period of two months for both groups.

Test	CONTROL		DIS		ID	
	%	<i>SD</i>	%	<i>SD</i>	%	<i>SD</i>
PRE	56.7	11.3	53.0	4.2	51.8	9.7
POST	61.9	11.1	62.8	9.4	79.7	9.3
DELAYED POST	63.3	14.0	60.4	8.2	80.1	8.3

Table 9. Percent correct identification at pre-test, post-test and delayed post-test per group (data from participants who completed all three tests).

#### 4. Discussion

The goal of this study was to evaluate the efficiency of two types of perceptual tasks for improving L2 speakers' ability to identify and produce target L2 vowels. The results show that HVPT positively affected the perception of L2 vowels by Spanish/Catalan L2 learners of English, and this improvement was facilitated by both methods tested, answering the first research question of the study. The ID group improved by 26.3 percentage points from pre to post-test and the categorical DIS group improved by 9.8. The amount of gain for the two experimental groups was similar to (DIS) or greater than (ID) the range of improvement usually reported in the phonetic training literature, that is, around 10%-15% (Jamieson & Morosan, 1986; Flege, 1989; Logan & Pruitt, 1995; Flege, 1995b; Iverson & Evans, 2009; Shinohara & Iverson, 2018). In addition, instances of generalization and retention of learning were found with both methods, as discussed below.

Globally, the findings of the present study provide further evidence that HVPT is effective and that both ID and DIS tasks can make a contribution to L2 learning (Iverson et al., 2012; Shinohara & Iverson, 2018). Further, the results suggest that categorical DIS tasks can be effective for training L2 vowel perception, even if to a lesser extent than ID tasks. The findings challenge previous views on the lower efficacy of discrimination tasks that were solely based on auditory discrimination tasks (Strange & Dittmann, 1984), and are more in agreement with training studies that reported that ID and categorical DIS were equally efficient for improving the perception of English final stops (Flege, 1995b), the perception of English initial stops (Carlet, 2017), the perception of coda nasals (Nozawa, 2015), the

perception and production of the English /r/-/l/ contrast (Shinohara & Iverson, 2018), and the perception of Thai tones (Wayland & Li, 2008). Thus, the current study supports previous findings about the efficacy of a categorical DIS task and extends them to vowel perception. The positive effect of categorical DIS tasks may be related to the fact that, contrary to the auditory DIS task, the categorical DIS task exposes learners to a greater range of acoustic variability, which in turn may promote L2 categorization (Polka, 1992).

Nevertheless, the greater gains obtained for the ID group suggest a potential superiority of ID over categorical DIS for training L2 vowel perception. This result is in line with the findings of the only previous study comparing ID and DIS tasks for training L2 vowel perception (Nozawa, 2015). It is possible that a task familiarity effect may have played a role in the better performance for the ID trainees. Recall that at pretest and posttest perception was tested by means of an identification task only, potentially creating an advantage for the ID trainees. This is a limitation of the current study, as discussed below. Nonetheless, the large and significant difference between ID and DIS may not be only the result of familiarity with the task. A possible explanation for this advantage may lie in the fact ID tasks may promote between-category sensitivity and thus be more efficient for category identification, as opposed to ID tasks, which may enhance within-category sensitivity (Jamieson & Morosan, 1986; Logan & Pruitt, 1995). Moreover, ID and DIS may also differ in that DIS tasks may tap into lower levels of phonological encoding that may not contribute greatly to category formation, whereas identification may involve the type of phonological encoding that is crucial for L2 categorization (Iverson et al., 2003; Iverson et al., 2008, Iverson et al., 2012).

Another possible explanation for the superiority of the ID over the DIS training method for L2 vowel perception might be connected to the presence of labels in the ID task, i.e. the response alternatives. The presence of labels may have provided learners with the chance to focus on phonetic form (i.e., phonetic symbols and/or orthography), which has been reported to impact speech perception (Saito, 2015). Note that while identification is a covert task, in which the single category presented in each trial is directly compared with a pre-existing memory representation, discrimination is an overt process, where the two items to be compared are physically present (Bohn, 2002). Thus, the nature of the task implies that the feedback provided was also different. ID feedback provided precise information about the category that the stimulus belonged to. By

contrast, the feedback provided to DIS trainees simply informed them about whether or not the two stimuli previously heard belonged to the same category. DIS trainees were not explicitly told which category each sound belonged to. Furthermore, alongside the phonetic symbol, each label or response alternative in the ID training task also contained two keywords exemplifying the target sounds (e.g., /i:/ - cheese/leaf; /ɜ:/ - earth/first). Thus, during each trial, the identification group was forced to relate the sound they heard to a given phonetic symbol and a familiar spelling and word. There may be a link between the use of phonetic symbols and orthographic representations and the generalization to real words.

As pointed out above, one limitation of the current study is the lack of a discrimination test in addition to the identification test, which means that only the ID trainees may have benefitted from a task familiarity effect. However, Flege (1995b) and Carlet (2017) also compared ID and DIS training and evaluated only identification and reported that DIS trainees did not differ significantly from ID trainees in the identification of English stop consonants, showing no task familiarity effect. Further, no task familiarity effect was evident in the results obtained by Nozawa (2015) on the identification of final nasals, as ID and DIS trainees obtained comparable results. Furthermore, a later training study using the same stimuli as the current study (Cebrian, Carlet, Gavalda & Gorba, 2017) tested vowel trainees in both abilities (discrimination and identification), and revealed that ID enhanced the identification of vowel sounds to a greater extent than the DIS method did, extending the findings of the current study. Interestingly, the ID method enhanced learners' vowel discrimination abilities to a similar extent as the DIS method did, also in line with previous findings (Wayland & Li, 2008). Hence, the preliminary findings of Cebrian et al. (2017) confirm the superiority of the ID training method for L2 vowel identification, as this method was able to enhance both perceptual abilities (identification and discrimination) either to a similar or to a greater extent than the categorical DIS method did.

The second research question involved the effect of high variability perceptual training on L2 vowel production. The results showed that, although numerically not large, there was a significant improvement after only 5 short sessions (30-mins) of perceptual training. This result corroborates previous findings that perceptual training may alter the production of L2 sounds, at least to some extent, without the need of explicit production training (Bradlow et al., 1997; Flege, 1989; Lambacher et al, 2005; Leng-eris, 2008; Iverson et al., 2012; Thomson, 2011; Pereira, 2014; Rato &

Rauber, 2015, Shinohara & Iverson, 2018). The production results add to the observed superiority of the ID method over the DIS method for training vowel sounds, as the improvement experienced by the DIS group reached marginal significance only. It is possible that the differences between ID and DIS tasks discussed above also account for the different results regarding production. An additional explanation could stem from the fact that since production assessment included real words, the orthographic representation present in the labels of the ID training might have played a role. Also, recall that ID training was the only method that promoted generalization to real word stimuli. Thus it may follow that the group that experienced an improvement in the perception of real words also showed evidence of gains in the production of real words. Taken together, these findings fit the predictions of the SLM and the NLM, which postulate that perception gains occur prior to production gains and the former is a prerequisite for the latter. However, it seems that the learners were at the stage where perception is more developed than production, since the perceptual gains were overall greater than the production gains in the study. Thus, this result provides further evidence that the improvements in both domains do not seem to occur in parallel (Bradlow et al., 1997; Pereira, 2014; Iverson et al., 2012; cf. Rochet, 1995; Shinohara & Iverson, 2018); that is, changes in perception and production seem to develop differently.

The third research question addressed the possible differences between ID and DIS regarding generalization and retention effects. According to Flege (1995b) “a high degree of generalization suggests that a training procedure has engendered the formation of a long-term memory representation that is more abstract than the sum total of the physical properties encountered in the training stimuli” (p. 435). In the case of generalization to novel non-word stimuli produced by familiar talkers, the gain obtained during training was maintained or even increased a week later by both groups, providing evidence of robustness of learning (Logan & Pruitt, 1995). This result emphasizes the reported benefits of HVPT (Logan et al., 1991; Iverson et al., 2012; Shinohara & Iverson, 2018; among many others) and adds to previous findings that attest that both training methods (ID and categorical DIS) are effective (Flege, 1995b). The outcome is different, however, when we consider generalization to real words. First, it is relevant to note that perception of real words was better than perception of non-words, already at pretest. This may indicate that learners found it easier to recognize the vowels when they were found in words that they recognized. This may be related to the interplay between

lexical and phonological categories. Solé (2013) found that L2 contrasts that are not easily distinguishable in non-words may be differentiated in real words, indicating that L2 phonological categories may be formed after lexical categories, which are learned as a whole. Secondly, the ID was the only group that outperformed the controls and, thus, the only group that generalized the learning acquired through training to real words. The DIS group's performance with real words at pre-test was numerically higher than the ID's (DIS: 78% vs. ID: 73%), and there was no change after training (DIS: 79% vs. ID: 86%). Methodological differences discussed above, such as the covert nature and presence of labels in the case of the ID task, may account for difference between ID and DIS with real word perception.

Finally, both ID and DIS training methods were found to promote retention of learning after a period of two months, in line with several previous studies showing long-term effects of training (Bradlow et al., 1997,1999; Lively et al., 1993; Wang, 2002; Wang & Munro, 2004; Nishi & Kewley-Port, 2007; Rato, 2014). According to Flege (1995b), if knowledge acquired during training is retained over time, it may indicate that robust L2 categories have been established in the L2 learners' perceptual space. Moreover, this effect adds to the potential of phonetic training as an L2 teaching tool. All in all, the results of the delayed post-test confirm that both training methods (ID and categorical AX DIS) were able to promote long term effects and are effective when training vowel perception, in line with Flege's (1995b) findings on the perception of final stops. Moreover, the effects were retained over time, which may be an indicator of L2 category formation (Flege, 1995b).

## **5. Conclusions and implications**

This study assessed the effect of two perceptual training methods (identification and same/different categorical discrimination) on the ability to identify and produce L2 vowels. The results showed positive changes in L2 learners' perceptual and production abilities as a result of high variability phonetic training (HVPT). Specifically, the present study provided evidence that both methods are effective, as both groups of trainees outperformed a group of untrained controls in the identification of trained sounds produced by untrained talkers, and both groups showed evidence of generalization to new non-word stimuli and retention of learning. However, the current study also evidenced that identification training was more effective in promoting generalization to perception of real words and



in improving vowel production, as judged by native speaker raters. In line with these results, a combination of both tasks (ID and categorical DIS) is suggested in order to enhance different perceptual abilities and maximize the effects of training. In fact, it has been suggested that discrimination tasks could be more suitable early in the learning process when the basic dimensions of variability are being discovered (Logan & Pruitt, 1995). Moreover, Pisoni and Lively (1995) explain that both types of training can be used in order to improve different perceptual skills. While identification training improves an “acquired equivalence”, discrimination training improves an “acquired distinctiveness” (p. 445). Shinohara and Iverson (2018) argue that although both ID and categorical DIS are effective as training methods, DIS training may be easier to implement with lower proficiency learners who may not have acquired different categories for L2 sounds yet and/or for young learners who may have trouble with the use of labels. However, other studies have provided evidence that ID is favoured over DIS by L2 learners since the latter is found to be harder and somewhat tedious (Flege, 1995b; Carlet, 2017). In brief, the results of this study show that HVPT can be an efficient tool to enhance learners’ perception and production abilities and that both ID and DIS may contribute to the learning process. Unfortunately, despite the success of HVPT, phonetic training methods are rarely implemented in the classroom. There is a need to bridge HVPT research and teaching practices, by making sure that this powerful perceptual tool is pedagogically implemented.

### **Acknowledgments**

This research was made possible by the PhD grant PIF (429-02-1/2011) to the first author, the research grants from the Spanish Ministry of Economy and Competitiveness (FFI2013-46354-P and FFI2017-88016-P) and by a grant from the Catalan Government (2017SGR34).

### **References**

- Aliaga-García, Cristina & Joan Carles Mora. (2009). Assessing the effects of phonetic training on L2 sound perception and production. *Recent research in second language phonetics/phonology: Perception and production*, 2-31. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Aliaga-García, Cristina, Joan Carles Mora & Eva Cerviño-Povedano. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznań Studies in Contemporary Linguistics*, 47, 1-14.
- Best, Catherine & Tyler, Michael. (2007). Non-native and second-language speech

- perception: Commonalities and complementarities. In Ocke-Schwen Bohn & Murray J. Munro (eds.), *Language Experience in Second Language Speech Learning*, 13-34. Amsterdam/Philadelphia: John Benjamins.
- Bohn, Ocke-Schwen. (2002). On phonetic similarity. In Petra Burmeister, Thorsten Piske & Andreas Rohde (eds.), *An Integrated View of Language Development: Papers in Honor of Henning Wode*, 191-216. Trier: Wissenschaftlicher Verlag.
- Bohn, Ocke-Schwen & Munro, Murray J. (2007). *Language Experience in Second Language Speech Learning*. Amsterdam/Philadelphia: John Benjamins.
- Bradlow, Ann R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/ - /l/ contrast. In J. G. Hansen Edwards, & M. L. Zampini (eds.), *Phonology and Second Language Acquisition*, 287-308. Philadelphia: John Benjamins.
- Bradlow, Ann R., David B Pisoni., Reiko Akahane-Yamada & Yoh'ichi Tohkura. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101(4), 2299-2310.
- Carlet, Angelica. (2017). L2 perception and production of English consonants and vowels by Catalan speakers: The effects of attention and training task in a cross-training study". Unpublished PhD dissertation. Universitat Autònoma de Barcelona.
- Cebrian, Juli. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics* 34, 372-387.
- Cebrian, Juli & Angelica Carlet. (2014). Second language learners' identification of target language phonemes: A short-term phonetic training study. *Canadian Modern Language Review* 70(4), 474-499.
- Cebrian, Juli, Angelica Carlet, Núria Gavaldà & Celia Gorba. (2017). L2 vowel learning through perceptual training: Assessing training method, task familiarity and metalinguistic knowledge. Paper presented at the 18th World Congress of Applied Linguistics, Rio de Janeiro, Brazil.
- Cebrian, Juli, Joan Carles Mora & Cristina Aliaga-García. (2011). Assessing crosslinguistic similarity by means of rated discrimination and perceptual assimilation tasks. In Magdalena Wrembel, Malgorzata Kul & Katarzyna Dziubalska-Kolaczyk (eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010*, Volume I, 41-52. Frankfurt am Main: Peter Lang.
- Cooke, Martin, María Luisa García-Lecumberri, John Maidment & Anders Ericsson. (2005). The web transcription tool. Retrieved February 18, 2017, from <http://www.wtt.org.uk/>.
- Council of Europe (2001). *The common European framework of reference for languages: learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.
- Earle, F. Sayako & Emily B. Myers. (2014). Building phonetic categories: an

- argument for the role of sleep. *Frontiers in psychology*, 5, 1192-1192.
- Flege, James E. (1989). Chinese subjects' perception of the word - final English /t-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America*, 86(5), 1684-1697.
- Flege, James E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, 89(1), 395-411.
- Flege, James E. (1995a). Second language speech learning: Theory, findings and problems. In Strange (ed.), 233-277.
- Flege, James E. (1995b). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, 16, 425-442.
- Flege, James E., Ocke-Schwen Bohn & Sunyoung Jang. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437-470.
- Flege, James E. (2003). Assessing constraints on L2 segmental production and perception. In A. Meyer and N. Schiller (eds.). *Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities*, 320-355. Berlin: Mouton de Gruyter.
- Grosjean, François. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28(4), 267-283.
- Hardison, Debra M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, 8(1), 34-52.
- Højen, Anders & James E. Flege. (2006). Early learners' discrimination of second-language vowels. *The Journal of the Acoustical Society of America*, 119(5), 3072-3084.
- Ingram, John C. & See-Gyoon Park. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, 25(3), 343-370.
- Iverson, Paul & Bronwen G. Evans. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*, 122(5), 2842-2854.
- Iverson, Paul & Bronwen G. Evans. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, 126(2), 866-877.
- Iverson, Paul, Melanie Pinet & Bronwen G. Evans. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(01), 145-160.
- Jamieson, Donald G. & David E. Morosan. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð-/θ/ contrast by francophones. *Perception & Psychophysics*, 40(4), 205-215.
- Kuhl, Patricia K. & Paul Iverson. (1995). Linguistic experience and the perceptual

- magnet effect. In Strange (ed.), 121-154.
- Lacabex, Esther G., María Luisa García-Lecumberri & Martin Cooke. (2008). Identification of the contrast full vowel-schwa: training effects and generalization to a new perceptual context. *Ilha do Desterro A Journal of English Language, Literatures in English and Cultural Studies*, 55, 173-196.
- Lambacher, Stephen G., William L. Martens, Kazuhiko Kakehi, Chandrajith A. Marasinghe & Garry Molholt. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26(02), 227-247.
- Lengeris, Angelos. (2008). The effectiveness of auditory phonetic training on Greek native speakers' perception and production of southern British English vowels. In *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics, ExLing 2008*, 133-136. Athens: ISCA.
- Lively, Scot E., John S. Logan & David B. Pisoni. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, 94(3), 1242-1255.
- Logan, John S., Scott E. Lively & David B. Pisoni. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89(2), 874-886.
- Logan, John S. & John S. Pruitt. (1995). Methodological issues in training listeners to perceive non- native phonemes. In Strange (ed.), 351-378.
- McClaskey, Cynthia L., David B. Pisoni & Thomas D. Carrell. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception & Psychophysics*, 34(4), 323-330.
- Muñoz, Carmen. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, 29(4), 578-596.
- Munro, Murray J. (2008). Foreign accent and speech intelligibility. In Jette G. Hansen Edwards, & Mary L. Zampini (eds.), *Phonology and Second Language Acquisition*, 193-218. Philadelphia: John Benjamins.
- Nishi, Kanae & Diane Kewley-Port. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech Language and Hearing Research*, 50(6), 1496-1509.
- Nobre-Oliveira, Denize. (2007). Effects of perceptual training on the learning of English vowels in non-native settings. In *Proceedings of the 5th International symposium on the acquisition of second language speech, New Sounds*, Vol. 5, 382-389. Florianopolis: Federal University of Santa Catarina.
- Nozawa, Takeshi. (2015). Effects of training methods and attention on the identification and discrimination of American English coda nasals by native Japanese listeners. In Scottish Consortium for ICPHS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow.
- Pereira, Yasna I. (2014). *Perception and production of English vowels by Chilean*

- learners of English: Effect of auditory and visual modalities on phonetic training* (Unpublished doctoral dissertation). University College London, London, UK.
- Piske, Thorsten, Ian R.A. MacKay & James E. Flege. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, 29, 191-215.
- Pisoni, David B. & Scott E. Lively. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In Strange (ed.), 433-462.
- Pisoni, David B., Richard N. Aslin, Alan J. Perey & Beth L. Hennessy. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 8(2), 297-314.
- Polka, Linda. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, 52(1), 37-52.
- Rato, Anabela. (2014). Effects of Perceptual Training on the Identification of English Vowels by Native Speakers of European Portuguese. *Concordia Working Papers in Applied Linguistics*, 5, 529-546.
- Rato, Anabela & Andreia Rauber. (2015). The effects of perceptual training on the production of English vowel contrasts by Portuguese learners. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow.
- Rauber, Andreia, Anabela Rato, Denise Kluge & Giane Santos. (2012). TP (Version 3.1).[Software]. *Brazil: Worken*. [[http://www.worken.com.br/tp\\_regfree.php?l=i](http://www.worken.com.br/tp_regfree.php?l=i)].
- Rochet, Bernard L. (1995). Perception and production of L2 speech sounds by adults. In Strange (ed.), 379-410.
- Saito, Kazuya. (2015). Variables affecting the effects of recasts on L2 pronunciation development. *Language Teaching Research*, 19(3), 276-300.
- Shinohara, Yasuaki & Paul Iverson. (2018). High variability identification and discrimination training for Japanese speakers learning English /r-/l/. *Journal of Phonetics* 66, 242-251.
- Solé, Maria Josep. (2013). Phonological vs. lexical categories in an L2. In *Proceedings of the 6th Phonetics and Phonology in Iberia Conference*, 58-59. Lisbon, Portugal, Lisbon University.
- Strange, Winifred (ed.). (1995). *Speech Perception and Linguistic Experience: Issues in Cross Language Research*. Timonium, MD: York Press.
- Strange, Winifred & Sybilla Dittmann. (1984). Effects of discrimination training on the perception of /r- l/ by Japanese adults learning English. *Perception & Psychophysics*, 36(2), 131-145.
- Thomson, Ron I. (2012). Improving L2 listeners' perception of English vowels: A computer- mediated approach. *Language Learning*, 62(4), 1231-1258.
- Wang, Xinchun & Murray J. Munro. (2004). Computer-based training for learning English vowel contrasts. *System*, 32(4), 539-552.
- Wayland, Rtree P. & Bin Li. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, 36(2), 250-267.

**Appendix 1 – Perception and production stimuli**

Training stimuli					
/æ-ʌ/		/ɪ-i/		/ɜ:-e/, /ɜ:-ɑ:/	
dadge	tadge	deege	teege	darge	targe
dudge	tudge	didge	tidge	derge	terge
pav	bav	peedge	beedge	parsh	barsh
puv	buv	pidge	bidge	persh	bersh
kak	gak	keedge	geedge	karch	garch
kuk	guk	kidge	gidge	kerch	gerch
zat	zad	jeet	jeed	zart	zard
zut	zud	jit	jid	zert	zerd
vap	vab	veep	veeb	jarp	jarb
vup	vub	vip	vib	jerp	jerb
vak	vag	veek	veeg	vark	varg
vuk	vug	vik	vig	verk	verg
Testing stimuli					
/æ-ʌ/		/ɪ-i/		/ɜ:/	
vab	vap	veeb	veep	jurb	
zad	zat	jeed	jeet	jerd	
vag	vack	veeg	veek	verg	
vub	vup	vib	vip	jurp	
zud	zut	jid	jit	jurt	
vugg	vuck	vig	vick	verk	
Generalization to real words stimuli					
/æ-ʌ/		/ɪ-i/		/ɜ:/	
cap	cab	feet	feed	hurt	heard
pup	pub	bit	bid		
Generalization to novel non-words stimuli					
/æ-ʌ/		/ɪ-i/		/ɜ:/	
dack	pag	fip	pid	vert	derg
dut	Jud	geep	keeb		
Production elicitation list					
/æ-ʌ/		/ɪ-i/		/ɜ:/	
cap	cab	bit	bid	hurt	heard
buck	bug	feet	feed		



## Perception of Brazilian Portuguese Nasal Vowels by Danish Listeners

Denise Cristina Kluge  
Federal University of Rio de Janeiro (UFRJ)

### Abstract

The word-final nasals /m/ and /n/ have different patterns of phonetic realizations across languages, whereas they are distinctively pronounced in English and Danish, in Brazilian Portuguese (BP) they are not fully realized and the preceding vowel is nasalized. Bearing in mind this difference, the main objective of this study was to investigate the perception of BP nasal vowels by Danish learners of BP. Two discrimination and two identification tests were used and taken by two groups composed of ten Danish learners of English, as a reference for comparison, and ten Danish learners of BP. General results showed both groups had similar difficulties in both discrimination tests. It was less difficult for the Danish learners of BP to identify the BP native-like pronunciation when presented in contrast to a non-native-like pronunciation.

### 1. Introduction

Many studies concerning the perception of second language (L2) sounds have discussed the influence of the native language (L1) on accurate perception of the L2 (Flege, 1993, 1995; Wode, 1995; Best, 1995; Kuhl & Iverson, 1995). Moreover, some L2 speech models have discussed the role of accurate perception on accurate production (Flege, 1995; Best, 1995; Escudero, 2005; Best & Tyler, 2007). According to some studies (Schmidt, 1996; Harnsberger, 2001; Best, McRoberts & Goodell, 2001; Best & Tyler, 2007), it is usually believed that, at least in initial stages of L2 learning, adults are language-specific perceivers and that they perceive L2 segments through the filter of their L1 sound system.

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 121-133). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.



As posited by Flege (1981), L2 sounds may be perceived in terms of those of the L1 by the learner, making this perception different from that of a native speaker. For example, sounds that are separate phonemes in an L2 might be merely allophones of the same phoneme in the L1. Furthermore, Flege states that this may influence the production of L2 sounds by a native speaker of this L1 because of the identical mental representation that this speaker has for the two sounds. Flege (1995) also posits in his Speech Learning Model that the perceived relationship between L1 and L2 categories plays an important role in correctly perceiving or producing L2 sounds. According to one of the hypotheses of this model, L1 and L2 sounds are “related perceptually to one another at a position-sensitive allophonic level” and acquisition of L2 sounds depends on the perceived dissimilarity between L1 and L2 sounds (Flege, 1995, p.239).

Kuhl (1993) proposes the Native Language Magnet model of speech perception and language development, which works with the concept of L1 phonetic prototypes, or the best exemplars of certain phonetic categories. These prototypes would act as perceptual magnets that pull the surrounding L2 sounds toward the same perceptual phonetic space occupied by the L1 prototype. She states that the nearer the L2 sounds are to the L1 prototype, the more difficult it becomes for the L2 speakers to discriminate L1 and L2 speech sounds.

Bearing in mind the perspective of both perception models, this article aims at investigating the perception of Brazilian Portuguese (BP) syllable-final nasal vowels by Danish learners of Brazilian Portuguese as a foreign language. In order to understand the difficulties Danish learners may face in BP with nasal vowels in word-final position, phonological differences between languages have to be considered regarding nasal vowels and nasal consonants in syllable final position. According to Fujimura and Erickson (1997), typically, nasal consonants have a place distinction between /m/ and /n/ as in English and in Danish (Ocke Bohn, 2013, personal communication). However, some languages have no place distinction for nasal consonants in syllable-final position, as Brazilian Portuguese, for instance. In BP, the nasal consonants /m/ and /n/ are not fully realized after a vowel in syllable-final position and sometimes not realized at all, and the preceding vowel is nasalized.

According to the literature, the degree of vowel nasalization differs between languages, from subtle as in English (Giegerich, 1992; Hammond,

1999; Ladefoged, 2006) to strong as in BP (Oliveira & Cristófaros-Silva 2005). It is important to state that although vowel nasalization can occur in languages such as English and Danish, for instance, nasalization of the vowel is not used to distinguish meaning in English (Ladefoged, 2005), thus vowel nasalization is not a distinctive feature. In BP, nasalization is quite an issue and has motivated different explanations and theories, but, in general and for the purpose of this study, it is assumed that: (a) phonetically, the nasal consonants /m/ and /n/ are not fully realized after a vowel in word-final position and sometimes not realized at all; and (b) the vowel assimilates nasalization from the following nasal consonant (Cristófaros Silva, 1999; Mateus; D'andrade, 2000; Câmara Jr., 1971; Kluge et al., 2009; Kluge, 2010). The differences regarding the pronunciation of word-final nasals /m/ and /n/ in Danish, English and in Brazilian Portuguese are extremely relevant to understand the difficulties that Danish learners of BP may have in the accurate discrimination/identification of BP syllable-final vowel nasals.

## **2. Method and procedures**

The data collection occurred from February 19th to March 07th, 2013 at Aarhus University. Two discrimination tests (same or different and AXB) and two identification tests (native vs. nonnative pronunciation and oral vs. nasal vowels) as well as a questionnaire for assessing biographical information about the participants, and a word recognition test were designed for this study and administered to 20 Danish listeners divided into two groups: (1) ten Danish learners of BP; and (2) ten Danish learners of English, as a reference for comparison.

The group of Danish learners of BP consisted of 10 undergraduate student from the Bachelor's degree program in Brazilian Studies at Aarhus University at the time of data collection: 2 men and 8 women, ages ranging from 20 to 29 (mean 23,2). Nine participants reported Danish as L1 and one participant reported Danish and Finish as L1. As regards foreign language, they all mentioned English as a L2. They all reported having no hearing problems. The questionnaire that assessed the participants' profile considering BP learning showed that (a) 8 participants started learning Portuguese after 20 years old (from 20 to 29); (b) 2 participants reported having contact to Portuguese since childhood but not regular/formal learning; (c) 7 participants have studied Portuguese for 6 to 8 months by the

time of data collection; and (d) 3 participants have studied Portuguese for 2,5 to 3 years by the time of data collection. Regarding their experience in Portuguese speaking countries, 3 participants said they had lived in Brazil for 1 to 6 months about 1 year before data collection and 1 participant had lived to Portugal for one month 2 years before data collection. The questionnaire showed the participants BP usage regarding hours speaking and listening to Portuguese per day in terms of percentage which varied from 10 to 40 per cent.

The group of Danish learners of English consisted of 10 undergraduate student from the Bachelor's degree program in English at Aarhus University at the time of data collection: 6 men and 4 women, ages ranging from 20 to 27 (mean 24,4). All participants reported Danish as L1 and one participant reported German as a L2 besides English. They all reported having no hearing problems. The questionnaire that assessed the participants' profile regarding English learning showed that (a) 1 participant started learning English at the age of 5; (b) 1 participant started learning English at the age of 8, (c) 5 participants started learning English at the age of 10, and (d) 3 participants started learning English at the age of 11-12. Two participants also reported they had lived in English speaking country for 6 -10 months about 5 years before data collection. The questionnaire also showed the participants reported spending from 10 to 80 percent of the day either listening to or speaking English.

The stimuli of the four perception tests were recorded in a phonetic lab at a university in Brazil by three female native speakers of BP ages ranging from 23 to 50 years old (mean 36) with no knowledge of Danish. All of them were advanced speakers of English and were phonetically trained to pronounce the target words with and without vowel nasalization, whenever necessary.

The stimuli of same or different discrimination test consisted of 5 two-syllable words: *sabão* – 'soap', *porém* – 'however', *assim* – 'so', *batom* – 'lipstick', *atum* – 'tuna'. Each word was recorded in two different conditions: with and without vowel nasalization. That is, in the word *sabão* – 'soap', for example, the word-final vowel was recorded as a nasal vowel as well as an oral vowel by the three talkers. In each trial, the participants heard two realizations of the same word and had to indicate if the words were the same or different regarding the final sound. Each word was spoken by a different talker within a trial. Each word appeared in 4 trials varying in position of appearance (first or second position) and contrasting vowel

nasalization or not (same or different). The test consisted of 40 trials (4 trials x 5 words x 2 repetitions).

The stimuli of AXB discrimination test consisted of 5 monosyllabic words: *não* – ‘no’, *bem* – ‘well’, *sim* – ‘yes’, *bom* – ‘good’, *pum* – ‘fart’. Each word was recorded in two different conditions: with and without vowel nasalization, like the other discrimination test previously described. In each trial, the participants heard three realizations of the same word produced by three different talkers and had to indicate which final sounds of the words they heard as the same by clicking in one of the three options: “first 2 words”, “last 2 words”, “three words”. In order to investigate whether the participants would perceive any difference regarding the three pronunciations of the target word in the trial, a third answer option was included. Each word appeared in 6 trials varying in position of appearance (first, second or third position) and contrasting oral and nasalized vowels or not. The test consisted of 30 trials (6 trials x 5 words).

The first identification test was a native vs. nonnative judgment test with 5 monosyllabic words *não* – ‘no’, *sem* – ‘without’, *fim* – ‘end’, *com* – ‘with’, *um* – ‘one’. As with the other tests, each word was recorded in two different conditions: with and without vowel nasalization. In each trial, the participants heard two realizations of the same word and had to indicate which pronunciation sounds more BP native-like by circling “1” (if it was the first they heard), “2” (if it was the second they heard); “both” (native-like pronunciations); or “neither” (native-like pronunciation). For this test, the stimuli were from two out of the three female talkers. Therefore, within the trial, each pronunciation of the target word was spoken by one of the talkers. Each word appeared in 4 trials varying in position of appearance (first or second position) and contrasting oral and nasalized vowels or not. The test consisted of 40 trials (4 trials x 5 words x 2 repetitions).

The second identification test contrasting oral and nasal vowels consisted of 3 pairs of syllables contrasting oral and nasal vowels *sá-sã*; *fã-fã*; *lá-lã*. The participants heard one realization of a syllable and had to indicate the vowel they heard: oral or nasal. The response alternatives were: *á* (oral vowel) and *ã* (nasal vowel). Each target syllable was pronounced by the three female talkers. The test consisted of 36 trials (6 syllables x 6 repetitions).

All the perception tests were designed and administrated using TP a free software to design perception test (Rauber et al, 2012). For all the four tests, the participants were only allowed to listen to each trial once before clicking on their answer. In order to avoid order effect, the stimuli were randomized for each participant for each test. The data was collected individually on a laptop computer by the researcher. It took from 20 to 22 minutes for the Danish learners of BP and from 10-12 minutes for the Danish learners of English. Instructions were given in English to both groups. Before starting each test, the participants did a familiarization test, that is, a short practice test before each of the four tests in order to get familiar to the task itself and clear any possible doubt they might have.

The order of data collection was: (1) Questionnaire; (2) Instructions and familiarization test; (3) Discrimination test: same or different; (4) Instructions and familiarization test; (5) Discrimination test: AXB; (6) Instructions and familiarization test; (7) Identification test: N vs. NN-like pronunciation; (8) Instructions and familiarization test; (9) Identification test: oral vs. nasal vowel; (10) Familiarity with the corpus. The Danish learners of English only did the discrimination tests (steps 1-5).

The statistical analysis was based on correct responses for each perception test as follows: (a) Same or Different Test: 40 trials x 10 participants= 400 responses for each group.; (b) AXB test: 30 trials x 10 participants= 300 responses for each group; (c) Native-like vs. Nonnative-like test: 40 trials x 10 participants= 400 responses for the Danish learners of BP; and (d) Oral vs. Nasal vowel test: 36 trials x 10 participants= 360 responses for the Danish learners of BP. Statistical significance (alpha level) was set at .05, and due to the limited number of participants and non-consistency between the results of skewness and kurtosis, the entire data were considered not normally distributed. Thus, non-parametric tests were used: Mann-Whitney (Inter groups) and Wilcoxon (Intra groups) using SPSS version 18.0. In this study, only significant results of the statistical tests are reported.

### **3. Results**

With regard to the first discrimination test, Same or Different test, contrasting the realization of the nasal and the oral vowels, Table 1 shows the correct responses in percentages for both groups of Danish: learners of BP and English.

	Danish Learners of BP			Danish Learner of English		
	Same	Different	Total	Same	Different	Total
P1	90	90	90	65	95	80
P2	70	60	65	70	85	77.5
P3	60	75	67.5	100	25	62.5
P4	85	60	72.5	70	85	77.5
P5	75	90	82.5	75	100	87.5
P6	65	55	60	100	70	85
P7	95	75	85	85	75	80
P8	25	100	62.5	85	60	72.5
P9	95	100	97.5	85	80	82.5
P10	60	95	77.5	80	80	80
Total	72	80	76	81.5	82.5	82
<i>SD</i>	21	17	12	12	21	7

**Note:** *SD*= Standard Deviation

Table 1. Responses of the Same or Different Discrimination Test in percentage (%).

Table 1 shows that accurate responses ranged from 25 to 95% for the same realizations of the same word for the English learners and from 65 to 100% to BP learners. As for the different realizations it ranged from 60 to 100% for the English learners and from 25 to 100% to the BP learners. Regarding intra group analysis, Wilcoxon tests revealed no significant differences regarding the same or different trial for all of the groups. Overall results showed that both groups had similar performance levels in the test and this was confirmed by Mann-Whitney tests as the results for inter group analysis showed no significant difference.

The second discrimination test, AXB, contrasted three pronunciations of the same word regarding the realization of the nasal vowel or not (oral vowel). Table 2 shows the correct responses in percentages for both groups of Danish: learners of BP and English, considering the three possible answers in: “first 2 words”, “last 2 words”, and “three words”.

	English Learners				BP Learners			
	first 2	last 2	Three	Total	first 2	last 2	three	total
P1	80	80	80	80	80	80	50	70
P2	70	90	80	80	50	80	70	66.7
P3	40	50	60	50	60	80	70	70
P4	60	50	70	60	60	80	40	60
P5	40	70	60	56	80	60	70	70
P6	40	60	70	56	60	100	40	66.7
P7	60	70	60	67	80	60	20	53.4
P8	70	90	50	70	50	70	60	60
P9	70	80	60	70	70	70	60	66.7
P10	80	30	55	55	80	50	50	60
Total	61	67	64.5	64.2	67	73	53	64.3
<i>SD</i>	15	19	10	10	14	16	6	6

**Note:** *SD*= Standard Deviation

Table 2. Responses of the AXB Discrimination Test in percentages (%).

Overall results showed that accurate responses ranged from 50 to 80% for the English learners and from 53.4 to 70% for the BP learners. Regarding intra group analysis, Wilcoxon tests revealed no significant differences for the English learners. Performing statistical analysis of BP learners, Wilcoxon tests revealed significant differences for the results of trials with contrast (first 2 words vs. last 2 words) vs trials with no contrast (three words) ( $Z=-2.203$ ,  $p=.028$ ). These results indicate that the BP learners were better at discriminating the BP nasal vowels when in contrast to oral vowel realizations, thus indicating an effect of trial type. With regard to inter group analysis, results of the Mann-Whitney test showed no significant differences, as with the same or different test.

The third test solely included the Danish learners of PB and was a native-like vs. a nonnative-like identification test contrasting the realization or not of the BP nasal vowel. Table 3 shows the correct responses in percentages, considering the trial in which the native-like realizations of the BP nasal vowels appeared in contrast to the nonnative-like one (“1”, “2”, “3”) and trials in which there were no contrast in pronunciation (“both” and “neither”).

	Trials with contrast	Trials without contrast
P1	95	90
P2	85	40
P3	75	15
P4	85	70
P5	100	80
P6	100	90
P7	80	60
P8	95	90
P9	90	85
P10	85	55
Total	89	67.5
<i>SD</i>	8.4	25

**Note:** *SD*= Standard Deviation

Table 3. Responses of the Native-like vs. Nonnative-like Identification Test in percentages by the BP learners (%).

Overall results showed that accurate responses ranged from 75 to 100% for trials with contrast of native-like vs. nonnative-like pronunciation and from 15 to 90% for trials with no contrast, thus showing a higher variability. Wilcoxon test revealed significant differences for the BP learners ( $Z=-2,812$ ,  $p=,005$ ) for trials with and without contrast, thus indicating that the Danish learners of BP show less difficulty in identifying the BP native-like pronunciation when presented in contrast.

As for the fourth test, the nasal vs. oral vowel identification test, Table 4 shows the result for the learners of BP.



	A	Ã
1	100	94
2	83	94
3	94	38
4	94	88
5	100	61
6	61	77
7	94	94
8	100	66
9	100	100
10	88	77
Total	91.4	78.9
SD	12	19

**Note:** SD= Standard Deviation

Table 4. Responses of the nasal vowel Identification test by the BP learners (%).

Overall results showed that Danish learners of BP were better at identifying the oral vowel (91.4%) than the nasal ones (78.9%). However, a Wilcoxon test was performed and revealed no statistically differences possibly due to the limited number of data

#### 4. Final considerations

The main objective of this small-scale study was to investigate the perception of Brazilian Portuguese syllable-final nasal vowels by Danish learners of Brazilian Portuguese as both languages differ in terms of nasalization specifically in syllable-final position. Two groups of Danish speakers took part in the study: one group of BP learners and a group of English learners as a matter of comparison. Both groups took two discrimination tests: a Same or different test and an AXB test. General results showed that both groups showed similar difficulties in both discrimination tests: Same or Different and AXB. The two identification tests were taken just by the BP learners.

For the native-like vs. non-native-like identification test, it was less difficult for the BP learners to identify the BP native-like pronunciation of the nasal vowel when it was presented in contrast to a non-native-like pronunciation, that is, an oral vowel realization. In the other Identification

test where the participants were asked to identify the oral and the nasal BP vowel, it was less difficult for the BP learners to identify the BP oral vowel than the nasal one; however statistical analysis showed no significance.

Regardless of the limited number of data of this small scale study, there are indications that there is a certain degree of L1 interference when Danish learners of BP perceive the BP nasal vowels in syllable-final position as predicted by Flege's model, for instance. Further studies could also analyze influence of phonological context such as the vowel or preceding context. Production and its relationship to perception may be also a great field of investigation.

## 5. Acknowledgment

I would like to thank Prof. Dr. Ocke Bohn for accepting me as visiting researcher at Aarhus University from January to April, 2013 as part of the Coimbra Group Scholarship Programme for Young Professors and Researchers from Latin American Universities. During my stay, under the invaluable supervision of Prof. Dr. Ocke Bohn I was able to conduct this study and learn so much from it. It was such an honor to work with Prof. Dr. Ocke Bohn. I also would like to thank the Coimbra Group for funding the present research.

## References

- Best, C. T. (1995). A direct realistic view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-206). Timonium, MD: York Press.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*(2), 775-794.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In: Bohn, Ocke-Schwen and Murray J. Munro (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34). Amsterdam: John Benjamins.
- Câmara Jr, J. M. (1971). *Problemas da lingüística descritiva*. Petrópolis: Editora Vozes
- Cristófaró Silva, T. (1999). *Fonética e fonologia do português: Roteiro de estudos e guia de exercícios*. São Paulo: Contexto.

- Escudero, P. (2005). Linguistic Perception and Second Language Acquisition. Explaining the attainment of optimal phonological categorization. Doctoral dissertation, Utrecht University.
- Flege, J. E. (1981). The phonological basis of foreign accent: a hypothesis. *TESOL Quarterly*, 15, 443-455.
- Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, 93(3), 1589-1608.
- Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-272). Timonium, MD: York Press.
- Fujimura, O. & Erickson, D. (1997). Acoustic Phonetics. In W.J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 65-115). Cambridge: Blackwell Publishers.
- Giegerich, H. J. (1992). *English phonology: An introduction*. Cambridge: Cambridge University Press.
- Hammond, M. (1999). *The phonology of English: A prosodic optimality theoretic approach*. Oxford: Oxford University Press. Chapters 1, 2, 3.
- Harnsberger, J. D. (2001). The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: A multidimensional scaling analysis. *Journal of Phonetics*, 29, 303-327.
- Kluge, D. C. (2009). Brazilian EFL learners' identification of word-final /m-n/: native/nonnative realizations and effect of visual cues. Unpublished doctoral dissertation. Universidade Federal de Santa Catarina: Florianópolis.
- Kluge, D. C.; Reis, M.S., Nobre-Oliveira, D., Bettoni-Techio, M.(2009). The use of visual cues in the perception of English syllable-final nasals by Brazilian EFL learners. In M.A. WATKINS, M.A.; RAUBER, A. S. & BAPTISTA, B.O. (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production* (pp. 141-153). Cambridge Scholars Publishing.
- Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics*, 21, 125-139.
- Kuhl, P. K. & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp.121-154). Timonium, MD: York Press.
- Ladefoged, P. (2005). *Vowels and consonants: an introduction to the sounds of Languages*. Second Edition. Oxford: Blackwell Publishing.
- Ladefoged, P. (2006). *A course in phonetics*. Fifth Edition. USA: Thomson Wadsworth.
- Mateus, M. H. M., & d'Andrade, E. (2000). *The phonology of Portuguese*. Oxford: Oxford University Press.

- Oliveira, J. C. & Cristófaros-Silva, T., (2005). Aprendizado de língua estrangeira: o caso da nasalização de vogais. Unpublished paper. Universidade Federal de Minas Gerais. Available at <http://www.cori.unicamp.br/jornadas/completos/UFMG/ND1010.doc>
- Schmidt, A. M. (1996). Cross-language identification of consonants. Part 1. Korean perception of English. *Journal of the Acoustic Society of America*, 99(5), 3301-3211.
- Wode, H. (1995). Speech perception, language acquisition and linguistic: Some mutual implications. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 321-350). Timonium, MD: York Press.



## Accent Matters in Perception of Voice Similarity

Mette Hjortshøj Sørensen  
Aarhus University

### Abstract

This study investigates how voice similarity is perceived by three different groups of listeners, namely by native listeners, by non-native listeners and by a group of listeners with no prior knowledge of the language. The study explores *whether* listeners can distinguish between voices and also *how similar* the listeners perceive the voices to be. The participants all listened to short recordings of 60 voice pairs of young male speakers speaking Danish and were asked to make a decision on whether they thought the voices sounded similar or not on a sliding scale. The results suggest that most of the listeners use the difference in fundamental frequency when deciding whether two voices sound similar or not. However, for the native listeners a change in regional accent seems to trump mean fundamental frequency as a deciding factor for judging voice similarity.

### 1. Introduction<sup>1</sup>

The speech signal carries tremendous amounts of information simultaneously. At the same time as the linguistic message is being delivered, indexical information about the speaker's identity, sex, regional origin, age, socioeconomic status, physical and emotional state is also present in the speech signal (Johnson, Westrek, Nazzi & Cutler, 2011).

---

<sup>1</sup> Parts of the findings reported in this contribution were presented at the IAFPA (International Association for Forensic Phonetics and Acoustics) Annual Conference 2010 in Trier (Sørensen 2010) and part of my doctoral research (Sørensen 2011). I am grateful for the comments and suggestions from the audience at IAFPA 2010. I would also like to thank Ocke-Schwen Bohn for the many discussions we have had about perception in general.

Studies on perception of second language (L2) sounds have long discussed the influence that the first language (L1) inevitably will have on perception of the L2 sounds (e.g. Flege, 1993; Wode, 1995; Best, 1995). It is generally accepted that adult learners are language-specific perceivers, at least in the initial stages of L2 learning (e.g. Best & Tyler, 2007). That is, the adult language learners process the L2 segments by means of their L1 sound inventory. These studies focus primarily on the phoneme inventory in the second language.

The present study explores whether listeners also listen through the filter of their native language when they are asked to judge voice similarity. It is clear that in spoken language, segmental information cannot be completely disentangled from the indexical information that is also present in the speech signal at the same time as the linguistic message (Johnson et al 2011). Although what counts as being indexical information in one language may be matter of phoneme identity in other languages, i.e. this is rather language specific (e.g. Gordon & Ladefoged, 2001). For example, in Jalapa Mazatec creaky voice is used to signal the difference between /jǎ/ meaning “he swears” and /já/ meaning “tree” (Kirk, Ladefoged & Ladefoged, 1993) whereas creaky voice primarily serves as indexical information in English. Hence, there will be a certain language-specificity to what counts as indexical information as well as sound inventory. Kreiman and Gerratt (2010) suggest that the native language of a listener does affect the listener’s sensitivity to voice characteristics as well as the perceptual strategy. Consequently, people are not surprisingly also more accurate when recognising voices in their native language compared to another language (Köster & Schiller, 1997).

Very little is known, however, about what listeners actually use as deciding cues or parameters when they listen to and apparently judge some voices to sound very similar and other voices to sound very different from one another. It may also differ what people actually consider as being part of the voice – whether it is laryngeal settings or whether some listeners also include supralaryngeal settings as part of their concept of ‘voice quality’.

Grønnum (2005) asserts that *intonation* is the strongest marker of dialects or regional varieties in Danish, and Kristiansen, Maegaard and Phrao (2011) also found that Danish speakers primarily seem to use intonation as a cue when identifying different types of Danish regional varieties. In fact, Danish is often described as being a relatively uniform language regarding variation at the segmental level (e.g. Grønnum, 1994;

Kristiansen, 2003). Segmental variation used to be a prominent part of Danish dialects in the past, but the segmental variation has been replaced by intonation as the more salient feature in modern Danish (Gregersen & Pharao, 2016). The term ‘regional accent’ will be used in the current study to stress that the differences found in the data are primarily differences in intonation patterns. In a study by Gooskens (1997) examining whether English and Dutch listeners rely more on segmental variation or intonation when identifying dialects, the results suggest that intonation also seems to be more important for the identification of English dialects whereas it appears less important for the identification of Dutch dialects.

Studies on recognition of *voices* also show that listeners have a higher success rate at *remembering* and *recognising* speakers who have either relatively high or relatively low fundamental frequencies (F0) compared to speakers with a more average fundamental frequency and this goes for English (e.g. Foulkes and Barron, 2000) as well as for Danish listeners (Sørensen, 2012). This suggests that – at least English and Danish listeners – appear to rely heavily on speakers’ F0 when listening to voices. Foulkes and Barron (2000) suggest that not only the mean F0 itself, but also the standard deviation (St. dv.) of F0 could have a correlation with the recognition rate in a speaker recognition test. Foulkes and Barron state that measuring the standard deviation is useful in some cases, as it enables a quantification of the F0 variation used by a speaker. According to Foulkes and Barron, speakers who are perceived as sounding monotonous most often would also have a lower standard deviation associated with their mean fundamental frequencies.

The aim of the present study is primarily to investigate whether voice similarity is perceived through the filter of the listener’s native language like e.g. segments are (e.g. Flege, 1993; Best, 1995). The study examines whether native listeners, non-native listeners, and listeners with no prior knowledge of the language in question focus on the same or different acoustic cues when they are judging voice similarity. That is, do people listen to speakers in other ways when they listen to other languages compared to their own native language? The present study then extends upon some of the previous research by exploring whether listeners can discriminate between voices, but also by investigating how similar or different the listeners perceived the voices in the study to be. The focus in this study will be on the possible correlation between mean F0 and perceived similarity of voices. That is, would a small measured



difference in fundamental frequency entail a small perceived difference between voices and would a larger measured difference in fundamental frequency between two voices entail a larger perceived difference between the voices?

Assuming that listeners focus on different cues in the voices depending on their familiarity with the language spoken, this may have an effect on whether voices are judged to be similar or not. The underlying assumption of this voice perception study is that the listeners with no prior knowledge of a given language will have to listen to the voice quality in a more global (as opposed to local) manner than the native listeners would. In other words, listeners with no prior knowledge of the language in question would probably solely make use of suprasegmental features, as they would have no prerequisite for what else to listen for – whereas the native listeners may listen for both subtle segmental and suprasegmental information, e.g. regional accent, intonation or other linguistic features when they perceive and judge voice similarity between speakers.

## **2. Method**

### **2.1 Stimuli**

The stimuli consist of recordings of spontaneous speech from 15 young Danish male speakers between 20 and 35 years of age. The speakers' F0 varied, but speakers with any other distinctive/characteristic voice qualities, like e.g. nasal, hoarse or creaky voice were excluded from this study. Furthermore, occurrences of any other linguistic cues to regional variety, e.g. regional vocabulary or grammatical constructions that are region specific were excluded as well. 12 of the young male speakers form a relatively homogeneous group from Eastern Jutland in Denmark, all speaking Danish with a regional (but not strong) accent. There are three additions to this otherwise homogenous group of speakers, namely one young male speaker from the Northern part of Jutland in Denmark and two young male speakers from the Copenhagen area in Denmark. These voices were added to the study to test whether the listeners would react to a change in the regional accent spoken. Small samples of 3 seconds of duration were extracted from the speakers and these were then presented in pairs. In total the stimuli consisted of 60 voice pairs of 2 x 3 seconds of speech.

## **2.2. Listeners**

Three groups of listeners participated in the study: A group of native listeners, a group of non-native listeners and a group of listeners with no prior knowledge of Danish. The first group was a group of 20 native listeners (21-40 years old) from Eastern Jutland in Denmark. The second group consisted of 20 non-native listeners with English as L1 (age 24-35 years old) who speak Danish as an L2 language at different levels of proficiency. It proved difficult to recruit participants with similar levels of proficiency in Danish, so the criteria for this group was that all the listeners had to be adult when arriving in Denmark, all of them lived in Denmark and all of them had first-hand knowledge of Danish. The third group, the listeners with no prior knowledge of Danish, were English L1 speakers (20-36 years old) from York in England and none of these speakers had any knowledge of Danish. All of the listeners from all of the groups self-reported normal hearing.

## **2.3. Procedure**

The speech perception software ‘Alvin’ (Gayvert & Hillenbrand, 2003) was used and modified to suit the present study. The listeners all listened to 60 voice pairs over high quality headphones (Sennheiser HD 280 Pro) on a laptop. The 60 voice pairs were played in random order and all of the voice pairs occurred twice in order to explore whether the listeners were consistent in their judgements throughout the study. After listening to each voice pair, the listener was asked on the screen to make a decision on how similar the voices just heard were on a sliding scale going from “very different” on one end to “very similar” in the other end. The listener would then move the slider accordingly on the screen and press ‘okay’ and after this the next voice pair would be played automatically and so forth. Order effects were checked for as well in the study. That is, some of the voice pairs were not only played twice, but also in reverse order.

## **3. Results**

As mentioned, previous research suggest that speaker’s F0 may be one of the important features when listeners notice and remember voices (e.g. Foulkes and Barron, 2000; Sørensen, 2012). For the current voice similarity perception study it was therefore also a priority to examine whether the actual measured difference in fundamental frequency was also reflected

by the *perceived* similarity, i.e. whether there was actually a correlation between *measured* difference in fundamental frequency and the listeners' ratings of voice similarity.

The scatter plot in Figure 1 shows the difference in mean F0 between the voices in all the voice pairs measured in Hz on the X-axis compared with the perceived difference between the voices in the voice pairs on the Y-axis. Low numbers on the Y-axis correspond to a small perceived difference between the voices and higher numbers correspond to a larger perceived difference.

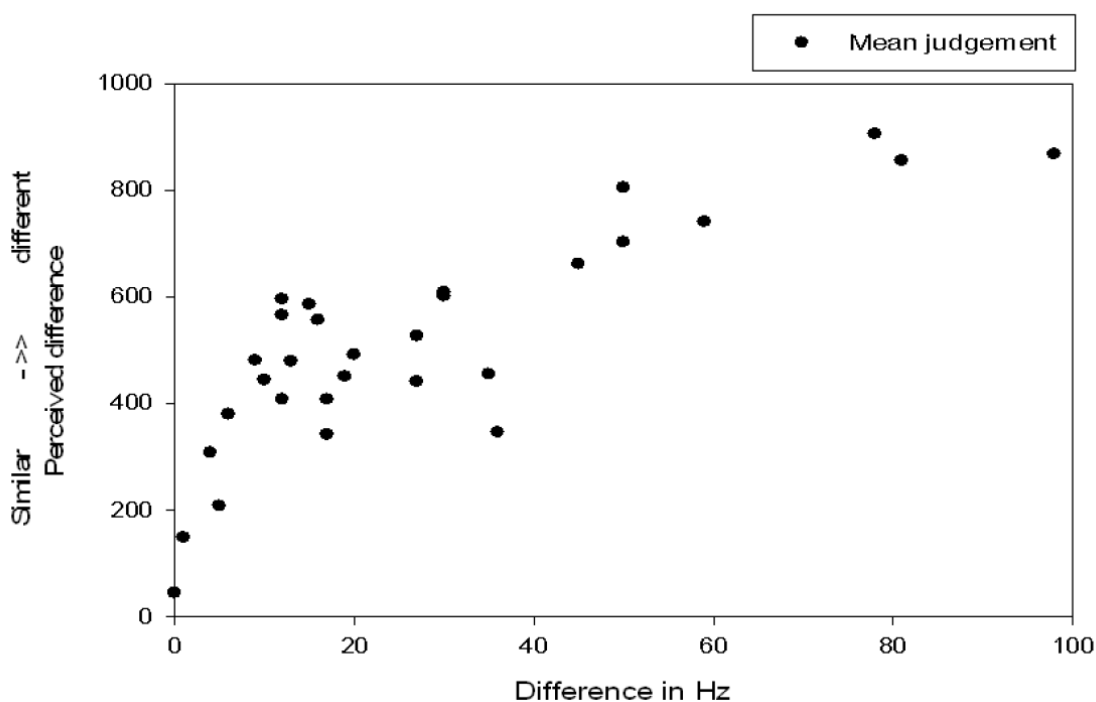


Figure 1. Results from the voice perception study showing correlation between the acoustic difference in mean F0 between the heard voices and the perceived difference between the voices.

Figure 1 shows the mean of all the listeners' trials from all the groups. The figure indicates that, in general, as the acoustic difference between the two voices in voice pair goes up, listeners will also perceive a larger difference. This was confirmed by correlation analysis (Pearson's  $r$ ) which showed that the correlation coefficient is  $r=.83$  ( $p<.001$ ). The results from the current voice perception study suggest that, in general, most of the listeners seem to use distance – or difference – in fundamental frequency as an important cue to judge voice similarity most of the time. Figure 2

shows the same results as are shown in Figure 1, but this time the results are divided into the mean scores for each of the three different groups of listeners.

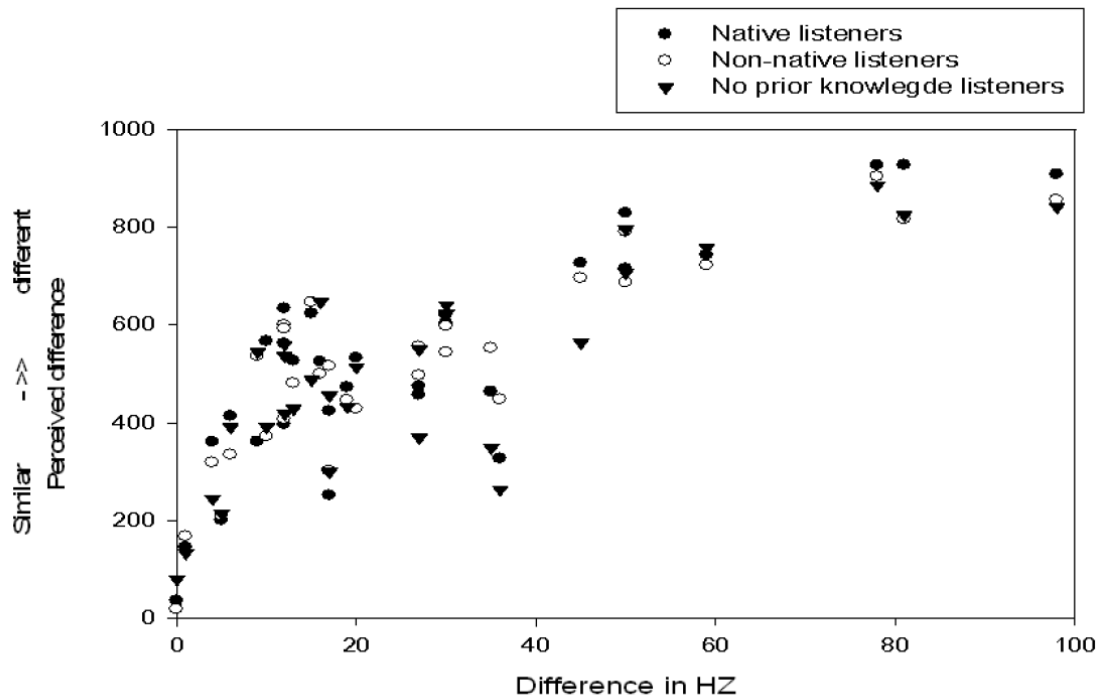


Figure 2. Results showing correlation between the measured difference in F0 between the heard voices and the perceived difference between the voices divided into the three different listener groups.

The results from Figure 2 suggest that there is a general correlation between difference between voices measured in Hz and the perceived similarity between the voices by all the three different listener groups. All three groups show a tendency to judge voices that are quite close measured in Hz to be perceptually similar. Voices that are further apart measured acoustically in Hz are generally also judged to be perceptually more different by all three groups of listeners.

The scatter plot in Figure 3 shows the mean of the first trials of all the listeners compared with the mean of all the listeners' second trial. The low numbers in the figure reflect a small perceived difference between the voice pairs and high numbers reflect a larger perceived difference. The results from the study suggest that the majority of the listeners in all three groups were consistent in their judgements from the first time to the second time they heard the same voice pair – regardless of their level of

knowledge of Danish.

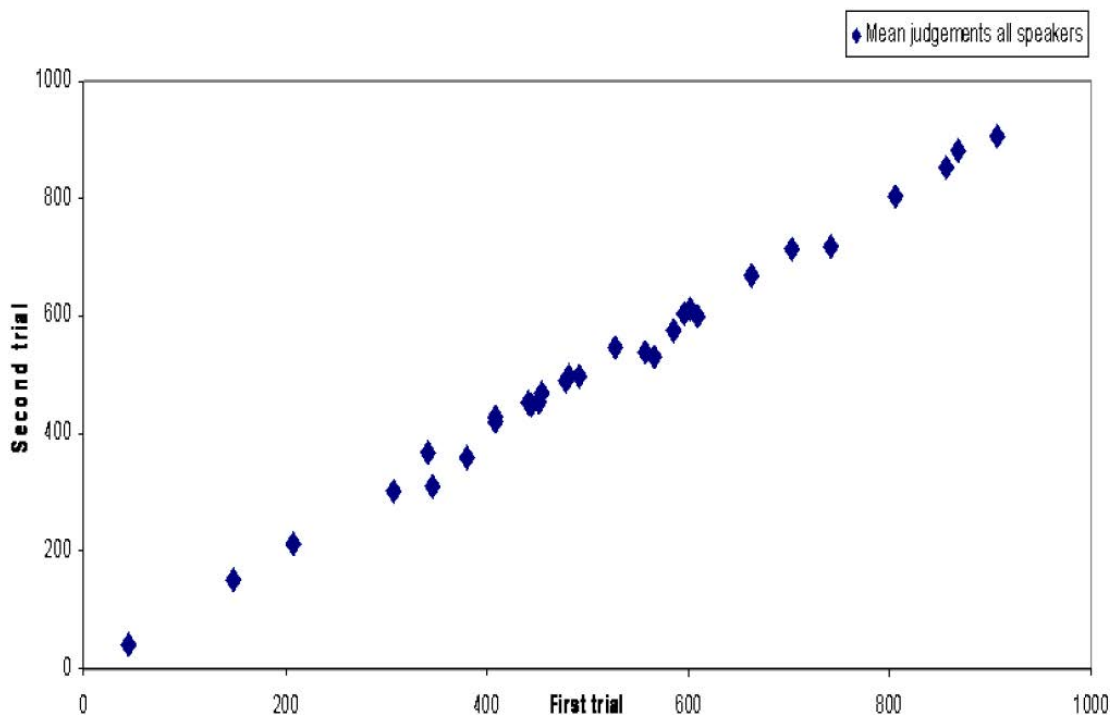


Figure 3. The mean of all the listeners' first trials correlated with the mean of all the listeners' second trial.

Figure 3 shows an almost straight diagonal line through the figure. This suggests that, generally, the listeners are consistent in their judgements from their first to their second trial. This impression was confirmed by correlation analysis (Pearson's  $r$ ) which showed that the correlation coefficient is  $r = .97$  ( $p < .001$ ). In general, there appears to be a strong correlation between the acoustic difference of the mean F0 and the perceived voice similarity.

There are, however, a few exceptions to the trend of a correlation between the acoustic difference of the mean F0 and the perceived voice similarity. Figure 4 shows the results for a single voice pair where the voices were relatively similar according to fundamental frequency. There was only a measured difference of three Hz between the average fundamental frequencies for the two speakers in this sample.

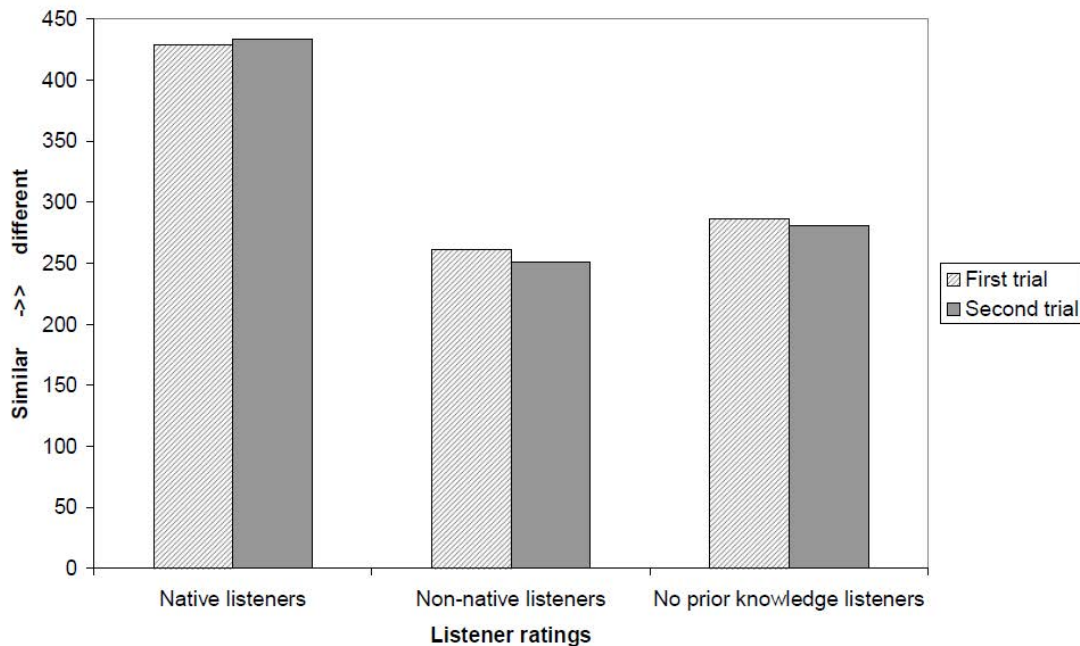


Figure 4. Results from a voice pair where the fundamental frequency is relatively similar, but the speakers speak with different regional accents.

The example in Figure 4 is particularly interesting because the two speakers in this example are from different parts of the country, namely one speaker from Eastern Jutland and the other speaker from Zealand (Copenhagen area). Apparently, the difference in regional accent between the two speakers strongly affects the way that the native listeners judge the voice pair. A one-way ANOVA was run and confirmed the visual interpretation of Figure 4 that the difference between the groups was significant,  $F(2,57)=54.422$ ,  $p=.0001$ . A larger difference was perceived by the native listeners than by the two other groups.

The group with no prior knowledge of Danish would have no prerequisite for what linguistic cues to listen for whereas the native listeners could make use of language specific segmental as well as suprasegmental cues. The results from the present study showed that there were more examples similar to the one in Figure 4. This suggests that there is something in the auditory signal that the native listeners perceive which the two other groups do not when they judge voice similarity. Since the two speakers in the example have similar F0 (only a difference of 3 Hz) a

possible explanation could be that the native listeners are more sensitive to the exact intonation pattern that would be distinct for the two speakers from the two different part of the country. Another explanation could be that native listeners are listening for subtle segmental cues when judging voice similarity after all. In similar examples the results for the non-native listeners were most often closer to the ones of the listeners with no prior knowledge of Danish than they were to the native listeners. This suggests that listeners listen to voices through their L1 filter and possibly not as sensitive to exact intonation patterns or subtle segmental variation in their L2.

The results from the study suggest that, as long as it is a homogenous group of speakers, then the native listeners seem to base their judgement of voice similarity on differences in mean fundamental frequency. However, a difference in regional accent seemed to trump mean fundamental frequency for the native listeners, making some voice pairs perceived to be more different from one another than the other two groups perceived them to be.

#### **4. Discussion**

In general, the listeners seem to judge voice similarity according to fundamental frequency – at least when the voice quality of the speakers are not very distinct, such as e.g. nasal, creaky or hoarse. However, for the native listeners this seemed to be the case only when speakers spoke with the same regional accent. When there was a change in accent, this affected the perceived difference and distance between the voices. Therefore it is important to keep in mind that language specific cues play a role for native listeners, whereas listeners with no prior knowledge of a given language listen in a more global manner and that non-native listeners resemble listeners with no prior knowledge more than they resemble native listeners.

The results from the present study suggest that, in general, as the measured difference between the standard deviation of the two voices in the voice pairs goes up it is also perceived as a bigger difference by the listeners ( $r=.624007$ ,  $p<0.01$ ). The results suggest that there is also some correlation between difference in the mean F0 *variation* and the perceived similarity between the voices by the different listener groups which is in line with suggestions made in previous studies, e.g. Foulkes & Barron (2000). The listeners show a tendency to judge voices with similar standard deviation measured in Hz to be perceptually similar as well. Voices that differ with more F0 variation are generally also judged to be perceptually

more different.

Several studies suggest that – besides fundamental frequency – average formant frequencies over longer stretches of speech also play a part when recognising voices (e.g. Nolan and Grigoras, 2005, Jessen 2008). Even though the first three formants are related to vowel quality produced, and hence have some constraints, there are still individual speaker differences in vowel articulation (Johnson, 2003). Not only are formant frequencies essential correlates of distinctions between different vowels and some consonants, but they also convey important speaker specific information (Jessen, 2008). As formant location depends on vocal tract characteristics, e.g. longer vocal tracts generally lead to lower formant frequencies, it is also possible that the formant frequencies can reveal important speaker specific pathological or habitual features in speech, e.g. a tendency to retract the tongue or a tendency to protrude the lips while speaking. It was beyond the scope of the present study to attempt assessing how this may influence the listeners rating of voices besides fundamental frequency, but there is of course a possibility that this could also be one of the features that the listeners used to decide voice similarity in the present study.

The results suggest that a change in regional accent make the native speakers judge the voice similarity to be more different as well. As mentioned in the introduction there may be different opinions of what constitutes ‘voice quality’ (Köster et al., 2007), hence, also whether some voices are similar or not. Some people could listen for laryngeal characteristics and others could also include articulatory setting as part of their concept of voice quality. In the current study, a change in regional accent caused native listeners to rate the voices to be more different than the other two groups. However, whether the native listeners are listening for specific intonation pattern of the regional accents or whether they are focusing on subtle segmental differences between the accents cannot be determined from the present results. It is still intriguing that a change in regional accent results in a much larger perceived difference between the voices than for other voice pairs with the same difference in F0 between the voices.

Kreiman and Gerratt (2010) also suggest that listeners may have individual listening strategies and that these strategies may be listening for different cues. However, if this was the case in the current voice perception study, much more random results across the listener groups would have been expected. The results from this study suggest that judging



voice similarity is not just a task that is particularly challenging for any of the groups, leading to inconsistent results. The shift in perceived voice similarity appeared instantly and consistently for the native speakers when there was a change in accent whereas the two other groups consistently rated the voices to be more similar in these instances.

## **5. Conclusion**

The aim of the present study was primarily to investigate whether voice similarity is perceived through the filter of the listener's native language like e.g. segments are (e.g. Flege, 1993; Best, 1995). Therefore the study focused on perceived voice similarity between presented voice pairs by different groups of listeners, namely by native listeners, by L2 listeners and by a group of listeners with no prior knowledge of the language.

The study furthermore explored how similar the listeners perceive the voices to be and the results from the study suggest that the majority of listeners use fundamental frequency as a key feature when rating how similar the voices sounded. When the regional accent remained the same, all three listener groups rated voice pairs with similar fundamental frequency to be similar and when there was a larger acoustic difference in fundamental frequency between the voices, the listeners also rated them as very different.

However, a few voices with different regional accent were also among the presented voice pairs in order to explore the affect that a change in accent would have on perceived voice similarity. The two non-native groups still rated voice pairs with similar fundamental frequency to be similar as before. The native group, however, noticed the change in accent and rated the voices as a lot more dissimilar and seems to trump fundamental frequency as the deciding factor when rating voice similarity. It is important to keep in mind that language specific cues play a role for native listeners, whereas listeners with no prior knowledge of a given language listen in a more global manner and that non-native listeners resemble listeners with no prior knowledge more than they resemble native listeners. The results suggest that listeners do actually listen through the filter of their native language – that this is not limited to sound inventory, but also applies when rating voice similarity. The findings from this study could have practical implications for several areas of applied phonetics.

## References

- Best, C. T. (1995). A direct realistic view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-206). Timonium, MD: York Press.
- Best, C. T. & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In: Bohn, Ocke-Schwen and Murray J. Munro (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam: John Benjamins.
- Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, 93(3), 1589-1608.
- Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-272). Timonium, MD: York Press.
- Foulkes, P., & Barron, A. (2000). Telephone speaker recognition amongst members of a close social network. *Forensic linguistics*, 7, 180-198.
- Gayvert, R. T. & Hillenbrand, J. M. (2003). Open-source software for speech perception research. *The Journal of the Acoustical Society of America*, 113(4), 2260-2260.
- Gooskens, C. S. (1997). On the role of prosodic and verbal information in the perception of Dutch and English language varieties. [Sl: sn].
- Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29(4), 383-406.
- Gregersen, F., & Phrao, N. (2016). Lects are perceptually invariant, productively variable: A coherent claim about Danish lects. *Lingua*, 172, 26-44.
- Grønnum, N. (1994). Rhythm, duration and pitch in regional variants of standard Danish. *Acta Linguistica Hafniensia*, 27(1), 189-218.
- Grønnum, N. (2005). *Fonetik og Fonologi*, 3. udg. København: Akademisk Forlag.
- Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass*, 2(4), 671-711.
- Johnson, E. K., Westrek, E., Nazzi, T. & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, 14(5), 1002-1011.
- Kirk, P. L., Ladefoged, J. & Ladefoged, P. (1993). Quantifying acoustic properties of modal, breathy, and creaky vowels in Jalapa Mazatec. *American Indian linguistics and ethnography in honor of Laurence C. Thompson*, 435-450.
- Köster, O., & Schiller, N. O. (1997). Different influences of the native language of a listener on speaker recognition. *Forensic Linguistics. The International*

- Journal of Speech, Language and the Law*, 4(1).
- Köster, O., Jessen, M., Khairi, F., & Eckert, H. (2007, August). Auditory-perceptual identification of voice quality by expert and non-expert listeners. In *Proceedings of the XVI International Congress of the Phonetic Sciences (ICPhS)* (pp. 1845-1848).
- Kreiman, J. & Gerratt, B. R. (2010). Effects of native language on perception of voice quality. *Journal of phonetics*, 38(4), 588-593.
- Kristiansen, T. (2003). Sproglig regionalisering i Danmark?. In *Nordisk dialektologi* (pp. 115-149). Novus forlag.
- Kristiansen, T., Maegaard, M., & Pharao, N. (2011). Det er intonationen, vi hører det på: Perceptionsstudier i genkendelse af moderne dansk med henholdsvis jysk og københavnsk aksang. In *Jysk, Ømål, Rigsdansk Mv*, (pp. 207-224). Peter Skautrup Centret for Jysk Dialektforskning.
- Nolan, F., & Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *International Journal of Speech Language and the Law*, 12(2), 143.
- Sørensen, M. H. (2010). Perception of voice similarity by different groups of listeners, *IAFPA 19th Annual Conference*, Trier, Germany.
- Sørensen, M. H. (2011). *Acoustic and perceptual aspects of speaker-specific differences in speech and their forensic implications*. Aarhus University. Doctoral thesis.
- Sørensen, M. H. (2012). Voice line-ups: speakers' F0 values influence the reliability of voice recognitions. *International Journal of Speech, Language & the Law*, 19(2).
- Wode, H. (1995). Speech perception, language acquisition and linguistic: Some mutual implications. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp.321-350). Timonium, MD: York Press.

# **BETWEEN SOUNDS AND GRAPHEMES**

Handling editor: Mette Hjortshøj Sørensen



## The Four Troublemakers in Danish Orthography

Henrik Jørgensen  
Aarhus University

### Abstract

This paper deals with those aspects of Danish orthography that makes it useless as a guide to actual Danish pronunciation. Danish has a bad reputation among prospective learners for being difficult to pronounce. Certain aspects, like the number of full vowels and the glottal stop, are difficult to cope with, but other aspects are certainly not. Part of the confusion lies in the fact that the orthography – rather than leading the student of Danish towards a proper pronunciation – systematically gives a false impression of what Danes do when they speak. The main areas are 1) the way graphemes associated with the plosives are used, 2) the unsystematic sound-grapheme relation of the short vowels, 3) the problems in deriving the correct length of a full vowel from the writing and 4) the absence of an orthographic equivalence to the glottal stop.

### 1. Pronunciation and spelling of Danish<sup>1</sup>

The theme of this article is to give an introduction to the main problems that arise for anybody trying to use the orthography as a guide to the pronunciation of Danish. Among prospective learners, Danish has a bad reputation; the language is considered difficult to learn and to cause endless trouble for people who try to learn it. Nevertheless some people succeed with the task, among them the person we celebrate with this book. Learning Danish is possible, but definitely not easy.

<sup>1</sup> This paper is based on a lecture given at SDU Odense (DK) and Harvard in 2014. I am grateful for comments and suggestions from the audiences in both places. Also, I wish to express my gratitude to Péter Ács, Budapest, for many years of discussion on related themes, and for his comments and suggestions for this paper.

Learning Danish pronunciation is in itself a difficult task. The pronunciation is characterized by many complicated assimilations, especially around the unstressed syllables (cp. Ács & Jørgensen, 1990; Ács, Fenyvesi & Jørgensen, 2008; Basbøll, 2005, pp. 293-322). The spectrum of consonants also varies in unpredictable ways; the actual inventory of consonant sounds being different in prevocalic and postvocalic positions (Basbøll, 2005, p. 42). Finally, the sheer amount of different full vowel phonemes is considerable; the analysis varies, but recent revisions by, among others, Hans Basbøll (Basbøll & Wagner, 1985; Basbøll, 2005) have brought the number to twelve. Along with this dirty dozen comes a number of combinatory variants, which means that Danish supersedes most other languages by having three different classes of vowels (back, front rounded and front unrounded) and at least five different height levels (Basbøll, 2005, p. 50).

In addition to all these complex factors, the prospective learner of Danish is faced with the fact that Danish orthography is quite complicated. Sometimes the euphemism “deep orthography” is used to characterize the situation; the situation is, put more bluntly, that the orthography has a long tradition, that the occasional changes of the orthography never were meant to deal with problems in the grapheme-to-sound correspondences, and that several aspects of the pronunciation have not been covered by orthography in any way. While such an orthography favours those who have put the best of their childhood years into mastering it, it is unfavourable to those who try to learn the language. In the words of the important Danish phonetician and grammarian Jens Pedersen Høysgaard (1698-1773): “It is no grace to a language to be filled up with rules that are only there to cause trouble to youngsters, to the simpletons and to foreigners who try to learn the language.”<sup>2</sup>

The intention with this paper is to give an outline of the kind of difficulties that Danish orthography causes to learners of the language. The four most important troublemakers are the following themes, where the Danish orthography gives no clue whatsoever to the actual pronunciation:

- 1) The graphemes *ptk* – *bgd* normally associated with the plosives
- 2) The vowel graphemes, esp. those covering short vowels
- 3) Vowel length
- 4) The glottal stop

---

<sup>2</sup> In Høysgaard’s original words: “Thi det er ingen Dyd ved noget Sprog, at det har mange unyttige og unødvendige Observationer at plage Børn, eenfoldige og fremmede med, som vilde legge sig derefter...” (Høysgaard, 1743, p. 207)

## 2. The effects of the four troublemakers

One of the main problems in Danish is the distributional asymmetry of many consonant graphemes. Many graphemes are associated with different sounds, according to whether they occur before or after the vocalic nucleus of the syllable. Part of the reason for this is economy. There is a different set of sounds before and after the vocalic nucleus; thus recycling may be necessary if we want to have a reasonably international alphabet. Another reason is that the sounds have changed their pronunciation in the different contexts after the writing tradition was established. As we shall see, both these factors have been active in shaping the sound/writing interface of Danish. The following discussion relies mainly on Basbøll (2004, 2005), Becker-Christensen (1988), Jervelund (2007), and above all Katlev (1980).

### 2.1. Plosives

Under this heading, I discuss the pronunciations associated with the six graphemes *ptk – bdg*, usually associated with plosive pronunciations in languages that use the Latin alphabet. The point is that these graphemes are associated with many different consonant qualities, not just plosives.

In Danish, the phonemes have standard pronunciations according to their position in the syllable. In principle, a consonant phoneme in Danish has four possible positions; before the vocalic nucleus, we may distinguish between the absolute front position with nothing preceding the consonant (C-) and a secondary position (cC-) with (at least) one other consonant preceding. After the vocalic nucleus, we may distinguish between absolute back position (-C) and a secondary position (-Cc) with (at least) one other consonant following. The following tables show how the graphemes are converted into pronunciation, according to the position in the (written and pronounced) syllable. The tables also includes certain variational phenomena.

Grapheme:	C-	cC-	-Cc	-C
p	[p]	[b]	[b]	[p]/[b] Non-standard: [w]
t	[t]	[d]	[d]	[t]/[d] or [ð]
k	[k]	[g]	[g]	[k]/[g]

Table 1. Plosives 'p', 't' and 'k'



The interpretation of this table is quite straightforward; only grapheme ‘t’ in -C position covers two distinct phonemes, /t/ and /ð/.<sup>3</sup> The actual distribution of the aspirated plosive phonemes is that they only appear in C-position as phonemically distinct from the unaspirated plosives. In all other positions, we find only the unaspirated sounds, although often represented by the ‘aspirated’ grapheme. This is the reason why the ‘unaspirated’ graphemes occur much more sparsely:

Grapheme	C-	cC-	-Cc	-C
b	[b]	--	[b] or [w]	[p]/[b]
d	[d]	--	(marginal)	[t]/[d] or [ð]
g	[g]	--	(marginal)	[k]/[g] or [i]/[w]

Table 2. Plosives ‘b’, ‘d’ and ‘g’

Thus the plosive graphemes must always be interpreted in relation to the position in the syllable. This is not unusual in Danish; it also happens, for instance, with the graphemes ‘r’, ‘v’ and ‘j’. All these three are pronounced as fricatives when prevocalic, but as semivowels when postvocalic, cp. Basbøll, 2005, p. 64. But, whereas this alternation is rather straightforward, the main problem with the plosive graphemes is that the contrast between what the grapheme would normally correspond to, and the actual sound is striking. When graphemes like ‘d’, ‘g’ and partly ‘t’ change from front to back position, they also change in three phonetic categories:

- 1) From punctual to continuous;
- 2) From unvoiced to voiced;
- 3) From contoid to vocoid.

In a hierarchy of sonority, the contrast between plosives and semivowels is considered strong, these two classes being at either end of the consonantal part of the hierarchy. Yet, in Danish, the three graphemes mentioned perform this change, as we shall see.

---

<sup>3</sup> There is a long discussion of the phonemic interpretation of [ð]. Here I have chosen a simple interpretation of this sound as a distinct phoneme, thereby disregarding a long tradition for including it as a positional variant of /d/, going back at least to Rischel 1970 (Consonant Gradation). Ács & Jørgensen (2016) discuss the reasons to give up this analysis.

<i>b</i>	C-	cC-	-Cc	-C
Plosive pronunciation	<i>by</i> (town) <i>bro</i> (bridge) <i>blad</i> (leaf)	--	<i>vable</i> <i>æble</i> (apple) <i>skæbne</i> (fortune) <i>erobre</i> (conquer) <i>krebs</i> (cancer)	<i>køb</i> (acquisition)
Semivocalic pronunciation	--	--	( <i>æble</i> )	( <i>køb</i> ) <i>kobber</i> (copper) <i>peber</i> (pepper)

Table 3. Asymmetry of plosive graphemes – pronunciation of 'b'

The grapheme 'b' presents only few problems. Semivocalic pronunciations (all of them with a [u]) are mostly varieties in casual speech<sup>4</sup>, except *kobber* (copper) and *peber* (pepper), where the semivocalic pronunciation is standard. It is odd, but not unsurmountable that out of the words given here with postvocalic plosive pronunciation, *æble* has a semivocalic variant but the others do not. In this case, a plosive pronunciation will always be acceptable.

<i>d</i>	C	cC-	-Cc	-C
Plosive pronunciation	<i>dør</i> (door) <i>dingle</i> (hang) <i>droppe</i> (drop)	--	--	<i>absurd</i> (absurd) <i>addend</i> (factor) <i>akkord</i> (chord) <i>ard</i> (type of plough) <i>bold</i> (ball)
Semivocalic pronunciation	--	--	-- <sup>5</sup>	<i>mad</i> (food) <i>mod</i> (courage) <i>bid</i> (bit) <i>slud</i> (slush) <i>vold</i> (violence) <i>aldehyd</i> <i>alkaloid</i> <sup>6</sup>

Table 4. Asymmetry of the plosives – pronunciation of 'd'

<sup>4</sup> The use of the semivocalic pronunciation alternative is common, especially among speakers born before 1965 regardless of other social or regional variation.

<sup>5</sup> A few word forms with '-ds' are pronounced [ðs]: *betids* (in due time), *andetsteds* (elsewhere) etc. They all derive from genitive forms of *tid* (time) and *sted* (place), and therefore they are not true examples of this spelling constellation.

<sup>6</sup> Although it looks like loanwords mainly have the [d] pronunciation, this is not handled consistently. Chemical terms often have [ð].

The real complication with this grapheme is the pronunciation in the -C position, where both plosive and semivocalic pronunciation are in play. Another complication is that a written ‘-d’ is not pronounced after ‘l’, ‘n’ and ‘r’ (*fald* (fall), *hold* (grip), *mand* (man), *grund* (ground), *hård* (hard), *mord* (killing)<sup>7</sup> and before ‘s’ and ‘t’ (*Mads* (name), *gods* (goods), *midt* (middle), *blandt* (among)<sup>8</sup>), the so-called “silent d”. In such contexts, the *d* often represents earlier conventions concerning certain now lost sounds that were represented in orthography of those days with a digraph. On the complicated details of these spelling forms and their interaction with the glottal stop, see Jensen 2016.

g	C-	cC-	-Cc	-C
Plosive pronunciation	<i>gave</i> (gift) <i>gåde</i> (riddle) <i>grufuld</i> (terrible)	<i>sgu</i> [swearword]	<i>gigt</i> <i>vægt</i> (weight) <i>hægte</i> (connect) <i>bygd</i> (village) <i>lægd</i> (military roll) <i>slags</i> (kind) <i>sigt</i> (sight) <i>magt</i> (power)	<i>grog</i> (rhum)
Semi-vocalic pronunciation	--	--	<i>smaragd</i> (emerald) <i>snegl</i> (snail) <i>fugl</i> (bird) <i>hagl</i> (hail) <i>tegn</i> (sign) <i>vogn</i> (wagon)	<i>bog</i> (book) <i>lig</i> (dead body; also as a derivative ending) <i>sag</i> (case) <i>borg</i> (castle)

Table 5. Asymmetry of the plosives – pronunciation of ‘g’<sup>9</sup>

This table demonstrates the same overall distribution as with the other two plosive graphemes: the cC- position is only represented by the swearword *sgu*, derived from a longer oath containing *gud* (‘God’), hence the deviant

<sup>7</sup> There are, however, some exceptions: *bold* (ball), *bande* (gang), *hærde* (make-hard), all of them with final [d].

<sup>8</sup> This last rule also applies when ‘t’ is an inflection: *hed* – *hedt* (hot); *sød* – *sødt* (sweet).

<sup>9</sup> This table does not take the use of ‘g’ in digraphs like ‘-ng’ into account.

sg-spelling. Here, the -Cc position is the complicated one, where both possible pronunciations are found. The -C position seems mostly to trigger semivowels; *grog* (rhum) is a loanword. It adds to the complications that the semivocalic pronunciation is dependent on the preceding full vowel: front vowels yield an [i] corresponding to the ‘g’, back vowels a [u].

## 2.2. Short vowels

As has been noted often (Basbøll & Wagner, 1985; Basbøll, 2005; Jervelund, 2007), the sound-to-letter correspondence for long vowels is unmarked in most cases. The short vowels, on the other hand, are complex. Sometimes one grapheme represents two or more phonemes, at other times two graphemes share a phoneme. When both these situations occur, the pronunciation gets rather complicated. With the front unrounded vowels, the two middle ones represent the same phoneme, but the high and the ones represent two:<sup>10</sup>

Grapheme	Sound	Examples
I	/i/	<i>pisk</i> (whip), ( <i>mini-</i> ) <i>Risk</i> (name), <i>mild</i> (mild), <i>sild</i> (herring), <i>vild</i> (wild), <i>skidt</i> (dirt)
	/e/	<i>disk</i> (counter), <i>fisk</i> (fish), <i>pil</i> (arrow), <i>vil</i> (vb. will), <i>midt</i> (middle)
E	/ɛ/	<i>fest</i> (party), <i>hest</i> (horse), <i>bedst</i> (best)
Æ	/ɛ/	<i>læst</i> (shoe tree), <i>næst</i> (next to)
A	/a/	<i>and</i> (duck), <i>hat</i> (hat), <i>fald</i> (fall)
	/ɑ/	<i>Anders</i> (name), <i>kaffe</i> (coffee), <i>kam</i> (comb) and in connection with -r-: <i>kram</i> (hug), <i>skrald</i> (garbage)

Table 6. Sound-grapheme correspondences of the short unrounded front vowels

The two front rounded vowels share the sound /ø/. Quite often, the phonological context gives no clue to the pronunciation (*bytte* vs. *nytte*; *dysse* vs. *kysse*):

<sup>10</sup> The notation of the vowels follows normalized versions in the table in Basbøll, 2005, pp. 45-47. I refrain from giving the strict non-normalized IPA forms in this paper, but they may be found in Basbøll’s table.

Grapheme	Sound	Examples
Y	/y/	<i>bytte</i> (exchange), <i>dytte</i> (honk), <i>hytte</i> (hut), <i>lytte</i> (listen), <i>pyt</i> (puddle), <i>dysse</i> (soothe), <i>hysse</i> (hiss, silence), <i>Sysse</i> (name)
	/ø/	<i>nytte</i> (be of use), <i>spytte</i> (spit), <i>kysse</i> (kiss)
Ø	/ø/	<i>bøtte</i> (bucket)
	/œ/	<i>bønne</i> (bean), <i>stønne</i> (groan)

Table 7: Sound-grapheme correspondences of the short rounded front vowels

The graphemes that represent the back vowels share four sounds in the most inconsistent way:

Grapheme	Sound	Examples
U	/u/	<i>bul</i> (tree-trunk), <i>bulle</i> (official letter), <i>skulle</i> (inf. of 'shall'), <i>tulle</i> (mess around), <i>kulle</i> (bald mountain)
	/ɔ/	<i>kul</i> (coal), <i>hul</i> (hole), <i>nul</i> (zero), <i>(for-)kulle</i> (turn into coal), <i>(gennem-)hulle</i> (to get filled with holes)
O	/o/	<i>mor</i> (mother), <i>foto</i> (photo)
	/ɔ/	<i>bombe</i> (bomb), <i>plombe</i> (dental filling)
	/ʌ/	<i>rombe</i> (rhomb), <i>hekatombe</i> (hekatombe)
Å	(/ɒ/)	<i>tårn</i> (tower), <i>år</i> (year), <i>hår</i> (hair) <sup>11</sup>
	/ʌ/	<i>bånd</i> (ribbon), <i>hånd</i> (hand)

Table 8: Sound-grapheme correspondences of the short back vowels

The situation concerning the short vowels can only be characterized as a complete mess. Due to sound changes and etymology, conventional spellings with a very inconsequent relation to the actual pronunciation prevail and leave the learner with almost no clue at all of what to do.

<sup>11</sup> This grapheme-sound correspondence is only relevant if the phonemic analysis has to catalyse an /-r/ in this position. Otherwise, these examples are simply long vowels, manifestation of a fourth back vowel phoneme /ɒ/, almost always corresponding to a digraph 'år'. This phoneme makes perfect commutations with /ɔ/: *å* (river) – *år* (year); *lå* (past tense of *ligge*) – *lår* (thigh). Although certain cases of this commutation apparently rely on grammatical relations, like *få-får* (infinitive and present of 'to get') or *gå-går* (same forms of 'to go'), commutations like *å-år* and *lå-lår* cannot be reduced to grammar in this way.

This problem, combined with the fact that the orthography only makes a dim distinction between long and short vowels (see sect. 2.3.), makes the orthography of the vowels almost impossible to use when actual pronunciation is attempted.

### 2.3. Vowel length

Vowel length is an important feature in Danish pronunciation; yet the orthography does very little to make clear when a vowel is long, and when it is short. Still, there are some main rules, which Becker-Christensen, 1988, p. 87 gives as a two-way system:

I: In syllables ending in a vowel and syllables with one postvocalic consonant: the vowel is LONG.

II: In syllables with two post-vocalic consonants: the vowel is SHORT

Before we may use this rule, there are some reservations. This rule applies only to monosyllabic words and words ending in a stressed syllable. Furthermore, the rule may first be applied when all inflections and derivatives have been removed. This makes life more complicated for learners, since one has to know the details of morphology in order to apply the rule.

However, this is not all there is to it. We find a number of exceptions to both rules, which makes the picture even more opaque. Exceptions to rule I, e.g. short vowel in VC-structures without glottal stop (cf. Becker-Christensen, 1988, p. 92 & 213):

- In front of plosives: *hat* (hat), *nok* (enough), *kat* (cat), *gok* (a stroke, blow), *tit* (often), *flok* (flock), *klik* (click), *smuk* (beautiful), *flot* (impressive), *glat* (even), *at* (that), *sat* (form of vb. 'to sit')
- In front of nasals: *han* (he), *hun* (she), *man* (pron. 'one'), *som* ('that' as relative)
- In front of semi-vowels: *og* (and), *jeg* (I), *dig* (obl. form of sg. 'you'), *sig* (refl. pronoun), *er* (is), *var* (was), *rav* (amber), *drev* (drive (IT)), *rev* (riff), *jer* (obl. form of pl. 'you'), *vor* (our)
- In front of [ð]: *glad* (happy), *mad* (food), *had* (hatred), *gud* (God), *bed* (bed of flowers), *fred* (peace)

Pronouns and other function words are well represented in this group: *og, jeg, dig, sig, er, var, at, som, det* (it), *sit* (reflexive-possessive), *jer* and *vor*<sup>12</sup>. The main reason seems to be that the prosody of the standard pronunciation has changed drastically after the first establishment of the writing tradition.

Furthermore, there are a number of exceptions to rule II, cp. Becker-Christensen (1988, p. 91):

- The vowel is long before *-rd* and *-ds*, and before *gC*: *Bord* (table), *kreds* (circle), *ligne* (look like), *fugl* (bird), *flegma* (phlegma), and as exceptions *karl* (farmhand), *vejr* (weather)
- The vowel is long before certain double consonants: *næbbet* (beak-the), *læggen* (thigh-the), *skægget* (beard-the), *ægget* (egg-the), *sjette* (sixth), *otte* (eight), *ætten* (family-the), *bredde* (broadness), *vidde* (width).
- The vowel is long before certain combinations of graphemes: *vable*, *æble* (apple), *skæbne* (fortune), *væbne* (vb. 'arm'), *erobre* (conquer), *sagte* (silent, soft-spoken), *ens* (identical), *besk* (bitter), *slesk* (wheedling), *træsk* (wily), *påske* (easter), *bæst* (animal, unpleasant person), *faste* (lent), *kiste* (coffin), *hoste* (cough), *pruste* (snort), *puste* (blow; the two last ones may be both long and short).

Inflectional forms are the reason for a number of (apparent) exceptions to rule II (short vowel when followed by two consonants). In many cases, one spelling form has two pronunciations, one following rule II and therefore short, one of them inflected, and therefore following rule I after subtraction of the ending *-t*:

- *Mast*: as a substantive ('mast') short, but as a verb ('mase', press) long
- *Læst*: as a substantive ('shoe tree') short, but as a verb ('læse', read) long
- *Lyst*: as a substantive ('pleasure') short, but as a verb ('lyse', give light) long
- *Kyst*: as a substantive ('coast') short, but as a verb ('kyse', to scare) long
- *Øst*: as a substantive ('east') short, but as a verb ('øse', to pour) long

In the central part of the vocabulary, such exceptions like e.g. short vowel in syllables ending in a vowel are quite frequent (Becker-Christensen,

<sup>12</sup> But not *den* (it) with glottal stop and *sin* (reflexive-possessive) with a long vowel (and glottal stop).

1988, p. 93):

- Many personal pronouns: *du* ('you' sg.), *vi* ('we'), *I* ('you' pl.), *de* (*De*) ('they' & 'you' polite sg. & pl.)
- Many interjections: *ja* ('yes'), *ha* ('ha'), *hurra* ('hurray'), *fy* (introducing a reproach), *nå* (expression of attention or doubt), *oho*, *hallo*
- Adverbs, conjunctions etc.<sup>13</sup>: *nu*, *da*, *så*, *thi*, *jo*

Many loan words generate exceptions as well:

- The solmisation *do*, *re*, *mi*, *fa*, *la*
- Loan words from French: *cha-cha-cha*, *gaga*, *charpi*, *fait accompli*, *hotel garni*, *kepi*, *maki* (including the French-inspired pronunciation of the capital of Finland, *HelsinKI*), *art deco*, *yoyo*, *vue/vy*, *revy*, (*portemonnæ*, *adjø*,) *miljø*.<sup>14</sup>

While many of these exceptions are marginal, several of the others deal with the core vocabulary. Together with the principle that the rules of prosodic interpretation of vowel graphemes do not apply until the stem has been stripped off its morphology, it is fair to conclude that the prosodic character of the vowels is almost inscrutable from orthography in Danish.

#### 2.4. The orthography and the glottal stop

According to the most recent and most comprehensive theory on the glottal stop, this prosodic feature is distributed according to the weight of the syllable, cp. Basbøll (1988, 1998, 2005). Therefore, the other prosodic features (vocalic length combined with certain voiced postvocalic consonants) determine where the glottal stop may occur. In principle, the glottal stop only occurs in the ultimate syllable of a stem; if there is an unstressed final syllable, usually no glottal stop occurs. However, the orthography gives no clues to this at all. No constellation of letters signals the glottal stop in any consistent way (Basbøll, 2005, p. 90).

The fact that the glottal stop is concomitant with other factors is probably the main reason why this phenomenon never attracted the interest

<sup>13</sup> Since most of these words do not correspond to similar words in English, no translation is given.

<sup>14</sup> This may be due to pronunciation habits created by "informed" speakers. The now obsolete and not quite polite *pø om pø* (fr. *peu en peu*) has a long vowel with a glottal stop, just as expected from a final vowel.



of orthographers (except Høysgaard in the 18th century). However, for a learner, the rules needed to identify the position of the glottal stop are so complex that they are hardly worth applying in teaching (except when teaching linguists). Thus, the absence of a spelling convention creates serious challenges for learners.

If Jutland had remained the core area of the kingdom (as it was at the dawn of Danish history, the capital being Jelling in Southern Jutland), things would have had to take a different course. All dialects in Jutland have APOCOPE, i.e. unstressed final syllables have been lost. Due to this sound law, Old Norse monosyllabic stems (later with glottal stop) and bisyllabic stems (later without glottal stop) form one monosyllabic group in this dialect comprising most of the current vocabulary of the language. However, the contact backwards in the Jutland dialects is intact; the old monosyllabic words retain the glottal stop, and the old bisyllabic words did not acquire it. Only the glottal stop will keep the two groups distinct and therefore any orthographic system for Jutland dialects will have to find a way of signalling the glottal stop; otherwise essential information is lost. Since, however, the standard orthography was based on the pronunciation in Sealand dialects, where no systematic apocope is found, this problem did not arise in Standard Danish.

Loss of final schwa is now spreading into Standard Danish (Brink & Lund, 1974, pp. 195-7), thus facing also the non-jutlanders with this sound-grapheme interface problem. Furthermore, most modern monosyllabic loan words from English (*boom*, *cool*, *cruise*) cannot be accommodated to modern Danish orthography. For theoretical reasons, the attempts at spelling conventions for Jutland dialects are of interest; they might provide us with a solution to a problem that will become more and more relevant due to the strong influx of English loan words.

Viggo Sørensen (2007, p. 54) gives an analysis of the situation in Standard Danish compared to Jutland dialects. He identifies three monosyllabic types in Standard Danish:

- Words with vocalic glottal stop (*bro* (bridge), *sne* (snow), *gry* (dawn), *fad* (tray), *nøl* (tarrying))
- Words with consonantal glottal stop (*land* (land), *rend* (mass), *vom* (stomach))
- Words without a glottal stop (*hat* (hat), *sæt* (set), *blot* (only), *rat* (steering wheel))

As we have seen, Danish orthography leaves no clues as to which of these types the learner is faced with, apart from the spelling conventions mentioned above. It is remarkable that the orthographic constellation (-)VC is represented in all three prosodic word types.

In the Jutland dialects on the other hand, Sørensen (2007, p. 57) identifies seven syllabic types:

1. Words with a tonal accent (only relevant in certain Southern Jutland dialects)
2. Words with vocalic glottal stop
3. Words with consonantal glottal stop
4. Words without a glottal stop
5. Words with a long vowel without glottal stop
6. Words with long consonant
7. Words with West Jutland glottalization

This is a general matrix. Only a few dialects in Southern Jutland have type 1, and the use of 7 is also restricted to parts of Western Jutland. However, types 2-6 are present all over the peninsula.

In Jutland, presence and absence of glottal stop<sup>15</sup> are the only distinctions of singular vs. plural with many monosyllabic words:

Hus (house)	[huʔs] – [hu:s]
Ben (leg)	[bieʔn] – [bi:en]
Gren (branch)	[græʔn] – [græ:n:]
Bro (bridge)	[broʔw] – [bro:w:]

Therefore, most Jutland dialects have a syllabic type unknown to Standard Danish until the monosyllabic English loan words arrived: monosyllables with a long vowel without glottal stop. The plural forms above display this.

In general, there has been little interest in a special orthography for Jutland dialects, even though the Standard Danish language taught in the schools must have seemed strange to small children in Jutland 200 years ago. Most practitioners of a special Jutland orthography were authors using dialect either in quotes or in whole narratives. The best known of them, poet

---

<sup>15</sup> Or similar prosodic oppositions, like the tonemic patterns in certain Southern Danish dialects, cf. <https://dialekt.ku.dk/dialektkort/>.

and vicar, Steen Steensen Blicher (1782-1848), devised an orthography for his short stories in dialect, mostly written in Central Jutland dialects. Here follows an overview of Blicher's method of rendering the prosody in orthography (after Sørensen, 2007, p. 55; 'GS' is short for 'glottal stop'):

- |  |   |
|--|---|
| 1. Short V +/- C :                       | <i>no special marking</i>   |
| 2. Long V + GS +/- C :                   | <i>Ve</i> (bar (carried) = baer; dør (door) = døer)                     |
| 3. Short V + C + GS:                     | <i>VCh</i> (vild (wild) = vilh; nem (easy) = nemh)                      |
| 4. Long V +/- C :                        | <i>Vh(C)</i> (plade (plate) = plahd; tørre (to dry) = tahr)<br>= löwnn) |
| 5. Short V + Long C :                    | <i>VCC</i> (nar (fool) = narr, levne (leave behind) = löwnn)            |
| 6. Short V + C + West Jutland GS: like 5 |   |

Certain aspects of Blicher's orthography are inconsistent. It seems irregular to signal the glottal stop in vowels with 'e' (2), but with 'h' with consonants (3), particularly when 'h' with vowels signify length (4). Many of Blicher's orthographic devices echo from orthography in early modern times (1500-1700, partly also older), where many similar spelling variants are found. In all likelihood, they were not used in a consistent way even then; what Blicher did, was to take unsystematic occurrences and give them a consistent meaning.

For Jutland dialect speakers, his orthography was, and may still be, intuitively useful. Whether an attempt at a modern spelling reform designed to eliminate the troublemakers would find Blicher's way of handling things useful, is quite another matter. However, he actually managed to solve the prosodic problems in a way that might be useful in a future orthographic reform.

### 3. Conclusion

The four problem areas that I have tried to identify in this paper, the plosive graphemes, the short vowels, the vowel length, and the glottal stop, are the main reasons why orthography is a very bad guide to Danish pronunciation. Hopefully, this overview serves to demonstrate how complicated the situation is. What Danes write, has very little to do with what they say.

Sometimes, the problem areas are language-internal matters. In almost all languages, short inherently unstressed words will often have a well-established orthographic form, and it is quite unlikely that it converges with the main tendencies in the sound/writing interface.

Other problems arise from loan words. German and French loan words are normally quite well integrated in present-day Danish orthography. The English loan words, on the other side, present unsurmountable problems to the orthographic system and at present no attempt is made to integrate them at all.

Obviously a regulation, especially of the prosodic form, is tempting – in theory. In reality, things look different. A thorough reform will change the orthography to a degree where contact with other Scandinavian languages and older written matters will become almost impossible.

My experience with Danish students is that they find it extremely difficult to distinguish prosodic features, although paradoxically they must perceive the effects of them. There are variational phenomena and developments underway; thus, there is no truly consistent norm to codify.

Therefore, my best guess is that nothing will happen with the Danish orthography, in spite of the state of affairs. If instructors want novice learners of Danish to sound Danish, they will still have to teach them pronunciation by the ear without the aid of a textbook. Furthermore, in the future written Danish will be utterly misleading when it comes to actual pronunciation. A spelling reform would probably make Danish less frustrating for foreigners, but due to the distance between spelling and pronunciation cause problems of other kinds. Høysgaard's negative judgment on complicated languages will continue to apply to Danish for a long time to come.

## References

- Ács, P. & Jørgensen, H. (1990). På afgrundens rand? - Nogle bemærkninger om konsekvenserne af en morfofonologisk udviklingsproces i moderne dansk. In I. Sooman (Ed.): *Vänbok till Otto Gschwantler*, (pp. 1-11). Wien: Verband der wissenschaftlichen Gesellschaften in Österreich.
- Ács, P. Fenyvesi, K. & Jørgensen, H. (2008). Dansk fonologi og morfologi set i lyset af de såkaldte naturlighedsteorier. *Nydanske studier*, 36, 10-37. <http://dx.doi.org/10.7146/nys.v36i36.13466>
- Ács, P. & Jørgensen, H. (2016). Hvorfor er dansk vanskeligt? Danske konsonantsegmenters to ansigter. *Skandinavisk Füzetek*, 10, 89-100.
- Basbøll, H. (1988). Mellem moræ og fonologi: nyt om stødet i moderne rigsdansk. *MUDS 2*, Aarhus Universitet, 37-48.
- Basbøll, H. (1998). Nyt om stødet i moderne rigsdansk – om samspillet mellem lydstruktur og ordgrammatik. *Danske Studier* 1998, 33-86.

- Basbøll, H. (2004). Et klassifikationssystem for stavemåder. P.S. Jørgensen & H. Jørgensen (Eds.), *På godt dansk. Festskrift til Henrik Galberg Jacobsen* (pp. 29-36). Århus: Wessel og Huitfeldt.
- Basbøll, H. (2005). *The Phonology of Danish*. Oxford: Oxford University Press.
- Basbøll, H. & Wagner, J. (1985). *Kontrastive Phonologie des Deutschen und Dänischen*. Tübingen: Max Niemeyer Verlag.
- Becker-Christensen, C. (1988). *Bogstav og lyd, bd. 1*. København: Gyldendal.
- Brink, L. & Lund, J. (1975). *Dansk rigsmål 1-2*. København: Gyldendal.
- Høysgaard, J.P. (1743). *Tres faciunt collegium*. København: Berling. Repr. in H. Bertelsen (ed.) (1979), *Danske Grammatikere* (pp. 187-215). København: DSL & C. A. Reitzel.
- Jensen, E. S. (2016). Den der skriver *d* i *mand*. *Nyt fra Sprognævnet* 1/2016, 1-4.
- Jervelund, A. A. (2007). *Sådan staver vi*. København: Dansk lærerforeningens Forlag & Dansk Sprognævn
- Katlev, J. (1980). Diverse lingvistiske parametre i retskrivningsspørgsmålet (pp. 173-200). *SAML* 6.
- Rischel, J. (1970). Consonant gradation: A problem in Danish phonology and morphology. In H. Benediktsson (Ed.), *The Nordic Language and Modern Linguistics* (pp. 460-480). Reykjavik: Vísindafélag Íslendinga.
- Sørensen, V. (2007). Når forfatterne skriver jysk – og hvad det fortæller om skriftsprog og lydskrift. *Ord og Sag*, 27, 44-61.

# Northumbrian Rounded Vowels in the Old English Gloss to the Lindisfarne Gospels

Johanna Wood  
Aarhus University

## Abstract

This paper<sup>1</sup> investigates the distribution of mid-front rounded vowels in the Northumbrian glosses to the Lindisfarne gospels. Rounding after /w/ is a dialect feature of late Old Northumbrian. Numerical counts for the distribution of the feature are merged with new data. The goal is to see whether the data support already hypothesized demarcations in the text. The main finding is that the gospel of Luke and the second half of Mark have the most frequent occurrences of this feature and therefore are the most conservative sections of the glosses.

## 1. Introduction

The debate regarding the authorship of the Old English Gloss to the Lindisfarne Gospels has maintained a continued presence in academic literature for at least 150 years. This paper contributes to that debate by further investigating the distribution of mid-front rounded vowels throughout the four gospels.

The Lindisfarne bible in Latin was written at Lindisfarne Priory on Holy Island and ascribed to the monk Eadfrith, who was Bishop of Lindisfarne between 698 and 721. The Lindisfarne community, after being

<sup>1</sup> Many thanks to Elly van Gelderen, Sten Vikner, and the participants in the Workshop on the Old English Glosses to the Lindisfarne Gospels (Arizona State University, May 26-27, 2017) for helpful comments. Thank you to Ocke Bohn for his cheerful collegiality and for motivating my interest in Old English vowel variants.

forced to flee to the mainland with their remaining treasures and relics due to Viking raids, eventually settled at Chester-le-Street, Durham where, around 970, interlinear glosses in the Northumbrian dialect were added to the Latin bible. The glosses are generally attributed to the priest, and later provost, Aldred, and the gospels include a colophon he wrote, describing his part in the enterprise, but even then it is a matter of dispute as to whether he is claiming authorship of the entire work or just the Gospel of John.

As early as 1857, K. W. Boutererk writes, “In der Glosse selbst sind mit Bestimmtheit zwei Hände zu unterscheiden” [In the gloss itself two hands may be differentiated with certainty] (as cited by Brunner 1947-8, p. 32, translation my own). This “multiple author” view is countered by paleographic evidence, most notably Ross et al. (1960), that claims the gloss to be the work of only one hand, in which case differences in ink color and grapheme size and spacing are attributed to the writing having taken place over different time periods and under different conditions. Nevertheless, the uneven distribution of linguistic features throughout is puzzling, given the “one author” view, and the gospel has provided substantial material for ongoing investigations. A prevalent view is that there was one glossator who drew on multiple exemplars, and in that way introduced the variant features. However, recently Cole (2016) reignites the debate in noting that “the commonalities between the linguistic and paleographical demarcations could indicate that the involvement of other hands in writing the gloss remains a possibility” (Cole, 2016, p. 187).

Although the paleographical evidence points to only one hand, Ross et al (1960) suggest a division into two main parts with a transition at ff. 203r–203v, that is, at the end of Luke. Evidence for this split cites the neat and compact script that follows, in contrast with the untidiness of that preceding the end of Luke; also notable is that here orthographic <u> is replaced by a more pointed form, <v> (Ross et al, 1960, p. 23), and this is used for ‘w’ instead of the runic letter form, *wynn*. (<ƿ>). However, demarcations based on orthographic evidence are not the only ones found. A series of investigations into linguistic features has established demarcations at various places, but demarcations that do not follow the gospel divisions. The feature that is the focus of this paper, mid-front rounded vowels, was noted in the first systematic attempt to investigate the distribution of features (Brunner, 1947) which looked at the spelling of the stems of the verbs, *wesan* ‘be’ and *cuepan* ‘say’. These vary between *wōēr-* or *wēr-* and *cuoep-* or *cuep-*, the <oe> spelling representing a rounded vowel. The observation is that forms of *wesan* in <oe> are rare in Matthew

and the first five chapters of Mark, and approximately equally frequent for the rest of the gospels. Forms of *cueþan* in <e> are comparatively rare after the first four chapters of Mark, but become frequent again in the first three chapters of John. Thus, there seems to be an uneven distribution of rounded vowels throughout the gospels, though the descriptive distribution is a little different when *wesan* and *cueþan* are compared. Two other observations in the same study, first that the accusative singular feminine form *ðyu* (as opposed to *ðiu*) does not occur after MkGl (Li) 5,32, and also that *heonu*, ‘behold’ often used to gloss Latin *ecce*, does not occur after MkGl (Li) 3,34 (although the alternative *heono* occurs throughout the glosses) lead to the overall conclusion that there is at least one break, at MkGl (Li) 5,40 and possibly others (Brunner, 1947, p. 35).

This observation regarding *wesan* and *cueþan* is regularly quoted but, as far as I am aware, has not been extended to other lexical items that show similar conditioned vowel rounding. Also, the raw figures quoted by Brunner (1947) are somewhat difficult to interpret and compare without recourse to the same 64 approximately equal divisions she constructed. Later work tends to identify possible section demarcations by chapter and verse number. Therefore, in this paper I take as a starting point already established demarcations, in addition to that at MkGl (Li) 5,40, that have been hypothesized in the subsequent literature. First, I rework the raw numbers for *wesan* and *cueþan* from Brunner (1947, p. 51) into percentages for the established sections. Next, I select other lexical items that show the same vowel rounding and investigate their distribution throughout the four gospels. Finally, I put my own data together with Brunner’s to find the overall picture. Section 2 discusses the Northumbrian dialect, the vowel system, and the variant features found in the Lindisfarne gospels which have been used for suggested demarcations. Section 3 reports the method and results, and section 4 is a conclusion. The purpose is to find whether there is uneven distribution of mid-front rounded and unrounded vowels and whether or not these examples support the already established demarcations in the gospels.

## 2. Variation in Old English: Northumbrian dialect

For convenience, the dialects of Old English are traditionally divided into four distinct areas that mirror the political structures of the time: Northumbrian, Mercian, West Saxon and Kentish, as shown in Figure 1. Early Northumbrian is represented by, for example, *Cædmon’s Hymn* and runic inscriptions on the Franks Casket and the Ruthwell Cross. The most



significant examples of Late Old Northumbrian are the interlinear glosses to the *Lindisfarne Gospels* (London, British Library, MS Nero D.iv), the *Durham Ritual*, (Durham, Cathedral Library, MS A.iv.19) and the parts of the *Rushworth Gospels* known as Ru<sup>2</sup> (Oxford, Bodleian Library, MS Auct. D.2.19) (Cuesta & Pons-Sanz, 2016, p. 1).

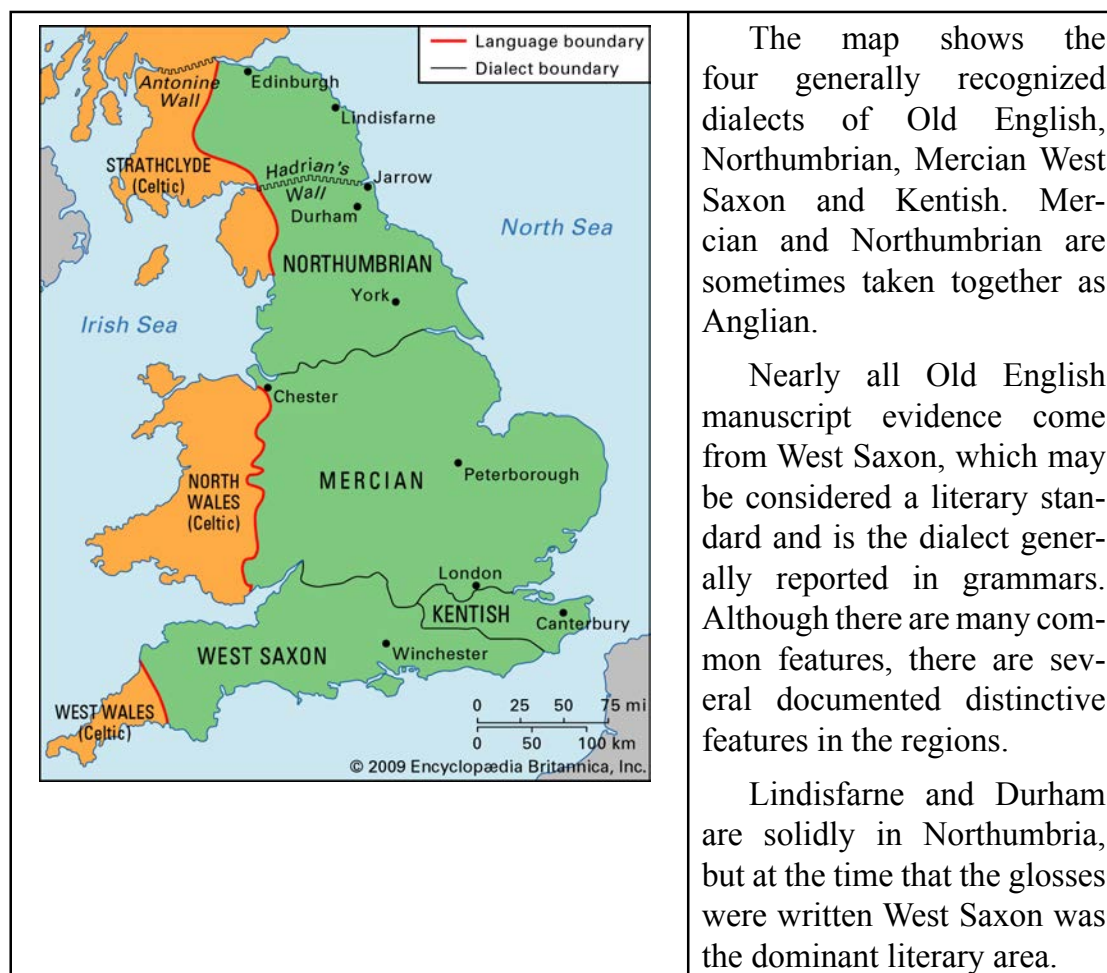


Figure 1. Old English dialect areas

A number of phonological, morphological, lexical and syntactic characteristics of Northumbrian have been identified, some exclusively Northumbrian, some more generally “northern” or Anglian, distinguishing these varieties from West Saxon. They include contracted negative verbs and adverbs (Levin, 1958; Wood, 2002; Van Bergen, 2008), Scandinavian vocabulary (Pons-Sanz, 2000), extensive use of 3<sup>rd</sup> person singular ‘s’ (Blakely, 1949; Cole, 2014, 2016), periphrastic and inflected genitive (Ledesma, 2016), and general early loss of inflectional morphology on nouns and verbs. (See also Cuesta et al, 2008, for a summary of features).

It is well known that English dialects show considerable variation in vowels, and somewhat less variation in consonants. I will assume that for the vowels in Old English, there is a direct correlation between grapheme and phoneme, as demonstrated by (Hogg, 1992, p. 85). Evidence for the sound system of Old English comes from investigations of contemporary scribal practice, and descriptions of the sound system rely on criteria such as “the later history of English, linguistic plausibility, etc.” (Hogg 2011, p. 12). The phonological contrasts of the Old English vowel system are: height, (high, mid and low), backness (front and back) and lip-rounding. Additionally, there is a quantitative contrast in vowel length, although this is not generally indicated orthographically in historical documents (Hogg, 1992, p. 85). (The tradition of modern editors, that I will adopt, indicates vowel length with a macron for long vowels). Rounding was contrastive only for the non-low front vowels.

In the 10<sup>th</sup> century, when the glosses were written, the mid-front rounded vowel that is the main focus of this paper, /ø(:)/, is present only in Northumbrian. “In W(est) S(axon) and K(en)t especially, /ø/ and /ø:/ are unrounded to /e/ and /e:/, and unrounding can be seen to a limited extent in Anglian) also. . . . In E(arly) W(est) S(axon) /ø/ remains only in *oele* alongside more frequent *ele*, [oil], and /ø:/ remains only in *ōēþel* alongside more frequent *ēþel*, [fatherland]. In L(ate) W(est) S(axon) only unrounded forms are to be found. Thus in these dialects we may assume that the unrounding was virtually complete by the time of the earliest texts” (Hogg, 2011, p. 121). Unlike West Saxon, these rounded vowels were still present in late Old Northumbrian texts.

The conditioning that favours vowel rounding is a preceding /w/; this is the result of the transfer of the rounding feature inherent in /w/ to the following vowel. So, one particular feature of later northern Northumbrian is a tendency to round /e/ and /e:/ to /ø/ and ø:/ after the back approximant /w/. The change is dated between *c.*800 and *c.*950 and is not present in early Northumbrian, as evidenced by, for example, Caedmon’s Hymn which has *uerc*<sup>2</sup> ‘work’; however the rounded vowels are frequent in Lindisfarne, represented orthographically by <oe>. Typical examples are *woer* ‘man’; *wōēron* ‘be’ (past ind.pl); *woeg* ‘way’; *woerc* ‘work’; *cwoeða* ‘say’; *cwōēdon* ‘they said’, *swoefen* ‘dream’; *twoelf* ‘twelve’; *woenda*

<sup>2</sup> The back approximant was sometimes represented orthographically by <u> or <uu> but more commonly by the runic symbol *wynn* (þ) for which editors usually substitute <w> to avoid confusion with the runic symbol *thorn* (þ).

‘go’; *wōēde* ‘garment’; *wōēpen* ‘weapon’; *hwōēr* ‘where’; *twōēge* ‘two’; *wōē* ‘we’ (Hogg, 2011, pp. 199-202). These contrast with orthographic <e> which, in all dialects, represents a mid-front unrounded vowel, both short and long (Hogg, 2011, p. 13). Hence, in West Saxon, *wer* and in Northumbrian *woer*; ‘man’ and so on. Although a scribe trained in the north would be expected to use the <oe> spelling, someone with a less conservative pronunciation, or influenced by West Saxon would have a tendency to use <e> in the words listed above, except with the verb ‘be’ which is a special case, as explained below.

A few remarks are in order for the verb ‘be’. In ‘be’, the West Saxon past tense uses orthographic <æ>, which in all dialects represents a low front unrounded vowel, both short and long, normally transcribed as /æ/ and /æ:/ (Hogg, 2011, p. 14), whereas <e> and <oe> variants are found in Lindisfarne. Differences between West Saxon and the other varieties are due to another sound change, raising of low front unrounded long vowels which took place in Anglian but not West Saxon. In all varieties of OE West Germanic [ɑ:] fronted to [æ:]. In West Saxon it remained as [æ:] but underwent subsequent raising to [e:] in Anglian (Mercian and Northumbrian). (Moore & Marckwardt, 1969, p. 25; Jones, 1989, p. 11). The lower pronunciation with [æ:] is typical of West Saxon and [e:] is typical of the north<sup>3</sup> as evidenced by variant spellings, for example, *englas/ængles* ‘angels’. This change only affected long vowels, hence *wæs* in the singular, but *weron* in the plural. Table 1 shows spelling variants for *be* found in the gospels.

Indicative sing.	Indicative pl.	Subjunctive sing.	Subjunctive pl.
wæs, wæss, uæs	weron woeron, uoeron, ueron,	were, wære woere, uere, uoere	uoere

Table 1. Alternative spelling of *was/were* in Lindisfarne

Note from Table 1 there is a need to distinguish orthography that merely represents conventional practice and not variant pronunciation, i.e. the choice between the rune *wynn* (ƿ) (represented by modern editors as <w>) and <u> depends on scribal practice, and the pronunciation is the same, as opposed to the spelling variation of the vowels, which usually represents pronunciation differences.

<sup>3</sup> A separate change in Kentish, the “Kentish Collapse” had a similar outcome, raising of [æ:] to [e:].

Orthographic variation as well as linguistic variation on all levels has been used to identify possible section breaks in the gospels, and the hypothesized breaks used in this study are shown in Table 2. For comparison purposes with studies that reference results by gospel name, note that section 1 could be considered most of Mathew, section 2 half of Mark, section 3 the second half of Mark, section 4 most of Luke and section 5 most of John.

Section	Sec 1	Sec 2	Sec 3	Sec 4	Sec 5
Chapters	- MtG1 (Li) 26,16	MtG1 (Li) 26,17 - MkG1 (Li) 5,40	MkG1 (Li) 5,41 - LkG1 (Li) 2,9	LkG1 (Li) 2,10 - JnG1 (Li) 3,13	JnG1 (Li) 3,14-end.
Skeat (1871-1887) pages	-Mt.215	Mt.215-Mk.41	Mk.41-Luke 29	Luke 29-John 54	John 54-

Table 2. Suggested section breaks taken from previous research

As mentioned above, Brunner (1947) identified a definite break in the 5<sup>th</sup> chapter of Mark and a possible one near the start of John. This is substantiated by Blakeley (1949, p. 91-94) who, in his investigation of the distribution of *-s* and *-ð* verbal endings, finds convincing support for dividing the text into four blocks with divisions at Mt. 26,16; Mk. 5,40 and Luke preface 2,9. The break at Mt. 26,17 is justified based on the lower proportion of verbal *-s* endings compared with *-ð* endings, as *-s* is more frequent in section 1 and also frequent in section 3 and somewhat less frequent in sections 4 and 5. In a more detailed study of *-s* endings, Cole (2016, p. 181) finds an increase in the rate of *-s* in section 3 followed by a drop in Luke which increases again from JnG1 (Li) 3.14 onwards.

The figures reported in Cole (2016, p. 181), which, she reports, are statistically significant, are reproduced in Table 3. As is well known, the *-s* verbal ending spread from the north to the other areas and can be considered the less conservative morphology.

	Sec 1	Sec 2	Sec 3	Sec 4	Sec 5
N (total)	975	194	318	947	619
N <i>-s</i>	794	55	185	209	261
% <i>-s</i>	<b>81%</b>	<b>28%</b>	<b>58%</b>	<b>22%</b>	<b>42%</b>

Table 3. % *-s* endings from Cole (2016, p. 181)

This extra division of John as a separate section is intuitively appealing, as several others have also singled out the gospel of John as potentially different from the others, and even suggested that the translation of John was the work of Bede (Elliott and Ross, 1972). Notable also in John is orthographic <v> for <u> (Ross et al 1960, p. 23) and the infrequent use of the rune *wynn*, as well as the use of different colored ink. Others have supported these divisions to a certain extent. For example, Luke shows a more frequent use of uncontracted negatives, Van Bergen (2008, p. 291), and only Luke and John use an intensive ‘self’, whereas all four gospels have a reflexive ‘self’ (van Gelderen, 2000; 2018). Recall also, that Ross et al (1960) suggest a paleographic division at the end of Luke.

### 3. Method and Results

#### 3.1 Method

First, I take the numerical results from Brunner (1947) and convert them into percentages as shown in Table 4. The sections in Table 2 are of unequal length so, in order to compare, the calculations show the percentage of <oe> in the overall total for each section. I then use a concordance programme to search for the following six lexical items; *hwōēr* ‘where’; *twōēge* ‘two’; *twoelf* ‘twelve’; *wōē* ‘we’; *woeg* ‘way’; *woerc* ‘work’ and their variants *hwēr*, *twēge*, *twelf*, *wē*, *weg*, *werc*. These are the most frequently occurring items of those that have a mid-front vowel following /w/. The data are taken from Skeat (1881-1887) and spot checked against the manuscript. In terms of orthography, <u> and <w> are treated as equivalent. For the four nouns, *twēge*, *twelf*, *weg*, and *werc*, which are inflected for case in Old English, I count all inflected and non-inflected forms. Compounds of *werc*, that is, *wercmenn* ‘workmen’ and *wercmonn* ‘workman’ are included. For *weg*, ‘way’, I also searched for *aweg* and *awoeg* ‘away’, but did not find examples. Percentages of <oe> for these six lexical items are calculated, enabling direct comparison with Brunner’s figures. The raw figures are shown in Table 5; the individual percentages in Table 6 and the overall percentages in Table 7. For *wesan* there are 7 examples with the vowel spelled <æ>. These will be discussed in section 3.3.

#### 3.2 Results

As can be seen from Table 4, which shows figures taken from Brunner’s research, the rounded vowel of forms of *wesan* is fairly infrequent in

the first two sections of Lindisfarne, as Brunner (1949, p. 35) describes. However, when it comes to an overall comparison with forms of *cueþan*, sections 1 and 2 are markedly different from each other. Forms in <e> are said to be comparatively rare after the first four chapters of Mark, but become frequent again in the first three chapters of John (Brunner 1947, p. 35). As can be seen from Table 4, this essentially sets sections 3 and 4 as markedly different from the others. However, there is little similarity between sections 1 and 2 as there was with *wesan*.

		Sec 1	Sec 2	Sec 3	Sec 4	Sec 5	Total
<i>wesan</i> 'be'	<e>	120	68	23	139	65	<b>415</b>
	<oe>	8	5	28	132	38	<b>211</b>
	%<oe>	6.3%	6.9%	54.9%	48.7%	36.9%	
<i>cueþan</i> 'say'	<e>	126	18	1	18	20	<b>183</b>
	<oe>	64	29	71	166	38	<b>368</b>
	%<oe>	33.7%	61.7%	98.6%	90.2%	65.5%	

Table 4. Frequency of rounded high front vowel <oe> vs unrounded vowel, <e>  
Raw numbers from Brunner (1947)

Assuming that /ø:/, the more marked form and the one that is already lost from West Saxon, is the most conservative pronunciation, comparison of Table 4 and Table 3 shows that the results for *wesan* and verbal *-s* taken together show an overall tendency towards a less conservative variety in section 1. After this, the correlation falls apart. For example, phonology figures in table 4 support a similarity between sections 3 and 4, the ones for morphology, in table 3, do not.

Turning now to the new data, the six selected lexical items, arranged alphabetically in Table 5, it is apparent that they are not as frequent as the forms of *wesan* and *cueþan* in Table 4. However, some general trends are apparent. Most notable is that, if the results for 'we' are ignored, there are very few examples of unrounded <e> in sections 3 and 4, a similar result to that found for *cueþan*. Also notable from the 'total' column is that there are more examples of rounded vowels than of unrounded ones for all the items except *hwēr* and *wē*.

		Sec 1	Sec 2	Sec 3	Sec 4	Sec 5	Total
<i>hwēr</i>	<e>	8	0	1	6	8	<b>23</b>
<i>hwōēr</i>	<oe>	0	0	0	2	1	<b>3</b>
<i>twēge</i>	<e>	3	0	1	1	1	<b>6</b>
<i>twōēge</i>	<oe>	23	4	10	26	5	<b>68</b>
<i>twelf</i>	<e>	7	4	0	0	3	<b>14</b>
<i>twoelf</i>	<oe>	4	1	11	12	3	<b>31</b>
<i>wē</i>	<e>	49	16	24	55	66	<b>210</b>
<i>wōē</i>	<oe>	3	0	5	12	2	<b>22</b>
<i>weg</i>	<e>	11	1	1	4	1	<b>18</b>
<i>woeg</i>	<oe>	11	2	13	16	2	<b>44</b>
<i>werc</i>	<e>	12	1	0	1	7	<b>21</b>
<i>woerc</i>	<oe>	2	0	2	14	25	<b>43</b>

Table 5. Instances of rounded high front vowel <oe> and unrounded vowel, <e>  
Selected lexical items

Why might *hwēr* and *wē* behave differently from the others? Possibly, little weight should be given to the results for *hwēr*, as it is not frequent enough for definite conclusions to be drawn. Puzzling, however, are the results for ‘we’, where there is an overwhelming absence of rounded forms. There are a number of possible explanations for the low incidence of rounding. First, *we* differs from the other lexical items on the list in being a function word, which means that the spelling may be more conventionalised. Also, the vowel is word final, unlike the other examples, and this may have a phonological effect, or even influence the orthography. Note also, the low overall figures in sections 2 and 3. This highlights the fact that sections 2 and 3 are relatively short, containing only 6 units and 7 units respectively of the 64 equal units that Brunner used, as opposed to section 1 which contains 24 units and, and sections 4 and 5 which contain 21 and 12 units respectively.

In Table 6, the percentages of <oe> for each section are shown, calculated from the figures in Table 5 and merged with the percentages in Table 4. As has already been noted, sections 3 and 4 stand apart. There is consistently a higher percentage of rounding in sections 3 and 4 for all the lexical items, Even for those lexical items where the overall total instances of rounding is low, such as ‘we’, sections 3 and 4 still have the most. What is not apparent at all is the clear break at MkGl (Li) 5,40, so marked for *wesan*.

		Sec 1	Sec 2	Sec 3	Sec 4	Sec 5
		-MtGl (Li) 26,16	MtGl (Li) 26,17 - MkGl (Li) 5,40	MkGl (Li) 5,41 – LkGl (Li) 2,9	LkGl (Li) 2,10 – JnGl (Li) 3,13	JnGl (Li) 3,14- end.
<i>wesan</i>	%<oe>	6.3%	6.9%	54.9%	48.7%	36.9%
<i>cueþan</i>	%<oe>	33.7%	61.7%	98.6%	90.2%	65.5%
<i>hw(ō)ēr</i>	%<oe>	-	-	-	25%	11%
<i>tw(ō)ēge</i>	%<oe>	88.5%	20%	90.9%	96.3%	83.3%
<i>tw(o)elf</i>	%<oe>	36.3%	80%	100%	100%	50%
<i>w(ō)ē</i>	%<oe>	5.8%	0%	17.24%	17.9%	2.9%
<i>w(o)eg</i>	%<oe>	50%	33.3%	92.8%	80%	66.6%
<i>w(o)erc</i>	%<oe>	14.3%	0%	100%	93.3%	78.1%

Table 6. Percentages of rounded high front vowel <oe> for selected lexical items

One factor that could have an influence is the difference in vowel length. The short vowel /e/ tends to be rounded more frequently than the long vowel /e:/ (Hogg, 2011, p. 199), which would go some way to explaining the low incidence of rounding in *w(ō)ē*. However, in the case of the numerals, ‘two’ and ‘twelve’ it might be expected that *tw(ō)ēge* with its long vowel would have less frequent rounding than *tw(o)elf*, but that is not so. In sections 1 and 5 it has considerably more, and in sections 4 and 5 there is little difference between the two numerals.

Finally, Table 7 shows the overall percentages when the figures in Tables 4 and 5 are combined.



	Sec 1	Sec 2	Sec 3	Sec 4	Sec 5
	-MtGl (Li) 26,16	MtGl (Li) 26,17 - MkGl (Li) 5,40	MkGl (Li) 5,41 – LkGl (Li) 2,9	LkGl (Li) 2,10 – JnGl (Li) 3,13	JnGl (Li) 3,14-end.
<e>	336	108	51	224	171
<oe>	115	41	140	380	114
	<b>25.5%</b>	<b>27.5%</b>	<b>73.3%</b>	<b>62.9%</b>	<b>40%</b>

Table 7. Total percentages of rounded high front vowel &lt;oe&gt;

Here, the differences between the 5 sections are clearly revealed. Overall there is much less use of the rounded vowel in the first two sections, supporting Brunner's original suggestion of a break at MkGl (Li) 5,40. There is a considerable increase in sections 3 and 4 followed by a reduction in the final section, supporting the often cited break at the end of Luke. Comparison of table 7 with table 3, shows conservative morphology and phonology in section 4 and more innovative morphology and phonology in section 1, but little correlation otherwise.

### 3.3 West Saxon <wær->

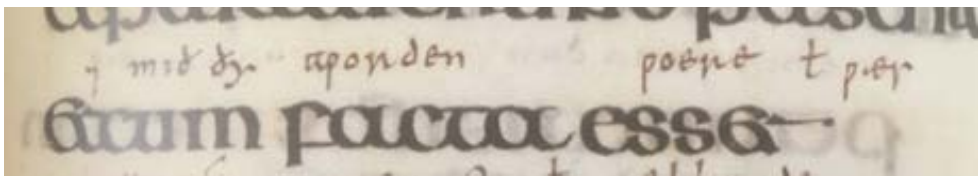
Recall that in section 3.1 it was mentioned that 7 examples were found with the vowel spelled <æ>.

As was explained in section 2, the lexical items under consideration are those that have either a short or long mid-front rounded vowel in Northumbrian (/ø(:))/, of which the southern equivalents are /e/ and /e:/. The one exception is the stem of *wesan*, which is *wōēr-* and *wēr-* in the north but *wǣr-* in West Saxon, thus <ǣ> being typical of West Saxon.

Surprisingly, Brunner (1947) does not mention the occurrence of West Saxon <æ> in Lindisfarne, even though she carefully documents her methodology to the extent of mentioning two examples of <eo>, which she judges as mistakes for <oe>. My electronic searches revealed 7 examples of this spelling. The question here is whether these are actually in the manuscript or are editorial mistakes; Cuesta (2016) is one of the most recent researchers to critique the editorial license taken by Skeat's (1871-1887) editions. If they are real examples penned by the glossator, the question becomes how they are distributed and what significance can be attached to their inclusion.

The examples are found in the following verses: MtGl (Li)13,35; MtGl (Li) 23,23 (both section 1); MkGl (Li).15,42; MtGl (Li) 16,11 (both section 3); LkGl (Li) 8,9; LkGl (Li) 15,20; LkGl (Li) 19,11 (section 4). There seems to be no regularity in this distribution, with examples in 3 of the 5 sections. The most significant observation that can be made is that there are no examples in John's gospel. It remains to be determined whether the manuscript confirms these 7 examples. I examined the remaining examples, and confirmed all but one as <æ>.

One of the easiest ways to see the orthographic difference between <oe> and <æ> in the manuscript is to examine a doublet, i.e. when a Latin word is given a double gloss, the two Old English words separated with the symbol '†'. Doublets are not unusual, and the gloss contains over 3000 in total (Kotake, 2006, p. 37). Such an example is shown in Figure 2, the doublet showing the Latin singular preterite subjunctive *esset*, glossed with both the indicative *wæs* and the subjunctive *woere*.



&	mið ðy¹.	aworden	woere † wæs	
<i>et</i>	<i>cum</i>	<i>facta</i>	<i>esset</i>	
and	when	done	were	(LkGl (Li) 22,14)

Figure 2. Doublet showing adjacent *woere* and *wæs*

Note the initial grapheme in each is *wynn* and that *wæs* ends with a long 's'. The vowel in *wæs* (past singular indicative of *wesan*) is always <æ>. (Recall that the short vowels do not raise in northern varieties, unlike /æ:/. See Table 1). The difference between the two vowels is clear in Figure 2. Next, compare the orthography in Figure 2 with that in Figure 3, which shows one of the 7 examples of <æ> under examination. This example also happens to be in a doublet. Here the vowels in *wæs* and *wære* are clearly similar to each other and similar to the *wæs* in Figure 2.

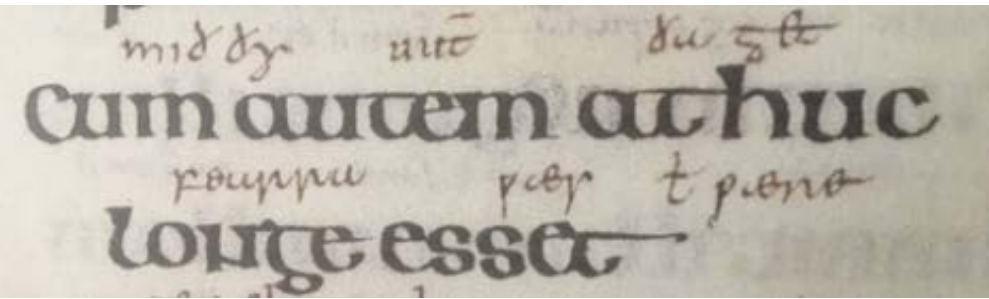
					
mið ðy	uitedlice	ða get	fearra	esset	
<i>cum</i>	<i>autem</i>	<i>adhuc</i>	<i>longe</i>	<i>wæs t wære</i>	
when	indeed	then	far	were	(LkG1 (Li) 15.20)

Figure 3. Doublet showing adjacent *wære* and *wæs*

I examined the remaining examples, and confirmed all but one as <æ>. They are all similar to the graphemes <æ> seen in Figures 2 and 3. The doubtful example is shown in Figure 4. Here there is even a <wæs> for comparison in the following line. As can be seen, this example is different when compared with the other examples and could also be taken as a hastily written <oe>.

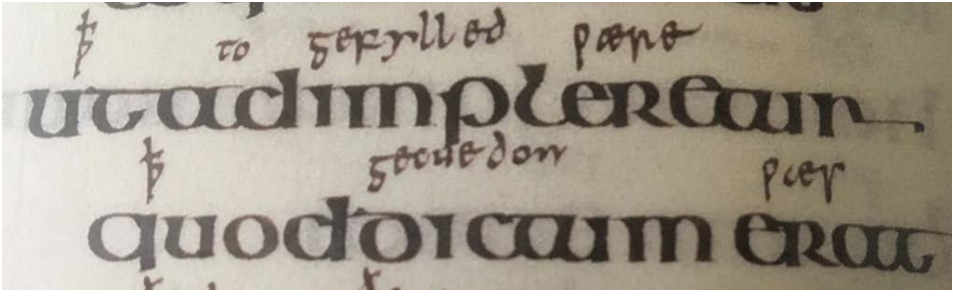
					
þæt <sup>2</sup>	to gefylled <sup>3</sup> wære	þæt	gecuedon	wæs	
<i>ut</i>	<i>adimpleretur</i>	<i>quod</i>	<i>dictum</i>	<i>erat</i>	
that	fulfilled were	that	Said	was	(MtG1 (Li) 13.35)

Figure 4. Doubtful example of *wære*, showing *wæs* on the line below

In conclusion it can be said that at least six of these seven examples are confirmed. How to interpret them remains a puzzle. There are not enough examples of <æ> to posit a southern influence and they are present in all the gospels except John. All that can perhaps be said is that the result for doublets underscores the singularity of John's gospel, already emphasized by much past research.

#### **4. Conclusion**

This investigation analyses the distribution of mid-front rounded vowels in the glosses to the Lindisfarne gospels. Rounding of these vowels when they follow /w/ is a dialect feature of late Old Northumbrian, one of the features that differentiates it from the dominant literary dialect of the 10<sup>th</sup> century, West Saxon. This feature is variable in the glosses of the Lindisfarne. The aim is to find whether there is uneven distribution that could either indicate a different glossator or a single glossator copying from existing exemplars and to look for correlation with other features to aid in confirming the demarcations. Already hypothesised demarcations in the gospels are presented and the investigation looks for evidence to support the division into sections. The data takes already published numbers for two verb stems. These are reworked into percentages together with new data taken from six of the most frequently occurring lexical items that have mid-front vowels following /w/.

Rounded vowels are found to be variably distributed throughout the different sections and also each lexical item showed different distribution. The most significant finding is that in sections 3 and 4 rounded vowels occur most frequently, setting Luke's gospel and half of Mark (sections 3 and 4) apart as more conservative than the other sections. This finds support in van Bergen (2008) who notes an increased use of uncontracted negatives, a feature found more often in northern varieties, in sections 3 and 4, i.e. MkGl (Li) 5,40 to the end of Luke, the data in section 2 being too sparse to reach any firm conclusion (van Bergen, 2008, p. 291). Another feature that sets Luke apart is that it has the lowest incidence of the –s verbal ending (Cole, 2014, and see Table 3 above) making section 4 also more conservative, though this does not extend to section 3.

The overall figures shown in Table 7 support the demarcation at MkGl(Li) 5,40 with sections 1 and 2 patterning similarly, though there is considerable variation when looking at each individual lexical item. It must be cautioned that in some of the sections the data is quite sparse. For future work it could be helpful to include other nouns and verbs as, even though many are infrequent overall, they would contribute to the overall result. Finally, seven instances of the vowel /æ:/ occurring in the past tense *wære* are investigated and found to occur sporadically in all the gospels apart from John, supporting the much cited singularity of John's gospel.

### Sources

- The Lindisfarne Bible*. (2002). Munich: Faksimile Verlag.
- Skeat, W. W. (1871). *The Gospel according to Saint Luke in Anglo-Saxon and Northumbrian versions synoptically arranged, with collations exhibiting all the readings of all the mss.* <https://catalog.hathitrust.org/Record/001357329>
- Skeat, W. W. (1878). *The Gospel according to Saint John in Anglo-Saxon and Northumbrian versions synoptically arranged, with collations exhibiting all the readings of all the mss.* <https://catalog.hathitrust.org/Record/001357326>
- Skeat, W. W. (1878). *The Gospel according to Saint Mark, in Anglo-Saxon and Northumbrian versions, synoptically arranged, with collations exhibiting all the readings of all the mss.* <https://catalog.hathitrust.org/Record/001925960>
- Skeat, W. W. 1887. *The Gospel according to Saint Matthew in Anglo-Saxon, Northumbrian, and old Mercian versions, synoptically arranged, with collations exhibiting all the readings of all the mss.* <https://babel.hathitrust.org/cgi/pt?id=inu.32000000891434;view=1up;seq=16>

### References

- Blakeley, L. (1949). *Studies in the language of the Lindisfarne Gospels*. PhD. University of Birmingham.
- Bosworth-Toller *Anglo-Saxon dictionary*. Digital Edition. <http://bosworth.ff.cuni.cz/>
- Brunner, A. (1947). A note on the distribution of the variant forms of the Lindisfarne Gospels". *English and Germanic Studies*, 1, 32-52.
- Cole, M. (2014). *Old Northumbrian Verbal Morphosyntax and the (Northern) Subject Rule*. NOWELE Supplement Series 25. Amsterdam: Benjamins.
- Cole, M. (2016). Identifying the author(s) of the Lindisfarne gloss: Linguistic variation as a diagnostic for determining authorship. In J. F. Cuesta & S. Pons-Sanz (Eds.), *The Old English Gloss to the Lindisfarne Gospels: Language, Author and Context* (pp. 169-188). De Gruyter Mouton.
- Cuesta, J. F. (2016). Revisiting the manuscript of the Lindisfarne Gospels. In J. F. Cuesta & S. Pons-Sanz (Eds.) *The Old English Gloss to the Lindisfarne Gospels: Language, Author and Context* (pp. 257-285). De Gruyter Mouton.

- Cuesta, J. F., Ledesma, M. N. R. & Silva, I. S. (2008). Towards a history of Northern English: Early and late Northumbrian. *Studia Neophilologica*, 80:2, 132-159, DOI: 10.1080/00393270802493217
- Cuesta, J. F. & Pons-Sanz, S. (Eds.) (2016). *The Old English Gloss to the Lindisfarne Gospels: Language, Author and Context*. De Gruyter Mouton.
- Elliott, C. O., & Ross, A. S. (1972). Aldrediana XXIV: The linguistic peculiarities of the gloss on St. John's Gospel. *English Philological Studies*, 13, 49-72.
- Hogg, R. (Ed.). (1992). *The Cambridge History of the English Language*. Cambridge; New York: Cambridge University Press.
- Hogg, R. (2011). *A Grammar of Old English. Volume 1: Phonology*. Oxford: Wiley- Blackwell.
- Kotake, T. (2006). Aldred's multiple glosses: is the order significant? In M. Ogura (Ed.), *Textual and Contextual Studies in Medieval English: Towards the Reunion of Linguistics and Philology* (pp. 35-51). Bern: Peter Lang.
- Ledesma, M. N. R. (2016). *Dauides sunu vs. filii david*: The genitive in the gloss to the Lindisfarne Gospels. In J. F. Cuesta & S. Pons-Sanz (Eds.), *The Old English Gloss to the Lindisfarne Gospels: Language, Author and Context* (pp. 213-238). De Gruyter Mouton.
- Levin, S. R. (1958). Negative contraction: An Old and Middle English dialect criterion." *Journal of English and Germanic Philology*, 57, 492-501.
- Moore, S. & Marckwardt, A. H. (1951). *Historical Outlines of English Sounds and Inflections*. Ann Arbor: George Wahr.
- Pons-Sanz, S. M. (2000). *Analysis of the Scandinavian Loanwords in the Aldredian Glosses to the Lindisfarne Gospel*. Studies in English Language and Linguistics: Monographs 9. Valencia: Department of English and German Philology, University of Valencia.
- Ross, A. S. C., Stanley, E. G. & Brown, T. J. (1960). Some Observations on the gloss and the glossator. In Kendrick, T. D., T. J. Brown, R. L. S. Bruce-Mitford, H. Roosen-Runge, A. S. C. Ross, E. G. Stanley & A. E. A. Werner (eds.), *Evangeliorum Quattuor Codex Lindisfarnensis, Musei Britannici Codex Nero D.IV*. Volume II: *Commentariorum libri duo, quorum unus de textu evangeliorum latino et codicis ornatone, alter de glossa anglo-saxonica* (pp. 5-33). Olten/Lausanne: Graf.
- Van Bergen, L. (2008). Negative contraction and Old English dialects. *Neuphilologische Mitteilungen*, 109, 275-312.
- Van Gelderen, E. (2000). *A History of English Reflexive Pronouns*. Amsterdam: John Benjamins.
- Van Gelderen, E. (2016). Reflexive pronouns in the Lindisfarne glosses. Ms. Arizona State University.
- Wood, J. L. (2002). Negative contraction, dialect and the AB language: A note on Levin (1958). *Journal of Germanic Linguistics*, 14(4), 357-368.



# **PROSODY**

Handling editor: Anna Bothe Jespersen





## **Native and Non-native English Speakers' Assessment of Nuclear Stress Produced by Chinese Learners of English**

Congchao Hua  
Hubei University, City University of Hong Kong

Bin Li  
City University of Hong Kong

Ratree Wayland  
University of Florida

### **Abstract**

This study compared naïve native and non-native English speakers' assessment of nuclear stress produced by Chinese learners of English and explored the effects of prosodic cues on their assessment. Adopting rapid prosody transcription (RPT), naïve raters comprising 36 highly proficient non-native English speakers and 30 native English speakers rated 176 sentence recordings produced by six Chinese learners of English. Results revealed that the native and non-native raters made generally comparable judgements and their ratings were reliable compared with expert rating. However, ratings by the two groups differed significantly on 20 sentences. Acoustic analysis showed that while native speakers relied on duration when identifying nuclear stress in learners' English, non-native speakers relied on both duration and intensity.

## 1. Introduction

Nuclear stress in English is particularly important for marking information focus (Dickerson, 1989; Lu et al., 2012). It marks the speaker's selection of priority in thought groups, and thus facilitates information processing by the hearer (Fouz-González, 2015). Misplaced nuclear stress often affects comprehensibility of both native and non-native English speech (Jenkins, 2002; Hahn, 2004; Luchini, 2005; Ingels, 2011; Frost, 2011).

In speech production, nuclear stress is realized through prosodic correlates such as F0, duration and intensity, yet roles that these cues play in stress marking vary across languages (Mennen, 2015). These cross-language variations may lead to difference in perception of nuclear stress by native speakers and second language (L2) learner.

Research has revealed that naïve first language (L1) listeners can reliably transcribe sentence stress in both L1 speech and L2 speech, yet it is not clear whether L2 listeners can also recognize sentence stress in a comparable manner. Thus, this study set out to compare L1 and L2 English speakers' perceptual judgement of nuclear stress produced by L2 English learners. In addition, an acoustic analysis was performed to identify effects of phonetic cues on L1 and L2 perception of nuclear stress.

## 2. Literature Review

### 2.1 Nuclear Stress in English and Mandarin Chinese

Nuclear stress in English refers to the stress associated with the nuclear tone in an intonation unit. Native speakers of English often follow a specific pattern for nuclear stress assignment. When a whole utterance is under focus (broad focus), nuclear stress is by default on the last content word (Crystal, 1969; Roach, 1991; Cruttenden, 1997). This is proved to be true by Alternberg (1987), who reports that 88% of the utterances in the London-Lund corpus have their nuclear stress on the last content word.

Example 1 --What happened?  
--The baby is crying.

Example 2 --What's wrong?  
--He cheated us.

Here the underlined syllable in each example carries nuclear stress and is the focus of the whole information structure. In Example 1, nucle-

ar stress falls on the last word as it is the last content word in the utterance, whereas in Example 2, nuclear stress falls on the penultimate word, which is also the last content word in the utterance. However, there are some exceptions to the last-content-word rule.

Example 3 --Have you been to the lake?

--We walked around it.

Example 4 --What's the news?

--I met the president this morning.

In Examples 3 and 4, nuclear stress does not fall on the last content word. In example 3, it falls on 'around', which is a function word, and in Example 4, it falls on 'president', which is the penultimate content word. We will turn to these exceptions in detail later.

The above are all examples of broad focus. There is another type of focus: narrow focus. Narrow focus signifies contrast to known information or emphasis of new information. Often the word under narrow focus carries nuclear stress, despite its grammatical category or semantic weight.

Example 5 --Did you see Jane?

--No, I talked with Jane.

Example 6 --Who won the game?

--We won the game.

In Example 5, 'talked' carries nuclear stress because it contrasts with 'see', and in Example 6, 'We' carries nuclear stress because it directly answers the question 'Who' and is therefore emphasized. In both examples, nuclear stress does not fall on the last content word. To express contrast or emphasis in English, nuclear stress can fall on any word under focus.

Apart from the default pattern on the last content word, nuclear stress assignment involves over a dozen exceptional patterns that mainly include sentences ending with a function word, an early-stressed compound, a reflexive or reciprocal pronoun, a reporting phrase, a parenthetical, an empty word, a time or place adverbial, a phrasal verb ending with a preposition, a noun modifier, repeated information, and contrastive information, and event sentences and sentences containing a wh-object (Cruttenden, 1997; Wells, 2006).

The placement of nuclear stress is language specific. As a non-stress language (Selkirk & Shen, 1990), Chinese has less salient stress

than English (Yu & Andruski, 2011) and tends to stress the final syllable of a word or phrase (Chao, 1979). Unlike English, Chinese relies more on syntax for focus marking, and prosody is only a supplementary means of focus marking (Xu, 2004). In Chinese, broad focus tends to be marked after the main verb or towards the end of a sentence (Chen, 1995) with no phonological manifestation (Xu, 2004), whereas narrow focus can be achieved either syntactically or phonologically (Xu, 2004). In other words, not all focuses in Chinese are realized through nuclear stress and narrow focus in Chinese is more likely to be realized through nuclear stress than broad focus. The following are two examples about how narrow focus is achieved in Chinese.

Example 7 -- 是谁 赢了 比赛?

*Shishui yingle bisai?*

Who won the game? (Who won the game?)

--是我们 赢了 比赛。

Shi women yingle bisai.

It's we won the game. (It's we that won the game.)

Example 8 --谁 赢了 比赛?

*Shui yingle bisai?*

Who won the game? (Who won the game?)

--我们 赢了 比赛。

Women yingle bisai.

We won the game. (We won the game.)

Example 7 shows the syntactic marking of narrow focus, where the focus ‘我们 women’ does not necessarily carry nuclear stress because there is the focus marker ‘是 shi’. In example 8, however, the focus ‘我们 women’ carries nuclear stress, as the absence of the focus marker ‘是 shi’ necessitates the prosodic marking of the focus.

The difference between the use of nuclear stress in English and Chinese often contributes to Chinese speakers’ misplacement or misuse of nuclear stress in English. For example, they tend to assign nuclear stress to the final word or syllable in an utterance (Yu & Andruski, 2011), to every word in an utterance (Juffs, 1990) or even to pronouns (Deterding, 2006). Such deviations in their English may lead to communicative problems as native English speakers as well as other non-native English speakers may misinterpret their intended message.

## **2.2 Acoustic Realizations of Stress in English and Mandarin Chinese**

The phonetic realization of nuclear stress in English has been widely investigated (Xu & Xu, 2005), including acoustic parameters such as F0 (pitch), duration, and intensity (Bolinger, 1986; Roach, 1991; Cruttenden, 1997; Pennington & Ellis, 2000; Chun, 2002; Ingels, 2011; Frost, 2011; Lu, Wang & de Silva, 2012). Acoustically, nuclear stress in English is indicated by a change in pitch height or pitch contour, a lengthening of the vocalic part in the stressed syllable, and an increase in intensity. It is 'generally accomplished by means of a co-occurrence of relatively extreme values of all three parameters' (Pennington & Ellis, 2000). A number of research suggests that pitch is the most indicative of nuclear stress in English, followed by duration and intensity (Lieberman, 1960; Roach, 1991; Cruttenden, 1997; Frost, 2011). However, there is also evidence suggesting a robust role of intensity in stress perception (Sluijter, van Heuven & Pacilly, 1997; Tamburini & Caini, 2005), and the co-dependent nature of duration to pitch increment (Bolinger, 1958; Ciszewski, 2012). In short, despite the disputes over the roles of phonetic cues to stress, a consensus is that pitch, duration and intensity are relevant cues and all contribute to English stress, but with decreasing importance (Roach, 1991).

Similarly, stress in Chinese is also realized through changes in pitch, duration and intensity. As a non-stress language, Chinese seldom marks focus with nuclear stress. When focus in Chinese is marked with stress (often contrastive stress for narrow focus), the pitch range of the element under focus is drastically expanded and that of the elements following the focus is greatly compressed (Shih, 1988; Xu, 1999; Yuan, 2004; Kabagema-Bilan, Lopez-Jimenez & Truckenbrodt, 2011), just as in English (Jin, 1996; Xu, 1999; Chen, 2003; Liu & Xu, 2005).

Unlike pitch, which has attracted wide attention, duration and intensity in Chinese stress have been relatively under-researched. Both Jin (1996) and Yuan (2004) report a lengthening of the syllable under stress and an increase in its intensity. Jin (1996) further claims that a stressed Chinese syllable is always longer but louder only in the sentence-final position. Likewise, Yuan (2004) found that syllable lengthening is especially salient for sentence-final stress, and the intensity of the stressed syllable is the highest and drops drastically thereafter. These findings are confirmed by Swerts and Kraemer (2004), who report that a stressed syllable in Chinese is the longest in sentence-final positions and that intensity rises and drops drastically after the stressed syllable.

Chen (2003) and Chen and Gussenhoven (2008), however, emphasize that the role of pitch at the sentence level is greatly weakened in Chinese. They found that the duration of word under contrastive focus is directly related with the degree of emphasis, yet pitch only varies in the focus and non-focus conditions but does not indicate the degree of emphasis.

In sum, previous research reveals that duration, intensity and pitch all contribute to sentence stress in Chinese, but with different importance.

### **2.3 Rapid Prosody Transcription (RPT)**

Speakers of different languages perceive and interpret the acoustic parameters of pitch, duration, intensity and vowel quality differently in oral communication (Beckman, 1986; Low & Grabe, 1999; Pennington & Ellis, 2000). As a result of L1 influence, L2 learners tend to use cues to English stress in a different manner from its native speakers. Consequently, native speakers often find it hard to rely on prosody to interpret L2 learner speech (Gray, 2015; Ingels, 2011).

Studies on speech prosody have proposed and tested various methods to the evaluation of L2 learner's English. Among the most recent development, Rapid prosody transcription (RPT) (Cole et al., 2010, 2016) emerges as an effective method. It refers to assessing prosody by a group of naïve listeners (listeners with no phonetic or phonological knowledge) and the percentage of listeners who have assigned a prosodic feature (e.g., prominence or intonation boundary) to a certain word or position in an utterance will be the rating score for that feature.

RPT has been proven an effective method for marking prominence and intonation boundary in different languages with different transcribers. For example, Cole et al. (2010) and Cole et al. (2016) found RPT effective for marking prominence and intonation boundary in American English by American English speakers; Smith (2011, 2013) and Roux et al. (2016) found RPT effective for marking prominence and intonation boundary in French by native French speakers; Pintér<sup>1</sup> et al. (2014) report that RPT ratings of L1 English by L1 and L2 English speakers were comparable; Smith and Edmunds (2013) report that L1 English speakers' RPT for L1 English and L2 English are both reliable. In addition, Smith (2009) compared native French speakers' RPT with expert transcription and found that their results are significantly correlated.

Previous findings have confirmed that naïve native speakers are able to make reliable judgements about both L1 and L2 prosody, so are naïve

L2 speakers about the target language prosody. However, it remains untested whether RPT can be applied with L2 speakers to assess L2 prosody, and whether there is a high degree of correspondence between L1 and L2 speakers' judgements. Variations in prosody across languages and L2 acquisition both suggest that naïve L1 and L2 speakers may differ in their assessment of L2 prosody. Therefore, this study aimed to answer the following research questions:

1. Do L1 and L2 English speakers yield comparable results when assessing nuclear stress produced by Chinese learners of English?
2. If there are discrepancies between the ratings by L1 and L2 English speakers, what acoustic cues contribute to these discrepancies?

### **3. Research Method**

#### **3.1 Participants**

##### **Speakers**

Recordings of learner speech were from six English majors (1 male and 5 female) at a provincial university in central mainland China: three were first-year students (intermediate level English learners), and three were third-year students (advanced level English learners). They were between 18 to 22 years old. All speakers came from the same province and reported using Mandarin Chinese as their primary language in everyday communication.

##### **Raters**

The raters were 36 L2 English speakers and 30 L1 English speakers. The L2 English speaking raters (henceforth L2 raters) all spoke either Mandarin or Cantonese as their first language, had received postgraduate education related to English language (either in linguistics or literature), and had been studying or/and using English for over 15 years. These raters were between 22 to 45 years old. Ten were male and 26 were female. They were all highly proficient in English and reported using English frequently in their daily communication.

The L1 English-speaking raters (henceforth L1 raters) were all from the U.K. and spoke standard British English. They were between 23 to 50 years old. Twenty of them were male and 10 were female. None of them were fluent in Mandarin Chinese, though some had learned basic Chinese and could speak a little.



None of the raters reported having received systematic training in English prosody. The L2 raters participated on a voluntary basis, and the L1 raters each were paid 30 RMB yuan for their participation.

### Expert

The first researcher served as an expert for nuclear stress rating. As a non-native English speaker, she had majored in English phonetics and phonology and had been systematically trained in English prosody. She had taught intermediate to advanced English learners at a Chinese university for over ten years and her own English proficiency was the highest at C2 Mastery for foreign language learners<sup>1</sup>.

### 3.2 Stimuli

The stimuli included 30 sentences selected from the recording of a reading task done by each of the six participants chosen from two university classes, totaling 176 sentences (four sentences were of bad quality and thus excluded). The reading task was to assess the learners' mastery of nuclear stress and contained two parts: sentences in isolation and a dialogue. The dialogue and sentences were adapted from Wells (2006). The stimuli produced by each learner included 15 sentences in isolation and 15 in context (i.e., the dialogue) (See the appendix).

The 15 sentences in isolation represent all typical types of nuclear stress placement summed up in Wells (2006), including the default pattern (nuclear stress on the last content word) and 14 exceptions to the default pattern where nuclear stress does not fall on the last word in an utterance. These 14 exceptions (13 types) include: one event sentence, one wh-object sentence, two contrastive sentences (one long contrastive sentence broken into two parts), and 10 other sentences respectively ending with different components: a function word, an early stressed compound, a time adverbial, a reporting phrase, a parenthetical, an empty word, repeated information, a noun modifier, reflexive pronoun and a phrasal verb ending with a preposition.

---

<sup>1</sup> C2 Mastery is the highest among the six reference levels (A1-2, B1-2, C1-2) of language proficiency, according to *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment* by the Council of Europe (<https://rm.coe.int/1680459f97>) and *The Core Inventory for General English* by the British Council in 2017 (<https://www.equals.org/resources/the-core-inventory-for-general-english/>).

The 15 sentences taken from the dialogue represent some of the types, including five default pattern sentences, two contrastive sentences, two ending with a time adverbial, one ending with a function word, one ending with an empty word, one ending with an early-stressed compound, one ending with an early-stressed compound and a parenthetical, one ending with repeated information, and one ending with a post-modifier.

The task was designed in this way to assess if the participants had awareness of nuclear stress in English and if they could apply such awareness in context. However, this is not the focus of this study and the findings concerning these learners' awareness and application of nuclear stress in English is not reported here.

Three sentences produced by two native British English speakers (1 male, 1 female) were also included in the stimuli. The three sentences were all taken from the dialogue mentioned above. The recordings of these native speaker sentences were adopted from Wells (2006).

### **3.3 Procedure**

The recordings of the learners' reading of the isolated sentences and the dialogue were firstly split into individual sentences. This yielded 180 sentence recordings (6 participants x 30 sentences), of which four were of bad quality and excluded. The 176 stimulus sentences were incorporated into 3 questionnaires designed on Qualtrics ([www.qualtrics.com](http://www.qualtrics.com)), each containing 60 sentences produced by two L2 English learners and the same three sentences produced by the two native speakers. In addition, questions about the rater's age, first language, and confidence level in rating were also included. For the L2 raters, information about their years of English learning and experience in English pronunciation learning was also elicited.

The questionnaires were distributed online to the target raters, who listened to the sentences individually and clicked on the word that they heard as the most prominent in each sentence. Eleven to 13 L2 raters and 10 L1 raters responded to each questionnaire. The expert rater rated all the 176 learner sentences. The recordings were randomized and rated twice by the expert rater with a two-week interval.

When all ratings were completed, the expert ratings were first compared and converted. Then the RPT results were converted and compared with the expert rating. Lastly, the acoustic cues contributing to their discrepancies were explored.

### **3.4 Data Analysis**

First, we compared the ratings of the L1 and L2 raters. More specifically, we calculated the percentage of raters selecting a certain word as the most prominent for each word in each sentence recording. Then the percentage for each target word (the word supposed to carry nuclear stress according to theory, as underlined in the appendix) that was judged by the raters as carrying nuclear stress was converted to a numerical grade (0, 1, or 2) using the following coding scheme: Ratings higher than 60% were converted to 2, standing for good mastery of the nuclear stress production. Ratings lower than 60% but the highest in the sentence were converted to 1, representing partial mastery; ratings as the highest in the sentence but shared with other word(s) in the same sentence were also converted to 1. Other ratings lower than 60% were converted to 0, standing for non-mastery of the nuclear stress production.

Likewise, the expert ratings were also converted to 0, 1, and 2. A target word marked as carrying a nuclear stress in both expert ratings was given 2; that marked in one rating was given 1; and that not marked in either rating was given 0.

This conversion was necessary for direct comparison between the expert rating and the naïve raters' ratings. The expert rated each target word as 0 (not carrying nuclear stress) or 1 (carrying nuclear stress) in each round of rating, whereas the two groups of naïve raters' ratings for each target word were in percentage (the percent of naïve raters choosing the target word as the most prominent). Thus, it would be difficult to compare the numbers (0 or 1) with the percentages. The conversion of the ratings mentioned above was a solution to this problem and makes the comparison possible.

All three sets of scores, that is, scores from the expert, the L1 raters and the L2 raters, were compared using Kendall's tau correlation coefficients in SPSS 20.0 to assess the intra-rater and inter-rater reliabilities. This non-parametric statistic was chosen because not all of the three sets of scores were in normal distribution.

Secondly, we calculated the discrepancies (in percentage, non-converted) between the L1 and L2 raters. Then the 176 sentences were ranked ordered according to the degree of discrepancies (indexed by percentage scores). Sentences containing target words with L1-L2 discrepancies equal to or above 33.3% (meaning one third of the raters in each group were in disagreement) were identified for acoustic analysis. Likewise, sentences containing target words with high L1-L2 agreement

(above 80% in both L1 and L2 ratings) were also selected. In total 20 sentences with great L1-L2 discrepancies and 20 sentences with high L1-L2 agreement were selected for acoustic analysis to explore further the relationship between L1 and L2 ratings.

The acoustic data collected included the following:

- 1) Duration: duration of the target words in the high-agreement sentences, that of words with ratings above 20% in the high-disagreement sentences, and also duration of the entire sentences. Duration ratio was then calculated by dividing word duration by sentence duration.
- 2) Fundamental Frequency (F0/pitch): F0 range for each word. Values of F0 peak, F0 valley, and mean F0 of the target words in the high-agreement sentences and of words with ratings above 20% in the high-disagreement sentences. F0 slope, calculated by dividing F0 range by word duration, and F0 ratio, calculated by dividing the mean F0 of each word by that of each sentence.
- 3) Intensity: intensity range for each word. Values of peak, valley, and mean of target words in the high-agreement sentences and of the words with ratings above 20% in the high-disagreement sentences. Intensity ratio, calculated by dividing the mean pitch of each word by that of each sentence.

All the calculations were done with raw values, and then z-normalized for cross-sentence and inter-speaker comparison. A series of Pearson product-moment coefficients were computed to explore the correlations between ratings and these cues. In addition, independent-samples *t*-tests were run to compare the acoustic characteristics of the exemplar sentences with high agreement with those of the sentences with high discrepancies.

## 4. Findings

### 4.1 Comparison between Ratings by Expert, L1 Raters and L2 Raters

To assess the reliability of RPT with naïve raters, recordings of three sentences read by native British English speakers were included in the rating task. Results showed that ratings for the three sentences were highly consistent, with 75%-95% of the L1 and L2 raters choosing the target words as the most prominent in each of these sentences. This high level of consistency among the raters on native production can serve as a bench mark against which different performances by the L2 English learners can be measured.

For the learners' recordings, ratings were less consistent, as expected. For some sentences, there was no agreement between L1 and L2 raters on the most prominent word, while for others their agreement could reach 100%.

Kendall's tau correlation coefficients were run to compare the two ratings by the expert for the 176 learner's sentences as well as the three sets of scores (expert rating, L1 speaker rating and L2 speaker rating) for these sentences. Results indicated that there was a strong positive correlation between the first expert rating and the second one ( $\tau_b = .842, p < .001$ ), representing high intra-rater reliability. For inter-rater reliability, there was a strong positive correlation between L1 and L2 ratings ( $\tau_b = .610, p < .001$ ), but a moderate positive correlation between the expert rating and the L2 rating ( $\tau_b = .484, p < .001$ ), and between the expert rating and the L1 rating ( $\tau_b = .353, p < .001$ ). The inter-rater reliability averaged at .482, which was moderate. Thus, the intra-rater reliability was higher than the inter-rater reliability.

#### 4.2 Effects of Learner Proficiency

Intra- and inter-rater reliability for all three groups of raters is shown in Table 1. On average, the intra-rater reliability was high and the inter-rater reliability was moderate. However, both types of reliabilities varied as a function of talker, i.e., with the L2 learners' proficiency level. Specifically, higher degrees of reliabilities were obtained for higher proficiency learners (Talkers 4, 5, 6), and lower reliabilities for lower proficiency learners (Talkers 1, 2, 3). As shown in this table. The intra-rater reliabilities for Learners 1, 2, 3 varied from .70 to .942, which was in a lower range in comparison to those for Learners 4, 5, 6, varying from .801 to 1. The average inter-rater reliabilities followed the similar pattern: those for the lower proficiency learners were moderate (between .30 and .50) to high (above .50), and those for the higher proficiency learners varied at a higher range from .520 to .621. Besides, for all the six L2 English learners, inter-rater reliability was lower than intra-rater reliability. Among the six correlations between expert rating and L2 speaker rating, two were high at .843 (Learner 2) and .650 (Learner 6), one was low at .190 (Learner 1), and the rest three were moderate at .329 (Learner 3), .470 (Learner 4), and .464 (Learner 5). Among the six correlations between expert rating and L1 speaker rating, one was high at .613 (Learner 4), four were moderate at .450 (Learner 2), .349 (Learner 3), .386 (Learner 5), .440 (Learner 6), and one was low at .144 (Learner

1). Learner 1 was exceptional among all the learners. Her f0 contours extracted in Praat (Boersma & Weenink, 2018) were rather flat with little variation, which in part explains the greater disagreement between the ratings of the expert and of the two groups of naïve raters.

Learner	Intra-rater reliability		Inter-rater reliability				Average		
			Expert-L2		Expert-L1			L1-L2	
	$\tau_b$	$p$	$\tau_b$	$p$	$\tau_b$	$p$		$\tau_b$	$p$
1	.700	.000	.190	.243	.144	.374	.690	.000	.341
2	.942	.000	.843	.000	.450	.007	.442	.006	.578
3	.810	.000	.329	.044	.349	.033	.547	.001	.408
4	.866	.000	.470	.008	.613	.001	.781	.000	.621
5	.801	.000	.464	.006	.386	.023	.710	.000	.520
6	1	.000	.650	.000	.440	.010	.632	.000	.574

L1: native English speaker raters; L2: non-native English speaker raters; Average: the average of expert-L2, expert-L1 and L1-L2 correlations

Table 1 Intra-rater reliability and inter-rater reliability by learner

In addition, ratings by L1 and L2 raters agreed better than ratings by the expert and either group of naïve raters for all learners, except for Learner 2. Correlations between L1 and L2 speaker ratings for learners 1, 3, 4, 5, 6 were all high, at above .50. For Learner 2, the L1-L2 correlation was moderate at .442, which was the lowest among all L1-L2 correlations and much lower than the expert-L2 correlation of .843.

A Pearson product-moment coefficient was also run to test if sentence length affected judgement, as one may infer that longer sentences meant more challenges for raters as they would be faced with more choices. Results disputed such an inference ( $r=.134$ ,  $p=.076$ ). Therefore, it is safe to conclude that RPT ratings were not affected by sentence length.

### 4.3 Effects of Acoustic Cues

Although there were high correlations between L1 and L2 speaker ratings, the two groups of naïve raters disagreed greatly on 20 of the 176 sentences rated. On the other hand, the two groups agreed almost perfectly on another 20 sentences. These 40 sentences were chosen for acoustic analysis to uncover what may have led to the (mis)matching in perceptual judgement.

A series of Pearson product-moment coefficients were computed to explore the relations between acoustic cues (duration,  $f_0$ , intensity) and L1 and L2 speaker ratings (in the original percentage). The results showed high correlations between duration and ratings, suggesting that the raters relied on temporal parameters in locating nuclear stress. Specifically, the correlation between word duration (z-normalised) and L2 speaker rating was moderate,  $r=.478$ ,  $p<.001$ , between word duration and L1 speaker rating was strong,  $r=.505$ ,  $p<.001$ , between duration ratio and L2 speaker rating was moderate,  $r=.471$ ,  $p<.001$ , and between duration ratio and L1 speaker rating was also moderate,  $r=.498$ ,  $p<.001$ .

Regarding intensity, no significant correlation was found for L1 ratings, suggesting that L1 raters did not rely on intensity to identify nuclear stress. L2 ratings, however, was slightly correlated with intensity, as indicated by a slight positive correlation between L2 speaker's rating and maximum intensity ( $r=.298$ ,  $p=.018$ ), intensity range ( $r=.272$ ,  $p=.046$ ), mean word intensity ( $r=.287$ ,  $p=.023$ ), and mean sentence intensity ( $r=.297$ ,  $p=.018$ ).

Surprisingly, we did not find any correlation between the naïve raters' ratings and  $f_0$  correlates including  $f_0$  peak,  $f_0$  valley and mean,  $f_0$  slope and  $f_0$  range. This suggests that both groups of naïve raters did not rely on pitch variations for judging the placement of nuclear stress.

Next, a series of independent-samples  $t$ -tests were run to compare the acoustic cues (duration,  $f_0$ , intensity) of the most rated words in the 40 sentences. For the 20 sentences with great L1-L2 rater discrepancy, all words with a rating of above 20% by at least one group were identified, yielding 48 words. The values of acoustic parameters of these 48 words were compared with those of the 20 target words in the 20 high-agreement sentences.

Results revealed that the two groups of words differed significantly in duration. The target words in the sentences with high agreement ( $M=0.70$ ,  $SD=0.84$ ) were significantly longer than the target words in the sentences with great discrepancy ( $M=-0.27$ ,  $SD=0.93$ ),  $t(66)=4.0$ ,  $p<.001$ ,  $d=1.09$ . Duration ratios confirmed that target words in sentences with high agreement ( $M=0.35$ ,  $SD=0.08$ ) were comparatively longer in their hosting sentences than those in the group of sentences with low agreement ( $M=0.24$ ,  $SD=0.11$ ),  $t(66)=3.66$ ,  $p<.001$ ,  $d=1.14$ .

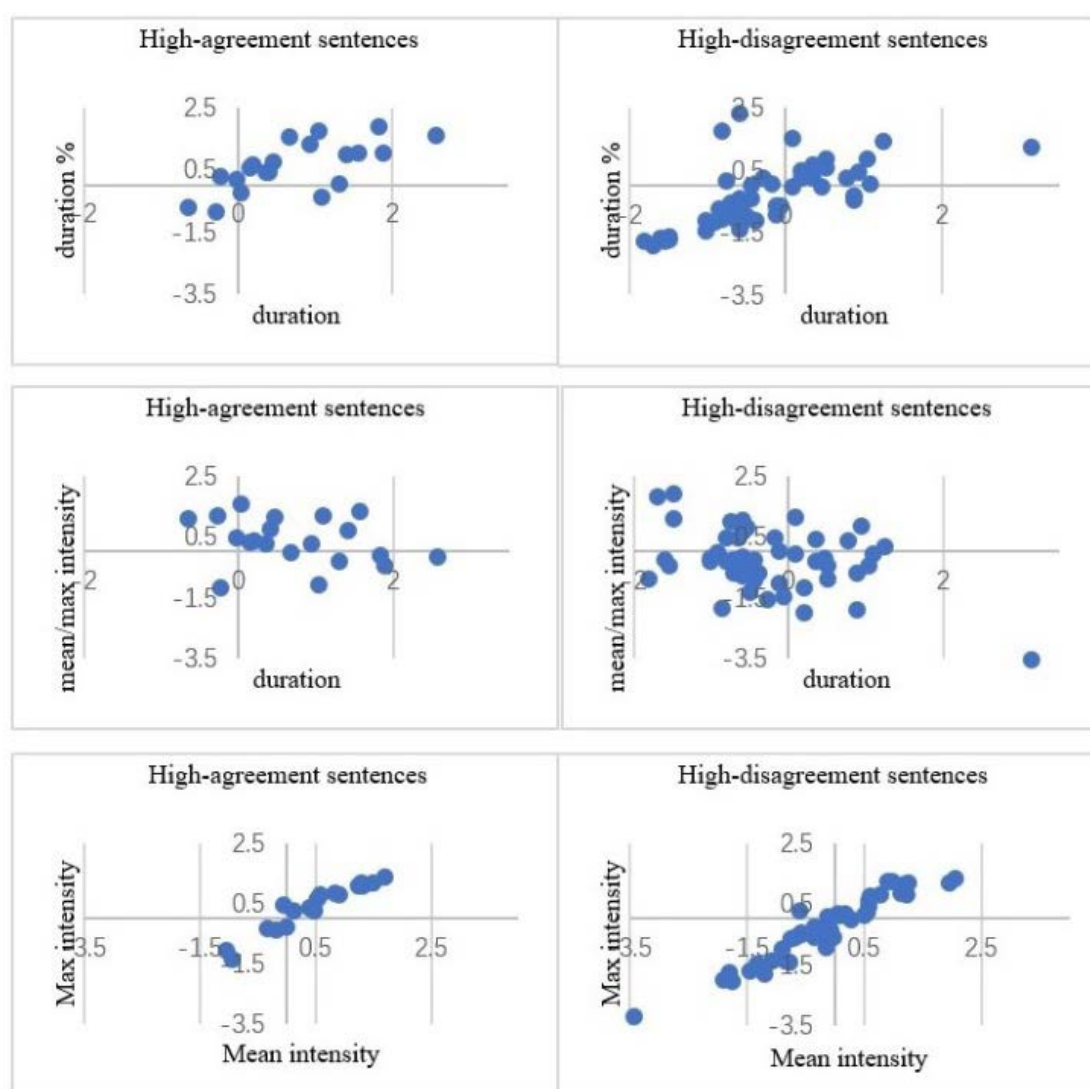


Figure 1. Scatterplots of z-normalised duration, duration ratio, maximum intensity, and mean intensity of the words in the two groups of sentences.

The two groups of words also differed significantly in intensity. More specifically, the target words in the high-agreement sentences had higher maximum intensity ( $M=0.51$ ,  $SD=0.76$ ) and higher mean intensity ( $M=0.45$ ,  $SD=0.78$ ) than the target words in the high-disagreement sentences ( $M=-0.21$ ,  $SD=1.03$ ), ( $M=-0.19$ ,  $SD=1.03$ ),  $t(66)=2.84$ ,  $p=.005$ ,  $d=0.80$ ,  $t(66)=2.15$ ,  $p=.055$ ,  $d=0.70$ , respectively.



However, the two groups of words did not differ in any  $f_0$  dimensions. This echoes with the patterns in perceptual judgement where neither the L1 raters nor the L2 raters seemed to use pitch in their judgements.

These differences in acoustic parameters between the two groups of sentences are illustrated with the scatterplots in Figure 1. As shown in the scatterplots, most of the target words in the high-agreement sentences have a z-normalized duration and a z-normalized duration ratio of above 0, which means that they are longer than the mean word duration in the two groups and occupy a larger portion of the total sentence duration. In comparison, over half of the words with ratings above 20% in the high-disagreement sentences have a z-normalized duration and a z-normalized duration ratio of under 0, meaning they are shorter than the mean duration and take up a smaller portion of the entire sentence.

A similar pattern is found with mean intensity and maximum intensity of words. While the z value of the mean intensity and maximum intensity of most target words in the high-agreement sentences are above 0, that is, higher than the means for the two groups, those of most words in the high-disagreement group are below 0, lower than the means.

This means that when there were robust acoustic cues (duration and intensity, in this case), L1 and L2 raters found it easy to locate nuclear stress and therefore their ratings matched; when these acoustic cues were obscure, variations occurred in their perception and judgement.

## **5. Discussion**

The moderate to high correlations between the expert score, the L1 rater score and the L2 rater score confirm the reliability of RPT among both native and non-native raters for assessing nuclear stress in L2 learner English. Naïve English speakers, native and non-native alike, can assess the nuclear stress in learner speech in a comparable manner.

However, RPT consistency across groups is affected by the learner's English proficiency level. Ratings by experts and by naïve raters were more reliable for high proficiency learners, but less so for low proficiency learners. For example, Learner 1 had the lowest intra-rater and inter-rater reliabilities. Acoustic analysis revealed few intonational fluctuations in her reading of the sentences and dialogue. Thus, the disagreement between the expert rating and naïve English speaker ratings can be attributed to the expert's awareness of and reliance on the acoustic cues associated with stress. While the presence of such cues made it easy

for the expert rater to decide on nuclear stress, lack of robust cues may have posed a problem. The naïve raters, by contrast, were less explicitly aware of such roles of acoustic cues and their rating may have been more psychoacoustically-based. Therefore, they were less affected by the presence or absence of acoustic cues when making judgements about nuclear stress in a sentence. Based on this finding, the low agreement within RPT can be an indicator of an L2 English learner's poor mastery of nuclear stress.

The agreement between expert rating and the naïve raters' ratings, though not strong for all the six learners, lends support to Smith's (2009) finding that expert rating and L1 speaker rating are comparable. The agreement between L1 and L2 raters' performances echoes with Pintér et al.'s (2014) finding that RPT results for L1 English prosody by L1 and L2 raters are comparable, yet we have taken a step further by proving that RPT results for L2 English prosody by L1 and L2 raters are also comparable, at least to a certain extent.

Another major finding of our study is that both L1 raters and L2 raters relied on duration for assessing nuclear stress in L2 English learners' speech. This dependence on temporal cues for nuclear stress supports previous findings on the phonetic realization of stress in both English and Chinese (cf., Roach, 1991; Jin, 1996; Cruttenden, 1997; Yuan, 2004; Swerts & Krahmer, 2004).

However, apart from duration, L2 raters also relied heavily on intensity for the task, but L1 raters did not. Since the production and perception of stress are correlated yet independent, this difference can be justified from two perspectives. One possibility is that the learners produced nuclear stress with the same acoustic realizations as native English speakers, but L1 raters were not strongly sensitive to intensity because intensity is the least robust cue for stress in English, whereas L2 raters were more sensitive to it due to the important role of intensity for stress in Chinese. However, if this was the case, a question arises for the role of pitch variations in L1 ratings since pitch is the most important cue for stress in English.

The absence of the role of pitch in both groups' judgements suggests that the nuclear stress produced by the learners was acoustically different from that by native English speakers. The Chinese-speaking English learners may have relied on duration and intensity but not pitch to realize stress in their English speech, as many pronunciation teaching materials describe stressed English words as being longer and louder

(e.g., Baker, 2009). Duration may have a far greater contribution to stress than intensity in these learners' English speech. Consequently, L1 raters relied only duration as a cue for nuclear stress location in these learners' speech.

In either case, there is evidence for L1 influence, either in the learners' or the raters' performance. The reliance on duration as a cue for stress by L2 learners and L2 raters and the absence of the role of pitch support Chen's (2003) and Chen and Gussenhoven's (2008) findings that unlike in English, duration is more important than pitch as a cue for stress in Chinese. The reliance on intensity echoes with previous research findings that intensity, together with duration, contributes greatly to stress in Chinese (Yuan 2004; Swerts & Kraemer 2004).

Given the absence of pitch as a cue in both L1 raters' and L2 raters' judgements of nuclear stress, it is highly likely that the L2 English learners did not make use of pitch to signal nuclear stress. This is worth L2 English teachers' attention. They need to raise their students' awareness of pitch as a cue for stress production to improve their production (and quite likely, their perception as well) of English stress.

## **6. Conclusion**

This study adopted the Rapid Prosody Transcription (RPT) and acoustic analysis to examine production of nuclear stress in L2 English. Naive and expert raters who were L1 or L2 speakers of English provided perceptual assessments of stress placement, which was then correlated with acoustic findings to evaluate the robustness of phoetic cues to nuclear stress. Comparable ratings from L1 and L2 naive rater groups confirmed the reliability and effectiveness of RPT in assessing L2 speech prosody. Besides, correlation patterns between perceptual results and phonetic features revealed that L1 and L2 raters may rely on different acoustic cues in making perceptual judgements. The former group seemed to use duration only, while the latter deployed both duration and intensity in locating nuclear stress. The variation in perceptual reliance could be attributed to L1 raters' lack of sensitivity or L2 raters' sensitivity to certain cues in L2 English. Future research may further examine the perceptual reliance by increasing learner diversity such as recruiting L2 learners from various proficiency levels and language backgrounds. More diverse L2 production could also contribute to maximizing the potentials of RPT as an effective and reliable method to assess L2 prosody.

**Acknowledgement:** This study received support from the Research and Development Project EDB(LE)/P&R/EL/175/12, under the Language Fund by SCOLAR, Hong Kong S.A.R.

### References

- Altenberg, B. (1987). *Prosodic patterns in spoken English: Studies in the correlation between prosody and grammar for text-to-speech conversion*. Lund: Lund University Press.
- Baker, A. (2009). *Ship or sheep? An intermediate pronunciation course* (3rd edition). Beijing: Beijing Language and Culture University Press.
- Beckman, M. E. (1986). *Stress and non-stress accents*. Dordrecht, the Netherlands: Foris Publications.
- Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Version 6.0.39, retrieved 10 April 2018 from <http://www.praat.org/>
- Bolinger, D. L. (1958). Intonation and grammar. *Language Learning*, 8(1), 31-37.
- Bolinger, D. L. (1986). *Intonation and its parts: melody in spoken English*. London: Edward Arnold.
- Chao, Y. R., (1979). *A grammar of spoken Chinese*. Bei Jing: China. Commercial press.
- Chen, R. (1995). Communicative dynamism and word order in Mandarin Chinese. *Language Sciences*, 17, 201-222.
- Chen, Y. (2003). *The phonetics and phonology of contrastive focus in standard Chinese* (Unpublished PhD dissertation). Stony Brook: State University of New York.
- Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in standard Chinese. *Journal of Phonetics*, 36, 724-746.
- Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam: John Benjamins Publishing Company.
- Ciszewski, T. (2012). Stressed vowel duration and phonemic length contrast. *Research in Language*, 10(2), 201-214.
- Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, 1(2), 425-452.
- Cole, J., & Shattuck-Hufnagel, S. (2016). New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology*, 7(1), 1-29.
- Cruttenden, A. (1997). *Intonation* (2nd edition). Cambridge: Cambridge University Press.
- Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

- Deterding, D. (2006). The pronunciation of English by speakers from China. *English World-Wide*, 27(2), 175-198.
- Dickerson, W. (1989). *Stress in the speech stream: The rhythm of spoken English*. Urbana, IL: University of Illinois Press.
- Fouz-González, J. (2015). Trends and directions in computer-assisted pronunciation training. In J.A. Mompean & J. Fouz-Gonzalez (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 174-195). Basingstoke, UK: Palgrave Macmillan.
- Frost, D. (2011). Stress and cues to relative prominence in English and French: A perceptual study. *Journal of the International Phonetic Association*, 41(1), 67-84.
- Gray, M. (2015). Training L1 French learners to perceive prosodically marked focus in English. In J.A. Mompean & J. Fouz-Gonzalez (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 174-195). Basingstoke, U.K.: Palgrave Macmillan.
- Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38(2), 201-223.
- Ingels, S. A. (2011). *The effects of self-monitoring strategy use on the pronunciation of learners of English* (Unpublished PhD dissertation). Urbana-Champaign: University of Illinois.
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23(1), 83-103.
- Jin, S. (1996). *An acoustic study of sentence stress in Mandarin Chinese* (Unpublished Ph.D. dissertation). Ohio: The Ohio State University.
- Jones, D. (1956). *The pronunciation of English*. Cambridge: Cambridge University Press.
- Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *International Review of Applied Linguistics*, 28(2), 99-117.
- Kabagema-Bilan, E., Lopez-Jimenez, B., & Truckenbrodt, H. (2011). Multiple focus in Mandarin Chinese. *Lingua*, 121, 1890-1905.
- Ladd, D. R. (1980). *The structure of intonational meaning*. Bloomington, IN: Indiana University Press.
- Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, 33(1), 37-63.
- Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustic Society of America*, 32, 451-454.
- Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica*, 62, 70-87.
- Low, E. L., & Grabe, E. (1999). A contrastive study of prosody and lexical stress placement in Singapore English and British English. *Language and Speech*, 42(1), 39-56.

- Lu, J., Wang, R., & de Silva, L. C. (2012). Automatic stress exaggeration by prosody modification to assist language learners perceive sentence stress. *International Journal of Speech Technology*, 15, 87-98.
- Luchini, P. L. (2005). A new approach to teaching pronunciation: An exploratory case study. *The Journal of Asia TEFL*, 2(2), 35-62.
- Mennen, I. (2006). Phonetic and phonological influences in non-native intonation: An overview for language teachers. *Working paper WP-9*. Edinburgh, UK: QMUC Speech Science Research Centre.
- Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory. In E. Delais-Roussarie, M. Avanzi & S. Herment (Eds.), *Prosody and language in contact: L2 Acquisition, Attrition and Languages in Contact* (pp. 171-188). Heidelberg: Springer Verlag Berlin.
- Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *Modern Language Journal*, 84(3), 372-389.
- Pintér, G., Mizuguchi, S., & Tateishi, S. (2014). Perception of prosodic prominence and boundaries by L1 and L2 speakers of English. In *Proceedings of INTERSPEECH 2014* (pp. 544-547).
- Roach, P. (1991). *English phonetics and phonology: A practical course* (2nd edition). Cambridge: Cambridge University Press.
- Roux, G., Bertrand, R., Ghio, A., & Astésano, C. (2016). *Naïve listeners' perception of prominence and boundary in French spontaneous speech*. Paper presented at Speech Prosody 2016, Boston, MA, USA.
- Selkirk, E., & Shen, T. (1990). Prosodic domains in Shanghai Chinese. In S. Inkelas & Z. Draga (Eds.), *The phonology-syntax connection* (pp. 313-338). Stanford: Stanford University.
- Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory, Number 3: Stress, Tone and Intonation* (pp. 83-109), Cornell University.
- Sluijter, A. M. C., van Heuven, V. J., & Pacilly, J. J. A. (1997). Spectral balance as a cue in the perception of linguistic stress. *Journal of Acoustic Society of America*, 101, 503-513.
- Smith, C. (2009). Naïve listeners' perception of French prosody compared to the predictions of theoretical models. *Proceedings of IDP 09* (pp. 335-349).
- Smith, C. (2011). *Perception of prominence and boundaries by naive French listeners*. Paper presented at ICPHS XVII, Hong Kong, 2011.
- Smith, C. (2013). French listeners' perceptions of prominence and phrasing are differentially affected by instruction set. *Proceedings of Meetings on Acoustics* (pp. 1-7).
- Smith, C., & Edmunds P. (2013). Native English listeners' perceptions of prosody in L1 and L2 reading. *Proceedings of INTERSPEECH 2013* (pp. 235-238).
- Swerts, M., & Kraemer, E. (2004). Congruent and incongruent audiovisual cues to prominence. *Proceedings of Speech Prosody 2004* (pp. 69-72).

- Tamburini, F., & Caini, C. (2005). An automatic system for detecting prosodic prominence in American English continuous speech. *International Journal of Speech Technology*, 8(1), 33-44.
- Wells, J. (2006). *English intonation: an introduction*. Cambridge: Cambridge University Press.
- Xu, L. (2004). Manifestation of information focus. *Lingua*, 114 (3), 277-299.
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 27, 55-105.
- Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33, 159-197.
- Yu, V. Y., & Andruski, J. E. (2011). The effect of language experience on perception of stress typicality in English nouns and verbs. *The Metrical Lexicon*, 6(2), 275-301.
- Yuan J. (2004). *Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling* (Unpublished Ph.D. dissertation). Cornell University.

### Appendix: Stimuli Sentences (Adapted from Wells (2006))

The underlined are the syllables that tend to carry nuclear stress in the sentences. The types of sentence are indicated in parentheses.

#### Sentences in isolation

- 1 I've just received a letter from her. (Ending with a function word)
- 2 You've told me what Emma wants, (Contrastive sentence Part 1)
- 3 what do you want? (Contrastive sentence Part 2)
- 4 I'm going to buy a new mobile phone. (Ending with an early-stressed compound)
- 5 Shall we walk to the restaurant? (Default pattern)
- 6 I'd prefer to go on foot. (Ending with repeated information)
- 7 You're looking rather pleased with yourself. (Ending with a reflexive pronoun)
- 8 How are you doing, he asked. (Ending with a reporting phrase)
- 9 I'll see you on Tuesday, then. (Ending with a parenthetical)
- 10 Let's go back to my place. (Ending with an empty word)
- 11 There's a mosquito on your finger. (Ending with a place adverbial)
- 12 What are you looking at? (Ending with a phrasal verb with a preposition)
- 13 Look at the tie he's wearing. (Ending with a noun modifier)
- 14 There's a train coming. (Event sentence)
- 15 Which route did you take? (Wh-object sentence)

**Sentences in context**

- 16 Are you planning to go away this year? (Ending with a time adverbial)
- 17 We've just been away. (Contrastive sentence)
- 18 We had a week in Cornwall. (Default pattern)
- 19 How was it? (Ending with a function word)
- 20 We had a marvelous time. (Ending with an empty word)
- 21 The only problem was the weather. (Default pattern)
- 22 It rained most of the time. (Ending with a time adverbial)
- 23 What did you do during all this rain? (Ending with repeated information)
- 24 The best thing we did was to go to the Eden Project. (Ending with an early-stressed compound)
- 25 What's that? (Default pattern)
- 26 It's a museum of ecology. (Default pattern)
- 27 I found it utterly fascinating. (Default pattern)
- 28 It's more like a theme park really. (Ending with an early-stressed compound and a parenthetical)
- 29 There's lots to do. (Ending with a noun modifier)
- 30 The children loved it (too). (Contrastive sentence)





## Effects of Semantic Information and Segmental Familiarity on Learning Lexical Tone

Angela Cooper  
Simon Fraser University, Canada

Yue Wang  
Simon Fraser University, Canada

Dawn M. Behne  
Norwegian University of Science and Technology, Trondheim

### Abstract

Languages such as Mandarin which utilize tone to contrast word meaning can present a challenge for learners whose native language does not use pitch contrastively. Acquiring tone words requires learners to contend with multiple dimensions of information, including segmental, tonal and semantic. The present work examined how these segmental and semantic dimensions influence the acquisition of non-native (L2) lexical tones. Native English participants completed Mandarin tone training where semantic information was either present or absent, and where the segments were familiar or unfamiliar to listeners. Pre- and post-test tone identification results revealed that L2 tone learning was inhibited for listeners who received semantic information during training; however, segmental familiarity did not significantly impact tone learning. These findings suggest that, at least at an initial learning stage, alleviating learners' processing load by reducing the number of dimensions of information provided during training facilitates the acquisition of L2 phonemic contrasts.

## **1. Introduction**

As language users acquire their native language (L1), the acoustic information relevant to phonemic distinctions within the L1 is weighted more heavily than less relevant information (e.g., Werker & Tees, 1984). Having been tuned to L1 phonetic information can be a formidable challenge for adult non-native (L2) learners when re-tuning their perceptual systems to the relevant acoustic cues necessary for discerning L2 phonemic distinctions (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997). Previous research has investigated a multitude of factors mediating the acquisition of L2 phonemic distinctions, with a particular focus on how L1 and L2 phonemic categories are perceptually related to one another (e.g., Best, 1995; Flege, 1995). While the majority of prior literature has focused on L2 segmental contrasts (e.g., Beddor & Strange, 1982; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Hallé & Best, 2007; Polka, 1991), a growing body of research has investigated the acquisition of L2 suprasegmental distinctions, specifically lexical tone (e.g., Gottfried & Suiter, 1997; Hallé, Chang, & Best, 2004; Wang, Spence, Jongman, & Sereno, 1999). In acquiring L2 tone words, learners must contend with the tonal contrasts as well as any novel segments within the syllable. Moreover, in addition to mastering the phonemic (tonal and segmental) components, learners must also map the phonemic form to a specific meaning. These multiple layers of linguistic information may result in an increased processing load for learners, which could potentially inhibit the acquisition process. The aim of the present study is to examine how these different dimensions of information (segmental, semantic) influence the acquisition of L2 lexical tones.

### **1.1. Processing load in L2 speech learning**

The automatic selective perception (ASP) model posits that the online processing of L2 sounds, particularly by late L2 learners, requires listeners to expend more cognitive resources in order to extract the necessary phonetic information to differentiate the contrasts than native language processing (Strange, 2011). According to this account, listeners process the auditory speech stream in one of two modes (or “ways of perceiving”, p. 460), phonological or phonetic, depending on a variety of factors including the listeners’ linguistic knowledge, the nature of the stimuli and task demands. The phonological mode is characterized as an automatic process, typically employed by adult listeners processing their L1. When in the phonological mode, listeners are posited to “ignore” context-dependent variation arising from, for instance, speaking rate or minor dialect differences, enabling

them to focus on and efficiently extract enough phonologically-relevant information sufficient to identify the appropriate word form. The phonetic mode of processing, on the other hand, involves focusing on context-specific phonetic information, where L1 listeners retrieve stored allophonic and phonotactic information, allowing them to adjust to an unfamiliar accent, for example. Compared with the phonological mode, the phonetic mode of processing is posited to involve more attentional focus and cognitive resources. L2 listeners are argued to utilize the phonetic mode of processing in the early stages of acquisition.

Despite the challenges of processing L2 contrasts, prior research has found that listeners' perception of non-native segmental and suprasegmental contrasts can improve with laboratory training, demonstrating that human perceptual systems retain a degree of plasticity over the lifespan (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Iverson, Hazan, & Bannister, 2005; Wang et al., 1999). If L2 language processing is taxing on the perceptual system, which can manifest as impaired comprehension in non-optimal listening conditions (e.g., Bradlow & Alexander, 2007), then alleviating the processing load during training, at least at the initial stage of learning, would likely enable listeners to allocate the necessary attentional resources to focus on the relevant phonetic details of the contrast they are trying to acquire. Indeed, prior work has demonstrated reduced identification accuracy of non-native pitch contours under high cognitive load conditions, specifically for listeners with relatively poorer perceptual abilities (Antoniou & Wong, 2015). One way to relieve the L2 language processing load could involve explicitly orienting listeners' attention to the appropriate phonetic information during training (Guion & Pederson, 2007; Hisagi & Strange, 2011; Pederson & Guion-Anderson, 2010). For example, Hisagi and Strange (2011) tested American English listeners' explicit versus implicit learning of temporally-cued contrasts in Japanese, manipulating whether or not listeners' attention was directed to the critical dimension. Listeners who were explicitly instructed to focus on the critical durational differences performed significantly better than those who did not receive such instructions.

When learning an L2, acquiring the ability to differentiate the phonemic contrasts of that language has the specific functional goal of distinguishing word forms and their associated meanings. That is, a native Japanese learner of English, for example, needs to learn to distinguish the phonemes /ɹ/ and /l/ in order to be able to form separate lexical entries for "rock" [ɹɔk] and "lock" [lɔk]. Acquiring an L2 lexicon, therefore,

involves encoding not only the relevant phonemic information about word forms but also their semantic information (i.e., word meanings). However, providing semantic information whilst attempting to acquire an L2 phonemic contrast and/or their associated phonetic differences may increase the processing load for the learner. Guion and Pederson (2007) found initial evidence in support of this notion. Native English listeners who were explicitly instructed to learn the meanings of words distinguished by Hindi stop contrasts performed poorer on a subsequent discrimination task relative to listeners who were instructed to attend to the specific Hindi stop sounds. The authors argue that actively attending to the semantic information resulted in increased processing load for the meaning-group and could have thus interfered with the learning of fine phonetic details. A similar proposal has been suggested for young infants acquiring words distinguished by native language contrasts (Stager & Werker, 1997). Fourteen-month-olds failed to detect phonetic detail in a word learning context, which they were capable of detecting in a syllable discrimination context. The computational demands of associating word forms with objects may divert resources away from processing lower-level phonetic information.

In contrast, some studies have suggested that lexical knowledge can actually facilitate the acquisition of phonemic contrasts for both young children and adults (Davidson, Shaw, & Adams, 2007; Hayes-Harb, 2007). For example, Hayes-Harb (2007) showed that native speakers' discrimination of a novel phonemic distinction in English (voiceless unaspirated stop [k] vs. prevoiced stop [g]) was improved by the inclusion of semantic information as compared to learners who received only auditory information about the contrast. Similarly, providing object referents were found to facilitate discrimination of the Hindi dental-retroflex contrast for 9-month-old English-learning infants (Yeung & Werker, 2009). Implicit semantic learning was posited to reinforce listeners' awareness that the subtle acoustic variations were in fact linguistically-relevant. The association of speech with categorical cues, such as distinct objects, may guide listeners to extract the relevant acoustic cues to help distinguish difficult phonetic contrasts.

Based on these prior findings, it is not clear whether the inclusion of semantic information during phonetic training inhibits or facilitates the learning of non-native contrasts. This discrepancy may relate to what task is required of listeners during training and at test. In Guion and Pederson (2007), listeners in a meaning-attending group were explicitly asked to learn sound-meaning pairings during training and then tested on their

ability to discriminate the sound contrasts (devoid of meaning information). In Hayes-Harb (2007), however, semantic information was present as an additional component during training, but listeners were not required to focus on learning the sound-meaning pairings. Not being asked to focus predominantly on the meaning could have freed up some attentional resources to extract some information about phonetic form.

### **1.2. L2 tone learning**

In addition to segmental contrasts, in lexical tone languages, identical syllables that differ in average fundamental frequency ( $f_0$ , perceived as pitch) or  $f_0$  contour can have distinct meanings (Yip, 2002). For example, in Mandarin Chinese, four distinct pitch contours are phonemically contrastive: 1) high-level, 2) high-rising, 3) low-dipping, and 4) high-falling (Chao, 1948). Similar to the challenges faced by segmental contrasts for L2 listeners, studies have shown that non-tone language listeners can find it difficult to identify and discriminate L2 lexical tone contrasts, though learners have been shown to improve their perception following perceptual training (e.g., Francis, Ciocca, Ma, & Fenn, 2008; Wang et al., 1999; Wayland & Guion, 2004; Wayland & Li, 2008). Moreover, while listeners are capable of improving their ability to distinguish non-native lexical tones, a variety of factors, including training structure, task demands and differences in individual abilities have been found to influence tone learning success. For example, variation during training, including talker variation (Perrachione, Lee, Ha, & Wong, 2011) and irrelevant variation of non-target phonetic features (Antoniou & Wong, 2016), can hinder perceptual learning, particularly for learners with poor perceptual abilities. Furthermore, in line with segmental work, attention has been found to be a significant factor in the acquisition of L2 lexical tones. Chandrasekaran, Yi, Smayda, and Maddox (2016) reported that focusing learners' attention on pitch direction specifically led to enhanced category learning of Mandarin lexical tones relative to attending to pitch height or no explicit instructions.

It is important to note, however, that tone learning differs from other types of perceptual learning, in that listeners must concurrently incorporate tonal, segmental and potentially semantic information, which may further increase the processing load for listeners. Given prior studies suggesting that increased processing load can worsen performance on L2 contrasts (e.g., Antoniou & Wong, 2015; Guion & Pederson, 2007), the need to incorporate three layers of information could make tone word learning particularly challenging for L2 listeners.

With respect to semantic and tonal information, Cooper and Wang (2012) explicitly trained listeners on distinguishing the meanings of Cantonese tone words, requiring implicit L2 tone learning, and found that listeners significantly improved their tone identification. On the other hand, further research showed that initial training explicitly focusing on tone (compared to the absence of such tone-only training) could enhance later learning of tone words (Cooper & Wang, 2013; Ingvalson, Barr, & Wong, 2013). These results indicate strong connections across tonal, segmental and semantic information processing during tone word learning. Together, these findings indicate a need for further studies testing the issue of processing load by directly comparing training which manipulates these different levels of processing.

Furthermore, regarding segmental and tonal information, many prior tone training studies utilize syllables containing segments from the listeners' L1, the implicit assumption being that unfamiliar L2 segments would have a negative influence on listeners' tone perception (Cooper & Wang, 2012; Francis et al., 2008; Hallé et al., 2004; Wayland & Li, 2008). Indeed, research investigating the integrality or separability of consonant, vowel and tone dimensions during speech processing has found that these dimensions are perceptually integrated for native Mandarin Chinese listeners; that is, when attempting to classify lexical tones, listeners were unable to ignore vowel or consonant variability (Lin & Francis, 2014; Tong, Francis, & Gandour, 2008). However, other research has shown integrated processing of tone and rime (vowels), but separate processing of tone and consonants by native English as well as native Mandarin listeners (Lin & Francis, 2014; Sereno & Lee, 2015). Subsequent questions thus arise about the contribution of segmental information to processing load during the acquisition of L2 tonal contrasts. Specifically, does the presence of unfamiliar segmental information (non-existent in L1) increase the processing load for listeners attempting to focus on tonal information, as listeners may be trying to process and categorize both segmental and suprasegmental components?

### **1.3. The present study**

The acquisition of L2 lexical tones is a unique case to test the role of processing load during perceptual learning, as learning words minimally contrasted by tone involves tonal, segmental, and semantic information, allowing us to examine both the separate and the combined effects of

these three factors, an approach which has not previously been explored. The present study investigated the hypothesis that alleviating learners' processing load would facilitate the acquisition of L2 phonemic contrasts, at least at early stages of L2 acquisition. If L2 listeners operate in a phonetic mode of processing when perceiving L2 speech, they should require more cognitive and attentional resources (Strange, 2011); therefore, providing training that may reduce processing load should enable them to devote sufficient resources to learn the relevant phonemic contrast. This issue was examined by either providing semantic information or only tonal information (Experiment 1), and the use of familiar or unfamiliar initial segments (Experiment 2) during the perceptual training of Mandarin lexical tones by native English listeners. That is, the task involved explicit L2 lexical tone learning and manipulated the implicit processing of semantic and segmental information.

## **2. Experiment 1: Role of semantic information in tone learning**

In the first experiment, two groups of native English listeners were administered a Mandarin tone training program, which either provided meanings for the words (Meaning group) or did not (Tone Only group). An identification task before and after training was used to assess improvement in identifying L2 lexical tones. By not including meaning as an extra information channel in the Tone Only condition, processing load may be reduced and facilitate learning; in which case, after training, participants in the Tone Only group should outperform the Meaning group. However, if providing semantic information reinforces that the  $f_0$  distinctions are lexically contrastive, then the Meaning group would be expected to outperform the Tone Only group,

### **2.1. Methods**

#### **2.1.1. Participants**

Twenty-six native Canadian English speakers were included in this study, with no prior experience with Mandarin or another lexical tone language. They self-reported normal hearing and had no musical experience within the last five years and less than 2 years of musical experience prior to that (e.g., Cooper & Wang, 2012; Wong, Skoe, Russo, Dees, & Kraus, 2007). Fourteen participants were included in the Tone Only group (nine females; *M age*=23 years) and 12 in the Meaning group (10 females; *M age*=21 years).



### 2.1.2. Stimuli

The stimuli used in the pre- and post-test tone identification (ID) task were 12 Mandarin monosyllables with four Mandarin tones, for a total of 48 tone words (Table I), all of which were produced by both of two native Mandarin speakers (1 male, 1 female). Half of the syllables contained initial consonants familiar to English, and half contained initial consonants that were unfamiliar. For the training phase, a second pair of Mandarin speakers (1 male, 1 female) each produced a different set of 6 Mandarin monosyllables for each of the four tones (Table I), containing initial consonants familiar and unfamiliar to English. Stimuli were recorded at a 44.1 kHz sampling rate using a SHURE KSM109 microphone in a sound-attenuated booth in the Language and Brain Lab at Simon Fraser University. They were RMS amplitude normalized to 65 dB and presented at a comfortable listening volume.

TEST SYLLABLES	
Familiar segment	Unfamiliar segment
<i>ka</i> [ka]	<i>zhuo</i> [tʂo]
<i>pou</i> [pou]	<i>xiong</i> [ɕion]
<i>fu</i> [fu]	<i>run</i> [ɹun]
<i>lan</i> [lan]	<i>zi</i> [tsi]
<i>nin</i> [nin]	<i>que</i> [tɕ <sup>h</sup> ue]
<i>ting</i> [tiŋ]	<i>chi</i> [tʂ <sup>h</sup> i]
TRAINING SYLLABLES	
Familiar segment	Unfamiliar segment
<i>ming</i> [miŋ]	<i>ri</i> [ɹi]
<i>yao</i> [jao]	<i>chun</i> [tɕun]
<i>te</i> [te]	<i>qiong</i> [tɕ <sup>h</sup> ion]
<i>wa</i> [wa]	<i>xue</i> [ɕue]
<i>kai</i> [k <sup>h</sup> ai]	<i>cuo</i> [ts <sup>h</sup> uo]
<i>lao</i> [lao]	<i>zhi</i> [tʂi]

Table 1. Syllables used in the pre-/post-test identification and training tasks.

For the Meaning group, the pre-/post-test and training sets of tone words were assigned meanings corresponding to common concrete nouns and represented by pictures, selected from a standardized set of 260 pictures, controlled for visual complexity and cultural familiarity (Snodgrass & Vanderwart, 1980). Figure 1 displays sample pictures presented to the Meaning group.



Figure 1. Sample pictures presented to the Meaning group.

### **2.1.3. Procedure**

Table II depicts an overview of the experimental setup for the test and training days. The pre- and post-training tone ID tests began with a two-part familiarization followed by the main task. In the first part of the familiarization, participants heard the four Mandarin syllables with tones individually while viewing its tone diagram displayed on 15-inch LCD monitors (Tone Familiarization). In the second part of familiarization, participants practiced the 4-alternative forced choice ID task, identifying the tone they heard by pressing the number corresponding to the appropriate visual depiction of its tonal pitch contour (tone diagram) and receiving feedback on the accuracy of their response as well as the correct answer (Task Familiarization). The familiarization task used productions of /fa/ by the female pre-/post-test talker (12 trials total). The main task was identical to the second part of the familiarization but without feedback, whereby listeners heard an item and identified the tone by pressing the number corresponding to the tone diagram. Participants identified 96 randomized stimuli (12 syllables x 4 tones x 2 speakers), presented with an inter-stimulus-interval (ISI) of 3 seconds.

<b>Pre-test</b>	<b>Training</b>			<b>Post-test</b>
Tone and Task Familiarization	<b>Session 1</b>	<b>Session 2</b>	<b>Session 3</b>	Same as Pre-test
	Training (2 blocks)	Training (2 blocks)	Training (2 blocks)	
Tone identification of 96 stimuli not used in training	Training test	Training test	Training test	
	<ul style="list-style-type: none"> <li>• The Meaning group viewed pictures associated with each tone word. The Tone Only group viewed a fixation cross.</li> <li>• Each training session contained 192 items (6 syllables x 4 tones x 2 speakers x 4 repetitions).</li> <li>• The training test contained stimuli received during training.</li> </ul>			

Table 2. Overview of experimental setup for test and training sessions.

The training program consisted of three separate training sessions within a 10-day period. Each training session consisted of two blocks followed by a training test. Each block began with a brief overview of the 4 tones, where listeners would hear each tone individually and view its associated tone diagram. Each block contained a different set of 12 training words (3 syllables x 4 tones x 2 speakers x 4 repetitions = 96 trials), presented with a 2-second ISI. Thus, each training session contained 192 trials for the two blocks. For the Meaning group, the assigned meaning of each item was depicted on the screen while the audio stimulus was played. Participants were not required to memorize the associated meanings of the pictures but were simply informed that each picture represented the meaning of the

item they heard. For the Tone Only group, a fixation cross was displayed during stimulus presentation. Training was similar to the familiarization task for the pre-/post-tests, whereby listeners responded to each stimulus by indicating the tone they heard, receiving feedback on the accuracy of their response. Feedback for the Meaning group consisted of seeing the assigned meaning of the item and tone number displayed, while feedback for the Tone Only group involved a display of the tone diagram and tone number. After both phases, participants completed a training test, identical in format to the pre-/post-training ID tests. They were tested on the 24 training words they received during training words they received during training (6 syllables x 4 tones x 2 speakers x 2 repetitions). All tasks were administered via E-Prime 1.0 on PC computers using AKG K1441 Studio headphones.

## **2.2. Results**

### **2.2.1. Pre-/post-test tone identification**

Tone identification accuracy on the pre- and post-training tone ID tests was calculated for each group (Figure 2) and submitted to logistic linear mixed effects regression (LMER) with contrast coded as a fixed effect of Training Type (Meaning vs. Tone Only), as well as a fixed effect for Test (Pre, Post) and their interaction. Random intercepts for Tone Word (each tone+syllable pairing) and Participant were included, as well as random slopes for Test by Participant and Training Type by Tone Word.

A significant main effect of Test ( $\beta=1.49$ ,  $SE \beta=0.15$ ,  $\chi^2(1)=42.99$ ,  $p<0.001$ ) was obtained, with listeners improving from pre-test (Mean proportion correct ID,  $M=0.32$ ) to post-test ( $M=0.65$ ). No effect of Training Type was obtained ( $p=0.35$ ); however, a significant Training Type x Test interaction was found ( $\beta=-0.76$ ,  $SE \beta=0.29$ ,  $\chi^2(1)=5.97$ ,  $p=0.01$ ). Follow-up LMERs for each test with Training Type as a fixed effect revealed no significant difference at pre-test ( $p=0.25$ ) but a marginally significant difference at post-test ( $\beta=-0.55$ ,  $SE \beta=0.28$ ,  $\chi^2(1)=3.68$ ,  $p=0.055$ ), suggesting a tendency for the Meaning group to perform less accurately than the Tone Only group at identifying lexical tones following training.

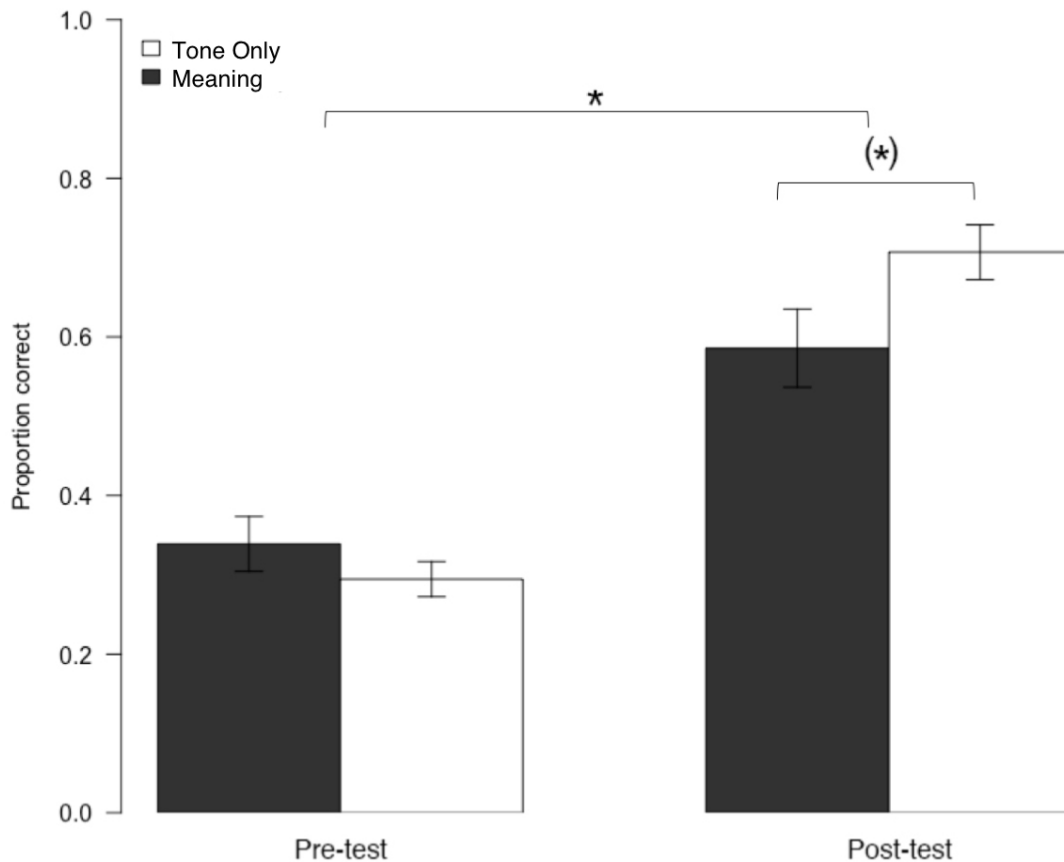


Figure 2. Mean proportion correct tone identification for Pre-test and Post-test by Training Group (Tone Only, Meaning). Asterisks denote significance ( $p < 0.05$ ), and asterisks in parentheses indicate marginal significance. Error bars indicate  $\pm 1$  standard error.

### 2.2.2. Training

To examine participants' trajectory of improvement over the course of training, tone ID accuracy scores for each training test (Figure 3) were submitted to a logistic LMER with Helmert-contrast coded as fixed effects for Session (A: 1 vs. 2 + 3; B: 2 vs. 3), a fixed effect for Training Type (Meaning vs. Tone Only), and their interactions. The Helmert coding, which is often utilized in cases where the levels of a categorical variable are ordered, for instance, from lowest to highest, reflected our prediction that listeners would improve as a result of training, with levels ordered from low (Training Session 1) to high (Training Session 3). Random intercepts for Participant and Tone Word were included, as well as by-participant random slopes for Session A and B.

Significant main effects of Session A ( $\beta=1.5$ ,  $SE \beta=0.14$ ,  $\chi^2(1)=42.09$ ,  $p<0.001$ ) and Session B ( $\beta=0.41$ ,  $SE \beta=0.11$ ,  $\chi^2(1)=12.26$ ,  $p<0.001$ ) were found, indicating that across groups, listeners were significantly improving after each training session. A significant main effect of Training Type was also obtained ( $\beta=2.02$ ,  $SE \beta=0.39$ ,  $\chi^2(1)=15.65$ ,  $p<0.001$ ), with the Meaning group ( $M=0.91$ ) significantly outperforming Tone Only group ( $M=0.70$ ) over the course of training. Finally, a significant Session A (1 vs. 2 + 3) x Training Type interaction was found ( $\beta=1.2$ ,  $SE \beta=0.29$ ,  $\chi^2(1)=13.07$ ,  $p<0.001$ ), with a significantly larger difference between the Meaning group relative to Tone Only group after the first session of training as compared to sessions 2 and 3, with superior performance by the Meaning group. The remaining interaction did not reach significance ( $\chi^2=2.22$ ,  $p=0.14$ ).

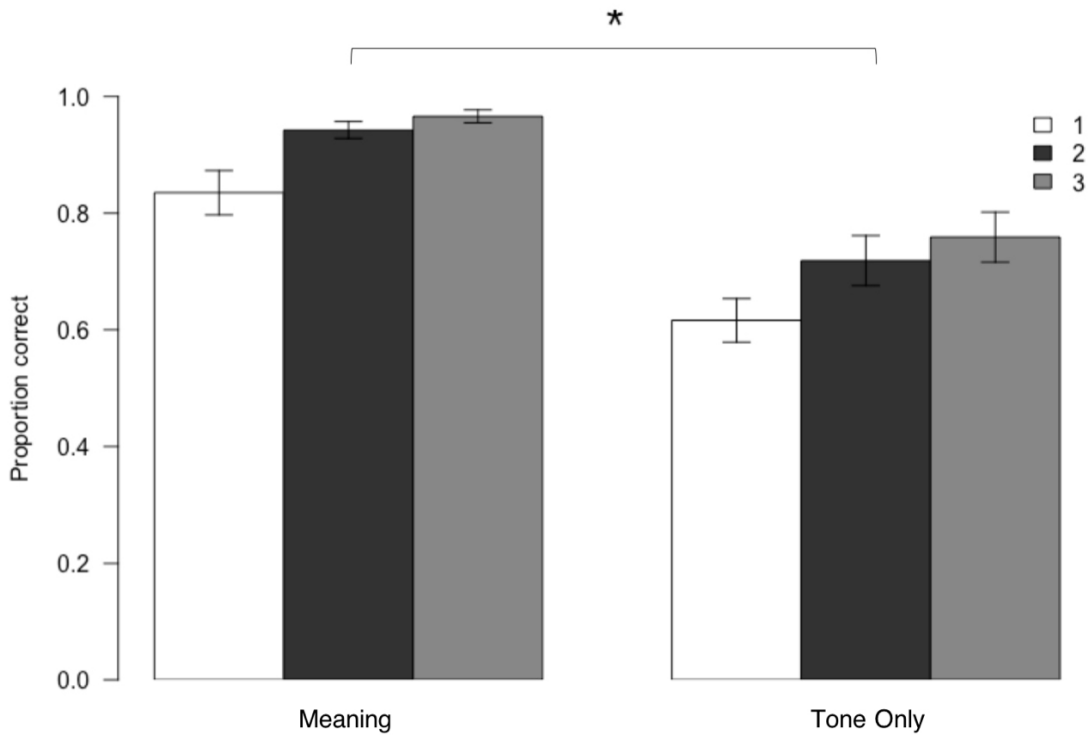


Figure 3. Mean proportion correct tone identification by training session (1-3) and training type (Meaning vs. Tone Only). The asterisk denotes significance ( $p<0.05$ ). Error bars indicate +/- 1 standard error.

Overall, the results of Experiment 1 indicate that providing semantic information during training facilitated the acquisition of the specific items used during training, as indicated by superior performance by the Meaning group on the training tests administered at the end of each training day. However, the inclusion of semantic information during the training phase appeared to ultimately inhibit the formation of generalizable tone categories, as the magnitude of improvement from pre- to post-test was smaller for the Meaning group relative to the Tone Only group.

### **3. Experiment 2: Segmental familiarity in tone learning**

In a second experiment, we compared the effects of tone training using segments familiar and unfamiliar to English trainees. When attempting to extract information about the nature of  $f_0$  contrasts in a new language, having to also process unfamiliar segmental information (non-existent in their L1) may increase the processing load for L2 listeners and thereby inhibit tone learning. This would predict that listeners who undergo tone training with syllables containing (familiar) segments existent in their L1 would outperform listeners trained on unfamiliar segments.

Given that tone and rime (vowel) dimensions are more integrally processed than tone and consonant dimensions (Serenio & Lee, 2015), initial consonants rather than vowels were manipulated in the present experiment, as a more separate dimension (i.e., consonant relative to vowel) would allow us to determine the effects the processing load of unfamiliar segments on lexical tone processing.

#### **3.1. Methods**

##### **3.1.1. Participants**

Thirteen native English listeners (9 females; *M age*=22 years) who did not participate in Experiment 1 but satisfied the same inclusion criteria as in Experiment 1 were recruited to receive tone training using segments familiar to them in English (Familiar group). Their results are compared to those from the fourteen participants in the “Tone Only” group in Experiment 1, since the training stimuli in Experiment 1 contained segments non-existent in English. In this experiment, this group is referred to as the Unfamiliar group.

##### **3.1.2. Stimuli and Procedure**

The pre- and post-test tone ID task was identical to Experiment 1, which included 48 items (12 syllables x 4 tones) produced by two speakers, half of which contained segments familiar to English listeners, and half

that were unfamiliar (Appendix A). For training, the same Mandarin training speakers as in Experiment 1 produced a new set of 6 Mandarin monosyllables with 4 lexical tones, containing initial consonants existent in English (e.g., [fu], [nin], [miŋ]), used for the “Familiar” training group. The training stimuli used for the “Unfamiliar” group (from Experiment 1) contained initial consonants specific to Mandarin (e.g., [ʃo], [ɰun], [ɕue], Appendix B). The total number of stimuli for each training session used in both groups was the same: 6 syllables x 4 tones x 2 speakers x 4 repetitions = 192. The length and format of training as well as training task procedure and feedback were the same as the Tone Only group in Experiment 1.

### 3.2. Results

#### 3.2.1. Pre-/post-test tone identification

Tone identification accuracy was calculated and compared with the performance of the Unfamiliar and Familiar groups (Figure 4) with a logistic LMER containing contrast-coded fixed effects for Segment Type (Familiar vs. Unfamiliar) and Test (Pre, Post) and their interaction, with random intercepts for participant and item, and a by-participant random slope for Test and by-item random slope for Segment Type.

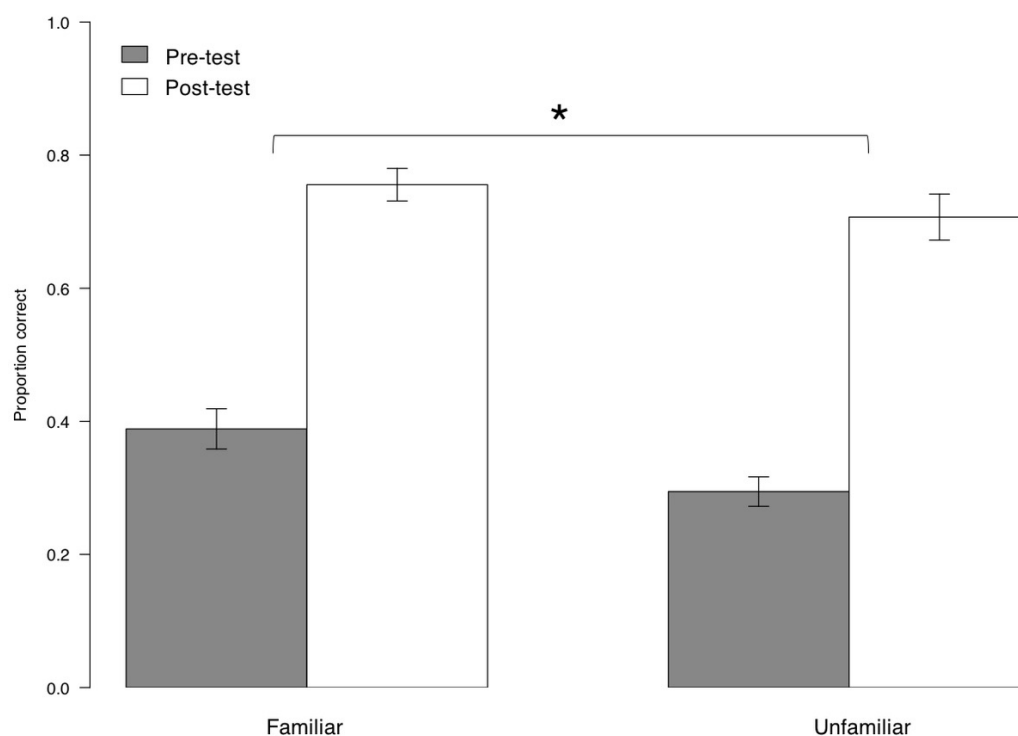


Figure 4. Mean proportion correct tone identification for Pre-test and Post-test by Segment Type (Familiar vs. Unfamiliar). The asterisk denotes significance ( $p < 0.05$ ) and error bars indicate  $\pm 1$  standard error.



A significant main effect of Test was found ( $\beta=1.76$ ,  $SE \beta=0.12$ ,  $\chi^2(1)=61.85$ ,  $p<0.001$ ), indicating an overall increase in listeners' ability to identify non-native tones following training. Segment Type was also a significant factor ( $\beta=0.34$ ,  $SE \beta=0.16$ ,  $\chi^2(1)=4.35$ ,  $p=0.04$ ), with listeners in the Familiar group ( $M=0.57$ ) outperforming the Unfamiliar group ( $M=0.50$ ) across pre- and post-tests. No Segment Type x Test interaction was found ( $\chi^2=0.67$ ,  $p=0.41$ ).

### 3.2.2. Training

Similar to analyses in Experiment 1, the Familiar group's tone identification performance following each training session was tabulated and compared to the Unfamiliar group (Figure 5). A logistic LMER was conducted with a contrast-coded fixed effect of Segment Type (Familiar vs. Unfamiliar) and Helmert contrast-coded fixed effect of Session (A: 1 vs. 2 + 3, B: 2 vs. 3) and their interactions, along with random intercepts for participant and item, and a by-participant random slope for Session and by-item random slope for Segment Type.

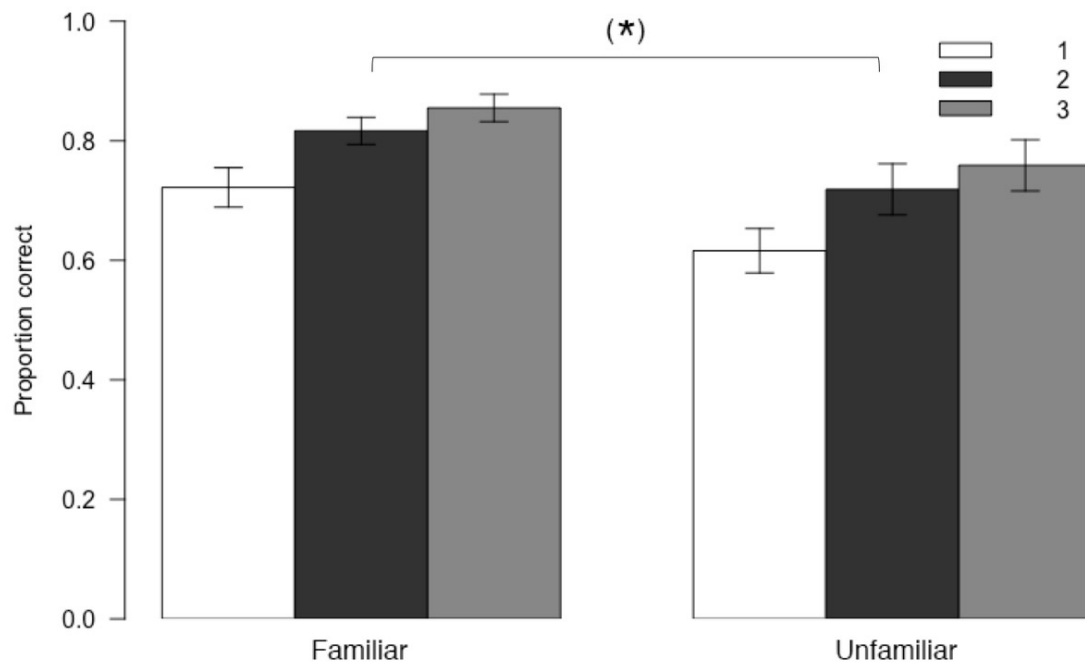


Figure 5. Mean proportion correct tone identification by training session (1-3) and training segment type (Familiar vs. Unfamiliar). The asterisk in parentheses denotes marginal significance ( $p=0.06$ ) and error bars indicate  $\pm 1$  standard error.

Significant effects of Session A ( $\beta=0.91$ ,  $SE \beta=0.11$ ,  $\chi^2(1)=34.26$ ,  $p<0.001$ ) and Session B ( $\beta=0.27$ ,  $SE \beta=0.08$ ,  $\chi^2(1)=9.86$ ,  $p=0.002$ ) were yielded, indicating that across groups, listeners' tone identification performance significantly improved after each training session. A marginally significant effect of Segment Type was obtained ( $\beta=0.53$ ,  $SE \beta=0.28$ ,  $\chi^2(1)=3.46$ ,  $p=0.06$ ), with the Familiar group outperforming the Unfamiliar group across training sessions. No significant interactions were found ( $\chi^2<0.13$ ,  $p>0.72$ ).

The results revealed that while listeners trained on items containing familiar consonants had overall higher tone identification accuracy during training and across pre- and post-tests, the amount of improvement as a result of training did not surpass listeners who were trained with unfamiliar segments.

### **3.2.3. Cross-Experiment Comparison**

In order to investigate the relative influence of both segmental and semantic information on tone learning, performance by listener groups from Experiments 1 and 2 were compared (Figure 6). Tone identification accuracy on the pre- and post-training tone ID tests was calculated for each group from Experiments 1 and 2 and submitted to logistic LMER. To examine the influence of implicit processing of semantic and segmental information on the explicit training of lexical tone, the LMER included Helmert-contrast coded fixed effects of Group (A: Unfamiliar-Meaning [UM] vs. Familiar-Tone Only [FTO] + Unfamiliar-Tone Only [UMN], B: FTO vs. UTO). It also included contrast-coded fixed effects for Test (pre, post) and Test Segment Type (familiar, unfamiliar). Random intercepts for Item and Participant were included, as well as a random slope for Test by item, to determine whether post-test accuracy increases stepwise from Unfamiliar-Meaning, Unfamiliar-Tone Only to Familiar-Tone Only. Additionally, "Participant" was included as a random factor in the statistical models in order to account for potential participant differences in performance. Therefore, while prior studies have found that individual perceptual and cognitive differences can have an impact on tone learning (Perrachione et al., 2011), the contribution of individual differences to the observed differences across groups is considered negligible in the current findings.

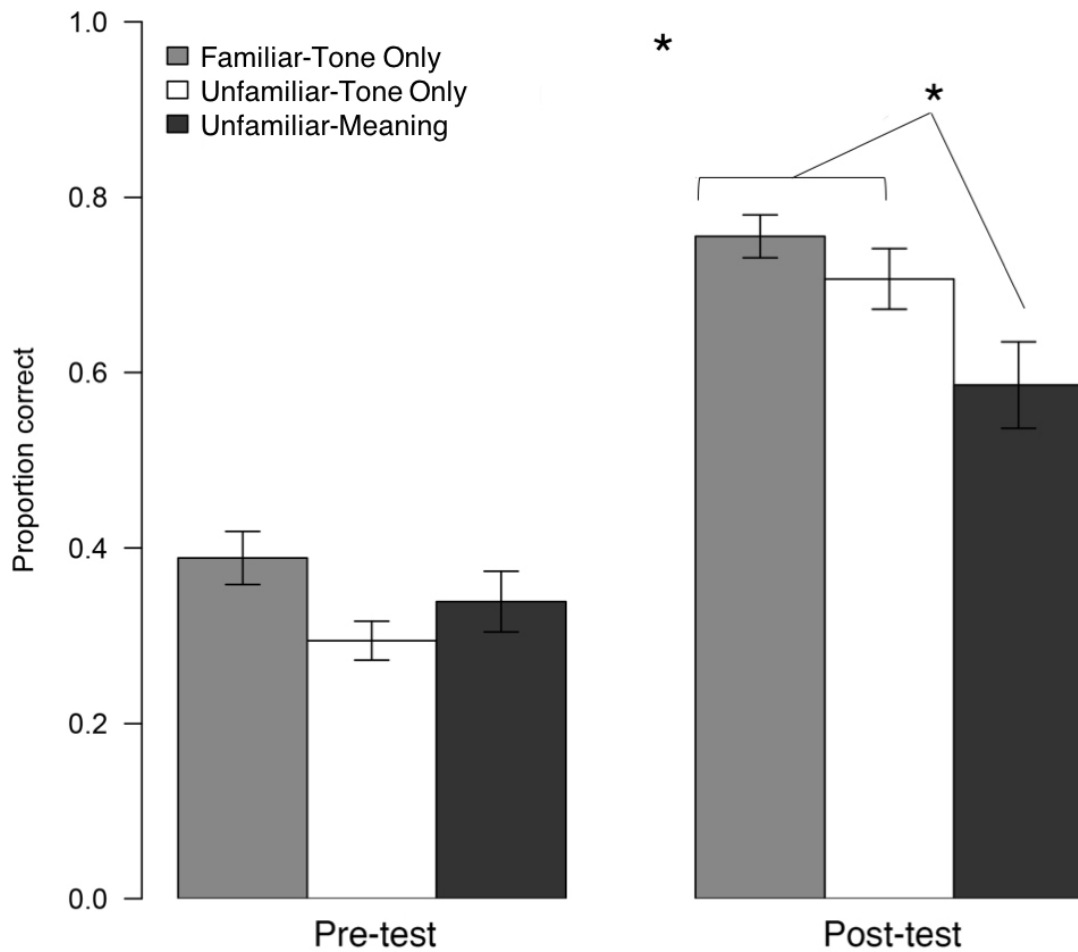


Figure 6. Mean proportion correct tone identification for Pre-test and Post-test by Group (Familiar-Tone Only, FTO; Unfamiliar-Tone Only, UTO; Unfamiliar-Meaning, UM). The top, centre asterisk denotes a significant pre- to post-test difference. Error bars indicate +/- 1 standard error.

A significant main effect of Test ( $\beta=1.52$ ,  $SE \beta=0.09$ ,  $\chi^2(1)=42.58$ ,  $p<0.001$ ) was obtained. A significant effect of Group A (UM vs. FTO + UTO) was also found ( $\beta=0.46$ ,  $SE \beta=0.21$ ,  $\chi^2(1)=4.7$ ,  $p=0.03$ ), along with a significant Test x Group A interaction ( $\beta=0.87$ ,  $SE \beta=0.14$ ,  $\chi^2(1)=37.729$ ,  $p<0.001$ ). Follow-up LMERS for each test with Group A as a fixed effect revealed no significant difference at pre-test ( $p=0.98$ ) but a significant difference at post-test ( $\beta=0.89$ ,  $SE \beta=0.31$ ,  $\chi^2(1)=7.7305$ ,  $p=0.005$ ), indicating that the Meaning group performed significantly worse than both Tone Only groups. A marginal effect of Group B (FTO vs. UTO)

was found ( $\beta=0.35$ ,  $SE \beta=0.18$ ,  $\chi^2(1)=3.806$ ,  $p=0.051$ )<sup>1</sup>; however, the Test x Group B interaction did not reach significance ( $p=0.18$ ). None of the effects or interactions involving Test and Segment Type were significant ( $p>0.14$ ), indicating that tone identification during pre- and post-tests was not influenced by whether the initial consonant of the test item was familiar or unfamiliar to listeners (e.g., listeners did not perform better identifying tones on the syllable [pou] vs. [ciou]).

Overall, these findings indicate that the inclusion of semantic information significantly inhibited the acquisition of L2 lexical tones, as the Unfamiliar-Meaning group performed significantly worse than both Familiar-Tone Only and Unfamiliar-Tone Only groups by the end of training. The familiarity of the segmental information provided during training did not appear to significantly improve post-training tone identification accuracy

#### **4. Discussion and conclusions**

The aim of the present study was to investigate the influence of linguistic processing load on the perceptual learning of L2 lexical tone contrasts. Compared with L1 speech sounds, speech perception of 12 sounds particularly by late L2 learners, requires listeners to expend more cognitive resources in order to extract the necessary phonetic information to differentiate the contrasts (Strange, 2011). Lexical tone provides a unique test case as the acquisition of a tone word involves three layers of information: tonal, segmental and semantic. In this study, we assessed the influence of processing load on the acquisition of lexical tone by examining the roles of semantic information and segmental familiarity.

Consistent with prior studies (e.g., Wang et al., 1999), the overall results revealed that tone identification training had a significant facilitative effect on native English listeners' ability to identify L2 lexical tones, with all groups significantly improving from pre- to post-test. Regarding the role of semantic information in tone learning, the current results show that listeners who received explicit semantic access (Meaning group) during training had significantly lower tone identification accuracy on the post-training test relative to those who focused on tone only (Tone Only group, Figure 2), even though their accuracy in identifying the tones during

---

<sup>1</sup> The significant FTO vs. UTO effect found in Exp. 2 is only marginal in this analysis, which may result from using a slightly different model in the cross-experiment comparison than in Exp. 1.

training was higher than the Tone Only group (Figure 4). These results are consistent with previous findings that the perception of difficult L2 segmental contrasts are worse after providing training and focuses learners' attention on semantic information than if they are told explicitly to focus on the speech sound differences (Guion & Pederson, 2007), and extends it to the perceptual learning of L2 suprasegmental contrasts. In the current study, even when not explicitly asked to pay attend to meaning or commit these meanings to memory, listeners may have automatically processed the information, diverting attention and resources away from extracting the relevant cues for distinguishing the lexical tone contrasts. This inhibition of perceptual learning may then have arisen from the increased processing load associated with processing both phonemic and higher-level semantic information (Strange, 2011). These results suggest that at least at the initial stages of learning, alleviating processing load improves the perception of L2 phonemic contrasts. Training with explicit focus on a single dimension, in this case tonal information, appears to be more beneficial than the inclusion of information from multiple linguistic dimensions, as it may alleviate the attentional and processing load associated with multi-domain linguistic information (Guion & Pederson, 2007; Werker & Fennell, 2004), especially for tone words that involve suprasegmental as well as segmental and lexical information.

Given that providing semantic information appeared to inhibit learning, why then was performance for the Meaning group significantly better over the course of training? One possible explanation for their superior performance on training tests is that trainees may have memorized the association between the whole entity of each training stimulus (i.e., cumulative segmental, tonal and semantic information) and the corresponding word object represented as a picture, rather than attending to the tonal patterns per se. This simple entity-picture match may have enabled them to better acquire the limited number of specific items they received during training. This interpretation finds support in previous research showing improvements in the word-meaning match task for those tone words used in training but no improvements in post-training tone identification involving new stimuli (Morett & Chang, 2015), indicating the effects of mnemonic labeling strategies rather than tone learning. Indeed, the improvements in entity-meaning association with trained words may not have facilitated the formation of generalizable tonal representations that would allow them to efficiently identify L2 tones on untrained items. Prior research on L2 speech learning posits that successful

learning is marked by the establishment of new L2 phonemic categories, and one way to test category formation is whether improvement from training can extend to new stimuli and talkers (e.g., Bradlow et al., 1997; Wang et al., 1999). The current results of better post-test performance by the Tone Only groups demonstrate evidence of more robust tone category formation relative to the Meaning group, at least in the short term. It should also be noted that while the inclusion of semantic information during training inhibited the formation of L2 lexical tone categories at the initial learning stage, it may nevertheless be advantageous for long-term learning, since different dimensions of linguistic information may affect learning at different stages (So & Best, 2010; Wu, Munro & Wang, 2014). It remains for future research to examine the long-term consequences of manipulating these different dimensions of information during training.

In addition to manipulating the semantic layer of information, the current study also examined the influence of the segmental dimension during lexical tone learning. One might expect that providing tones on syllables containing unfamiliar non-native segments would also serve to increase processing load, as perception might involve categorizing and integrating both L2 segmental and suprasegmental components. The results show that while listeners in the Familiar group did significantly outperform listeners in the Unfamiliar group across tests, the magnitude of improvement on tone identification from pre- to post-tests did not significantly differ as a result of training segment type (Figure 4). This lack of a robust facilitative effect of segmental familiarity on tone learning may have stemmed from unfamiliar L2 segments not being sufficiently taxing to process (at least not substantively more taxing than familiar segments), or a dimension more easily tuned out than visually-presented semantic information when focusing on identifying suprasegmental contrasts. This is consistent with prior work examining the influence of non-native (versus native) phonology on grammar processing in an artificial language (Finn, Kam, Ettliger, Vytlačil, & D'Esposito, 2013). Neural recruitment was found to differ as a function of whether the artificial language used native or non-native phones; however, no behavioural differences were ultimately observed.

As for why semantic but not unfamiliar segmental information increased processing load for listeners, one might ask if it was because conveying word meaning in the current experiment involved more complex information in the visual modality. While listeners in Experiment 2 saw a fixation cross during training, listeners in the Meaning group in

Experiment 1 received 24 different visual items on the screen over the course of training. However, the fact that the Meaning group significantly outperformed the other groups during training would suggest that *viewing* the pictures themselves did not enhance processing difficulty relative to the fixation cross. Rather, it could be the case that encoding semantic information into newly-forming lexical representations required more cognitive resources than processing unfamiliar segmental information (Figure 6).

Taken together, the current results are in line with the ASP model (Strange, 2011), which posited that L2 listeners' ability of perceive L2 phonemic contrasts is dependent not only on linguistic factors (e.g., the phonetic similarities between L1 and L2 phonemic categories) but also on cognitive factors, such as stimulus complexity, that may affect how much cognitive effort is being expended (that is, how heavy the processing load is during speech perception) and how much attention is paid to the relevant dimensions of the speech input (e.g., Antoniou & Wong, 2015; Chandrasekaran et al., 2016). According to this model, L2 learners at the beginning stages of acquiring a second language, such as those included in the present study, operate in a phonetic mode of processing, requiring greater attentional focus and cognitive resources. In line with this account, the present study suggests that having to process multiple dimensions of information concurrently may increase the amount of cognitive effort required of L2 listeners, relative to being able to focus on just one or two dimensions, and potentially divert attention away from the fine-grained phonetic information necessary to differentiate L2 phonemic contrasts. The different dimensions of a lexical tone word, which include segmental, tonal and semantic information, may exert differing degrees of processing load on L2 listeners. Specifically, the results indicate that providing semantic information during the acquisition of L2 tonal contrasts may have intensified the processing load, inhibiting tone learning, even though listeners were not explicitly asked to attend to the semantic information. On the other hand, being unfamiliar with the segmental information of the tone words, at least their initial consonant which is not integral to tone processing, did not intensify the processing load to a degree that interfered with tone learning.

While the eventual goal of any language learner is to acquire a lexicon of word forms to use in communicative exchanges, providing learners in their initial stages of learning with semantic information while they attempt

to acquire difficult phonemic contrasts appears to be disadvantageous. Thus, the current findings support our previous study (Cooper & Wang, 2013), suggesting that allowing learners to first form stable, delineated phonemic representations in the earliest phase of L2 learning through focused training on the relevant phonemic contrasts enables them to more easily acquire word meanings distinguished by those contrasts. It remains for future work to examine the long-term contributions of various levels of linguistic information, as L2 speech learning involves a dynamic process where different resources may be utilized at different stages of learning.

## **5. Acknowledgements**

Portions of this research were presented at the *17<sup>th</sup> International Congress of Phonetic Sciences* (2011) in Hong Kong. We thank Caitlin Annable, Mathieu Dovan, Alison Kumpula, Rebecca Simms, and Xianghua Wu from the Language and Brain Lab at Simon Fraser University for their assistance. This study was funded by research grants from the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grant-312457-2006, 2011).

## **References**

- Antoniou, M., & Wong, P. C. M. (2015). Poor phonetic perceivers are affected by cognitive load when resolving talker variability. *Journal of the Acoustical Society of America*, *138*(2), 571-574. <https://doi.org/10.1121/1.4923362>
- Antoniou, M., & Wong, P. C. M. (2016). Varying irrelevant phonetic features hinders learning of the feature being trained. *Journal of the Acoustical Society of America*, *139*(1), 271-278. <https://doi.org/10.1121/1.4939736>
- Beddor, P. S., & Strange, W. (1982). Cross-language study of perception of the oral-nasal distinction. *The Journal of the Acoustical Society of America*, *71*(6), 1551-61. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7108030>
- Best, C. T. (1995). The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The Development of Speech Perception* (pp. 167-224). MIT Press.
- Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339-2349. <https://doi.org/10.1121/1.2642103>



- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4), 2299-2310. <https://doi.org/10.1121/1.418276>
- Chandrasekaran, B., Yi, H.-G., Smayda, K. E., & Maddox, W. T. (2016). Effect of explicit dimension instruction on speech category learning. *Attention, Perception & Psychophysics*, 78(2), 566-582. <https://doi.org/10.3758/s13414-015-0999-x>.Effect
- Chao (1948). *Mandarin Primer*. Cambridge: Harvard University Press.
- Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *The Journal of the Acoustical Society of America*, 131(6), 4756-69. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22712948>
- Cooper, A., & Wang, Y. (2013). Effects of tone training on Cantonese tone-word learning. *The Journal of the Acoustical Society of America*, 134(2), EL133-EL139. <https://doi.org/10.1121/1.4812435>
- Davidson, L., Shaw, J., & Adams, T. (2007). The effect of word learning on the perception of non-native consonant sequences. *The Journal of the Acoustical Society of America*, 122(6), 3697-709. <https://doi.org/10.1121/1.2801548>
- Finn, A. S., Kam, C. L. H., Ettlinger, M., Vytlačil, J., & D'Esposito, M. (2013). Learning language with the wrong neural scaffolding: the cost of neural commitment to sounds. *Frontiers in Systems Neuroscience*, 7, 1-15. <https://doi.org/10.3389/fnsys.2013.00085>
- Flege, J. E. (1995). Speech Language Speech Learning: Theory, Findings and Problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233-277). Timonium, MD.
- Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. M. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, 36(2), 268-294. <https://doi.org/10.1016/j.wocn.2007.06.005>
- Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, 25, 207-231.
- Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning* (pp. 57-77). Amsterdam: John Benjamins Publishing Company.
- Hallé, P. A., & Best, C. T. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *The Journal of the Acoustical Society of America*, 121(5), 2899-2914. <https://doi.org/10.1121/1.2534656>
- Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32(3), 395-421. [https://doi.org/10.1016/S0095-4470\(03\)00016-0](https://doi.org/10.1016/S0095-4470(03)00016-0)

- Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, 1, 65-94.
- Hisagi, M., & Strange, W. (2011). Perception of Japanese Temporally-cued Contrasts by American English Listeners. *Language and Speech*, 54(2), 241-264. <https://doi.org/10.1177/0023830910397499>
- Ingvalson, E. M., Barr, A. M., & Wong, P. C. M. (2013). Poorer Phonetic Perceivers Show Greater Benefit in Phonetic-Phonological Speech Learning. *Journal of Speech, Language, and Hearing Research*, 56, 1045-1050. [https://doi.org/10.1044/1092-4388\(2012/12-0024\)Materials](https://doi.org/10.1044/1092-4388(2012/12-0024)Materials)
- Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, 118(5), 3267-3278. <https://doi.org/10.1121/1.2062307>
- Lin, M., & Francis, A. L. (2014). Effects of language experience and expectations on attention to consonants and tones in English and Mandarin Chinese. *The Journal of the Acoustical Society of America*, 136(5), 2827. <https://doi.org/10.1121/1.4898047>
- Morrett, L. M., & Chang, L. Y. (2015). Emphasizing sound and meaning in tonal language acquisition: A gesture training study. *Language, Cognition and Neuroscience*, 30, 347-353.
- Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *The Journal of the Acoustical Society of America*, 127(2), EL54-EL59. <https://doi.org/10.1121/1.3292286>
- Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, 130(1), 461-472. <https://doi.org/10.1121/1.3593366>
- Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America*, 89(6), 2961-2977.
- Sereno, J. A., & Lee, H. (2015). The Contribution of Segmental and Tonal Information in Mandarin Spoken Word Processing. *Language and Speech*, 58(2), 131-151.
- So, C. K., & Best, C. T. (2010). Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech*, 53(2), 273-293.
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388, 381-382.
- Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39(4), 456-466. <https://doi.org/10.1016/j.wocn.2010.09.001>

- Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*, 23(5), 689-708. <https://doi.org/10.1080/01690960701728261>
- Wang, Y., Spence, M. J., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, 106(6), 3649-58. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10615703>
- Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese Listeners to Perceive Thai Tones : A Preliminary Report. *Language Learning*, (December), 681-712.
- Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, 36(2), 250-267. <https://doi.org/10.1016/j.wocn.2007.06.004>
- Werker, J. F., & Fennell, C. (2004). Listening to Sounds versus Listening to Words: Early Steps in Word Learning. In D. G. Hall & S. R. Wazman (Eds.), *Weaving a lexicon* (pp. 79-109). Cambridge, MA, US: MIT Press.
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10(4), 420-422. <https://doi.org/10.1038/nn1872>
- Wu, X., Munro, M.J., and Wang, Y. (2014). Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *Journal of Phonetics*, 46, 86-100.
- Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, 113(2), 234-43. <https://doi.org/10.1016/j.cognition.2009.08.010>
- Yip, M. (2002). *Tone*. New York: Cambridge University Press.

## Focus on Consonants: Prosodic Prominence and the Fortis-Lenis Contrast in English

Míša Hejná & Anna Jespersen  
Aarhus University

### Abstract

This study investigates the effects of intonational focus on the implementation of the fortis-lenis contrast. We analyse data from 5 speakers of different English dialects (Ocke’s colleagues), with the aim of examining the extent to which different correlates of the contrast are used by each speaker, and whether the contrast is implemented differently across different levels of focal prominence (narrow focus, broad focus, de-accentuation). The correlates examined include three measures often associated with the contrast (pre-obstruent vowel duration, consonant/vowel durational ratio, rate of application of obstruent voicing), as well as a number of lesser-investigated phenomena. Firstly, we find that individual speakers utilise different phonetic correlates to implement the fortis-lenis contrast. Secondly, focus affects several of these, with the biggest effect found with consonant/vowel ratio, and the smallest with obstruent voicing.

### 1. Introduction

It has been frequently claimed that there is more variation in vowels than consonants (e.g. Bohn & Caudery, 2017, p. 63), possibly because “consonantal variation (in British English at least) tends to be used less as a way of marking local identity than vocalic variation does” (Trousdale, 2010, p. 116). An alternative claim may be that “[c]onsonantal features have been studied far less rigorously than vowel features” (Cox & Palethorpe, 2007, p. 342, who comment on the state of consonantal variation studies in Australian English; but see also Su, 2007, p. 6).

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 237-270). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.

These claims are nevertheless somewhat surprising at least in the context of the fortis-lenis contrast, which has traditionally received widespread attention from phoneticians and which has been reported to show a wide range of phonetic implementation strategies (e.g. Jansen, 2004; Kingston, Diehl, Kirk, & Castleman, 2008; Kohler, 1984; Toscano & McMurray, 2010). Limiting ourselves to English, this contrast is generally reported to be cued and signalled by differences in the voicing of the obstruent, the duration of the preceding vowel, and by the presence and duration of post-aspiration (Bohn & Caudery, 2017, pp. 74-81). Regarding plosives, previous work has tended to focus on post-aspiration in the form of VOT. VOT analyses have targeted primarily the word-initial position (as in *tock* vs *dock*; see e.g. discussions in Docherty, 1992; Iverson & Salmons, 2006). For obstruents more generally, analyses have centred on the preceding vowel duration non-initially (e.g. Fox & Palethorpe, 2007, p. 343; Wells, 1990). In comparison to VOT and preceding vowel duration, discussions of *voicing* tend to be limited (but see e.g. Scobbie, 2005; Smith, 1997). More recently, glottalisation and pre-aspiration have been reported as correlates of the contrast in some British English accents (Hejná, 2016b; Hejná & Kimper, In press; Gordeeva & Scobbie, 2013), and glottalisation also in Australian English (Penney, Cox, Miles & Palethorpe 2018). Furthermore, a range of other potential correlates of the fortis-lenis contrast in British English has been put forward (Hejná & Kimper, Submitted). Thus the implementation strategies of the fortis-lenis contrast are abundant in English varieties.

We propose that one way to tap into the strength of different correlates of the fortis-lenis distinction is through investigating its variation under different levels of prosodic prominence. Previous work on the fortis-lenis contrast has tended to make use of laboratory speech, in which the target word for the investigation is often prosodically focused (see e.g. the following studies, which use word lists and/or carrier sentences likely to encourage production with narrow focus and which represent but a few examples available: Gordeeva & Scobbie, 2013; Hejná, 2016b; Hejná & Kimper, In press; Scobbie, 2005; Smith, 1997). Apart from the fact that these conditions may not reflect the phonetic situation in more natural data regarding VOT, voicing, and the preceding vowel duration, it is also possible that the robustness of lesser-investigated potential correlates in such studies has a causative relationship with prosodic prominence (as discussed e.g. in Mücke & Grice, 2014). Controlling for this variable may thus yield insights into the degree to which individual correlates

of the fortis-lenis contrast are utilised by speakers of English. At the same time, such an investigation can enrich our knowledge of prosodic prominence. More specifically, non-focal prominence, such as word stress, is known to have different acoustic correlates across languages (Mennen, 2006; Leemann, Kolly, Li, Chan, Kwen, & Jespersen, 2015). While the contributions of syllable duration is known to play a role in the perception of prominence (Leemann et al., 2015; Hua, Li, & Wayland, this volume), little is known of the contribution of segmental variation to prosodic prominence in general, and to intonational focus in particular. As such, focusing on focused consonants therefore also extends our knowledge of the ways in which prosodic prominence depends on segmental variation.

### **1.1. Interactions between prosody and segments**

Studies and models of intonational variation, such as variation in intonational focus, tend to treat its physical manifestation as more or less synonymous with variation in  $f_0$  (Fundamental Frequency; see Niebuhr, 2013, for a discussion of this). However, intonation is crucially dependable on variation not just in  $f_0$  but also in the duration, intensity and phonatory characteristics of segments which form the underlay of suprasegmental variation (Nolan, 2006, p. 433). It has long been known that the intrinsic prosodic characteristics of individual segments can influence the course of  $f_0$  movements (Kingston, 1986; Kingston, Diehl, Kirk, & Castleman, 2008). Nonetheless, recent work has started to broaden our knowledge of the ways in which both the production and the perception of prosodic phenomena such as intonation rely on segmental variation. For instance, evidence from German suggests that local intonational categories are cued through variation in the levels of intensity and duration of accented syllables, and that listeners are able to make use of this variation to identify distinct pitch contours and decode their communicative content (Niebuhr & Pfitzinger, 2010; see also Niebuhr, this volume).

There is, however, limited evidence for the impact of intonational variation on segmental realisation. Niebuhr and colleagues have conducted a number of studies on German in which voiceless fricatives such as [f], [s], [ʃ], and [x] have been shown to have greater intensity and higher centres of gravity after rising  $f_0$  than after  $f_0$  falls, both medially and finally (Niebuhr, 2012; 2013, p. 10). Coarticulatory processes can also be affected: the assimilation of /s/ to /ʃ/ has been argued to be stronger with  $f_0$  rises than with falls (Niebuhr et al., 2011). Studies have also suggested

that listeners are able to tap into segmental variation as additional cues to intonational contrasts and to compensate for truncated  $f_0$  movements (Kohler, 2011). Limited work has nevertheless been carried out on non-fricative obstruents. One exception to this is Niebuhr (2008), who shows that the aspiration of utterance-final /t/ can be impacted by intonational contour type. Two falling contours were included, which were distinct in the alignment of the peak: early (H+L\* L%) or late (L\*+H L%). With early  $f_0$  peaks, /t/ aspirations were realised as short, intense and with low-frequency centres of gravity, while late-peak aspirations were longer, less intense and featured energy at higher frequencies. A subsequent perception study indicated that listeners made use of /t/ aspiration as a cue to the intended  $f_0$  contour.

This paper aims to extend previous findings on the effects of intonational variation on the realisation of consonants by homing in on the contribution of *prosodic focus* on obstruent realisation. The manifestation of focus can have three different outcomes on individual syllables. *Broad focus* assigns the whole utterance prominence.<sup>1</sup> In other words, no single element of the utterance receives special prominence; a nuclear intonation contour is thus assigned to the last metrically prominent syllable in the utterance. This level of focus is perceived by listeners as prosodically neutral (Cruttenden, 1996, p. 87). *Narrow focus* assigns the nucleus to any syllable in the utterance. A syllable receiving narrow focus is highlighted as carrying new information and thus receives a greater amount of prosodic prominence than the nuclear syllable under broad focus (though in phonetic terms, this distinction might not always be unambiguous; cf. Ladd, 1996, pp. 254-56). Assigning narrow focus to a syllable affects the prosodic makeup of the entire utterance, in that it de-accent all following pitch accents (Ladd, 1996, p. 225). *De-accenting* affects any material following a narrowly focussed syllable by reducing e.g. the pitch range of  $f_0$  movements and the duration of metrically prominent syllables, and may also affect the strength of boundary cues (Baltazani & Jun, 1999; Norcliffe & Jaeger, 2005). Realisational effects of de-accenting a syllable include reduced vowel durations, changes in vowel quality, and the non-elision of linking- and intrusive /r/ (Ladd, 1996, pp. 266-67). Yet studies have not focused on the fine-grained

<sup>1</sup> Note that “prominence” is used in this study to mean *metrical* prominence; whether and to which extent accented syllables always receive *physical* prominence, as expressed through e.g.  $f_0$ , durational and segment-realisation means, is not fully established. For a discussion definitions of prominence, see Wagner et al. (2015).

phonetic consequences of the deaccentuation of syllables on consonants, or compared consonantal realisations across different focus conditions (for an overview of papers discussing the effects of focal prominence on vowel realisation, see Mücke & Grice, 2014, p. 5).

## **1.2. Hypotheses and research questions**

This paper aims to answer three research questions:

- Is the implementation of the fortis-lenis obstruent contrast affected by focus?
- In which other ways does focus affect variation in obstruent realisation?
- Does the nuclear intonation contour affect variation in obstruents?

With regard to the first research question, we hypothesise that focus will affect the implementation of the fortis-lenis contrast in such a way that increased levels of prosodic prominence will correlate with higher strength of the consonantal correlates,<sup>2</sup> with a decrease in this strength as the prosodic prominence decreases (Kügler, 2008; Görs & Niebuhr, 2012). This hypothesis is based on the well-known durational differences between different levels of focal prominence (e.g. between broad and narrow focus, see Baumann, Becker, Grice, & Mücke, 2007), and on studies which find a correlation between increased focal prominence and exaggerated articulation, in which a narrowed focus domain is signalled through increased articulatory effort, e.g. through increased lip-rounding in the production of French vowels (Dohen, Loevenbruck, & Hill, 2006; de Jong, 1995) and increased peripherality in German vowels (Baumann et al., 2007).

As to the second research question, it is well-known that there is more variation in obstruent variation in non-foot-initial position (e.g. Smith, 2002), but little is known about other prosodic contexts. Based on such findings, we expect that increased levels of prosodic prominence will correlate with increasing differentiation between the fortis and lenis obstruent realisation. A related hypothesis has previously been suggested for pre-aspiration by Hejrná (2015, pp. 241-2), who proposes that pre-aspiration first innovates in lexically stressed rather than unstressed

---

<sup>2</sup> We follow MacWhinney's definition of correlate strength, which includes correlate/cue reliability and correlate/cue availability (2001, 2012).



syllables. The motivations for this may be related to the durational properties of the preceding vowel, whose quality may be less likely to be affected by pre-aspiration-induced breathiness when they are durationally long (Steriade, 1998, p. 214). If that is the case, pre-aspiration should be likely to interact with other types of prominence, such as focus. To the best of our knowledge, our study is the first to investigate this possibility.

Finally, we will briefly investigate the effects of different nuclear contours on obstruent realisation. This investigation was motivated by findings reported for German by Niebuhr et al. as discussed above. Hypotheses based on this work include a tendency for increased differentiation of the fortis-lenis contrast with high or rising intonational movements, as well as correlation between the complexity and duration of intonational contours. This has often been referred to in the literature (e.g. Ohala, 1978), but not investigated experimentally.

The structure of the study is as follows. We present our methodology for the study, where we also introduce the consonantal variables under consideration. Next, we show our results, examining both individual-specific patterns as well as those shared by our five speakers, and we discuss their implications for the study of the fortis-lenis contrast in English.

## **2. Methodology**

### **2.1. Data**

The data were recorded using a H4 Zoom Handy recorder with a C520 AKG headset microphone at a 44.1 kHz rate with a resolution of 32 bits. We collected a list of words embedded in carrier sentences with different prominence conditions (see below on the specifics of these carrier sentences). These words targeted foot-medial and foot-final fortis and lenis obstruents, namely the alveolar plosive and fricative fortis-lenis pairs /t-/d/ and /s-/z/, as shown in Table 1. Thus, each participant produced 24 word types, each repeated six times (ideally three times under narrow focus and three times in a de-accented prosodic environment). Efforts were made to include words with comparable lexical frequencies; however, in this study we were primarily concerned about excluding obviously low frequency items. We obtained 756 tokens for analyses in total.

/s/	/z/	/t/	/d/
<i>bus</i>	<i>buzz</i>	<i>mutt</i>	<i>mud</i>
<i>moss</i>	<i>Oz</i>	<i>lot</i>	<i>odd</i>
<i>lass</i>	<i>jazz</i>	<i>mat</i>	<i>mad</i>
<i>buses</i>	<i>buzzer</i>	<i>mutter</i>	<i>muddy</i>
<i>mossy</i>	<i>Ozzy</i>	<i>otter</i>	<i>odder</i>
<i>lassie</i>	<i>jazzy</i>	<i>matter</i>	<i>madder</i>

Table 1. Words with fortis-lenis alveolar obstruents.

The participants also read a number of distractors, which included *abbey*, *bleak*, *blob*, *blobby*, *goofy*, *Hobbit*, *hobby*, *hoof*, *lab*, *leak*, and *leek*. The order of the items was randomised, which also extends to the different carrier sentences. However, the order in which the words were presented to the participants was uniform across these participants.

We investigated the effects of focus on segmental realisation by incorporating target words into carrier sentences such as *Ocke is writing a paper on the word \_?*. In contrast to other studies of focus (e.g. Baltazani & Jun, 1999; D'Imperio, 2001; Xu & Xu, 2005), this experiment thus contained target words appearing at the end of the carrier sentence. That is, the location of the target word in the sentence was not changed across elicitations, and was not generally sentence-medial. We chose this context for two reasons: firstly, it allowed for an investigation of the effect of different nuclear intonation contours on the segmental material. Secondly, the phrase-final context made for a more conservative study. This is because English is known to have significant lengthening of both vowels and non-plosive consonants in phrase-final words (Klatt, 1975; Turk & Shattuck-Hufnagel, 2007). Differences in duration resulting from alterations of focal prominence will therefore be smaller in phrase-final than non-phrase-final positions, all other things being equal. As such, any durational-related differences in consonantal realisation resulting from different focus conditions in this position can therefore be considered robust.

We originally aimed to elicit two extremes in terms of focus: narrow focus, where the word under investigation is made metrically prominent, and de-accentuation, where the target word is made metrically non-prominent. The two focus levels were elicited by 1. capitalising

the target word (*Ocke is writing a paper on the word MOSS?*, where *moss* is the target), or 2. capitalising a non-target word in the carrier sentence (*OCKE is writing a paper on the word moss?*, where *moss* is the target). In order to avoid target words being judged by speakers as new information due to their continued change in static carrier phrases, and thus triggering narrow focus, the two content words in the carrier sentence (in the above sentences, *Ocke* and *paper*) were randomly varied with each new sentence, so that the subject of the sentence was either a pronoun or the name of one of Ocke Bohn's colleagues at the Department of English, and the object a type of research output (*paper*, *keynote*) which could plausibly be generated by these persons<sup>3</sup>. In order to further prompt the speakers to produce the intended level of focus, they were presented with powerpoint slides which contained cues to the appropriate pragmatic context (e.g. “not *Anna / Míša?*” or “not the word *wiggle / bumblebee / pumpkin?*”). Finally, the speakers were given a training set of example sentences prior to the start of the recordings in which the relevant pragmatic contexts were explained to them by the first author.

Despite the written cues and oral demonstrations, when presented with the carrier phrases intended to elicit narrow focus, several speakers varied between producing the sentences with narrow focus on the target word and with broad (i.e. “neutral”) focus. In this case, the target word, because of its placement at the end of the target sentence in this experiment, corresponded to the nuclear accent. Broad focus on the sentence in this case meant that the target word became more prominent than in the deaccented condition, but did not receive extra levels of metrical prominence due to narrow focus. In this way, we inadvertently ended up eliciting three levels of focus, whereby the target words were either de-accented, relatively neutral (but accented) or under narrow intonational focus.

## 2.2. Quantifying consonantal variation

In order to answer the research questions, we measured the duration of the pre-obstruent vowel, the duration of the obstruent, and the frequency of application of voicing. In addition, the following phenomena were annotated for their presence: post-aspiration, affrication, spirantisation, ejectives, glottal replacement, pre-aspiration, and flapping. All annotation was carried out manually in Praat (Boersma & Weenink,

---

<sup>3</sup> See Niebuhr & Michaud (2015) for further discussion of read speech and intonational elicitation.

1992-2017). Since none of the speakers displayed ejectives and glottal replacement, these phenomena are not commented on in what follows. Furthermore, glottal reinforcement presented challenges which were beyond the scope of this paper, and was therefore not quantified.

### Segmental annotation

On the whole, the onset and offset of the vowel were defined as the onset and offset of periodicity, which were determined on the basis of the soundwave. However, further context-specific criteria were used. Firstly, word-initial affricates (*jazz, jazzy*) are generally voiceless; however, they could be articulated as voiced at their offset, which means that a period of voiced oral friction could be found in our affricate-vowel sequences. In such cases, it was the offset of the oral friction that defined the onset of the vowel. Secondly, when a vowel was followed by a phonetically voiced fricative (e.g. *Oz, Ozzie*), it was again the onset of the oral friction of that fricative that defined the offset of the vowel. Next, in case of vowel-initial words, aperiodic glottalisation was excluded from the vocalic interval (*otter, odder, odd, Oz, Ozzie*), unless this glottalisation actually affected the whole vowel.<sup>4</sup> On the other hand, where glottalisation was found which was not due to vowel-initial word effects, this was always considered part of the vowel for practical rather than theoretical reasons: the boundary has to be placed somewhere in order to extract the measures and it is not always obvious whether the glottalisation is conditioned by the following obstruent.<sup>5</sup>

### Voicing

We identified the presence and the duration of voicing on the basis of the waveform. The onset of periodicity was considered the onset of voicing and its offset was considered the offset of voicing. As long as there was some voicing present in the obstruent (fortis or lenis), the token was

---

<sup>4</sup> This was done because it became obvious that glottalisation could reach fairly high durational values without obviously affecting those of the vowel. The entire vowel was sometimes glottalised, in which case glottalisation could not be excluded as this would have left us with no vowel to measure. Yet these cases were *very* infrequent.

<sup>5</sup> Although pre-glottalisation is a frequent feature of fortis plosives in a number of English accents and has, moreover, been found to function as a correlate of the fortis-lenis contrast in at least three accents of British English (see Hejná & Kimper, In press), this phenomenon was excluded from our analyses. This decision was made because it is problematic to distinguish utterance-final glottalisation, individual-specific global glottalisation, and subsegmental glottalisation (i.e. glottal reinforcement).

considered to contain voicing. Although we do not report on variation as to the extent to which the obstruent is voiced (e.g. 50% of the obstruent), the vast majority of the phonetically voiced obstruents are only partially voiced rather than fully voiced. As can be expected, there is variation in the exact proportion value across and within our speakers. We also find a considerable amount of variation as to where in the consonantal interval voicing occurs (only initially, only finally, both initially and finally, throughout the duration of the obstruent). However, due to the limitations of space, variation beyond presence/absence of voicing is not commented on further in this study.

### Pre-aspiration

Pre-aspiration is defined as a period of voiceless (primarily) glottal friction which occurs in sequences of a vowel and a phonetically voiceless obstruent, in line with Hejná (2015; 2016a, amongst others), as shown in Figure 1. It is therefore distinguished from local breathiness, which is very closely linked to voiceless pre-aspiration (see Hejná, 2016a, for more details).<sup>6</sup> We followed the same segmentational criteria as Hejná (2016b), who looked into the role of pre-aspiration in the fortis-lenis plosive contrast in Aberystwyth English. Regarding the segmentation criteria, see Hejná (2015, pp. 85-87). Post-aspiration is defined and identified in the same way as pre-aspiration, with the difference of post-aspiration occurring during the release phase of the obstruent rather than prior to it.

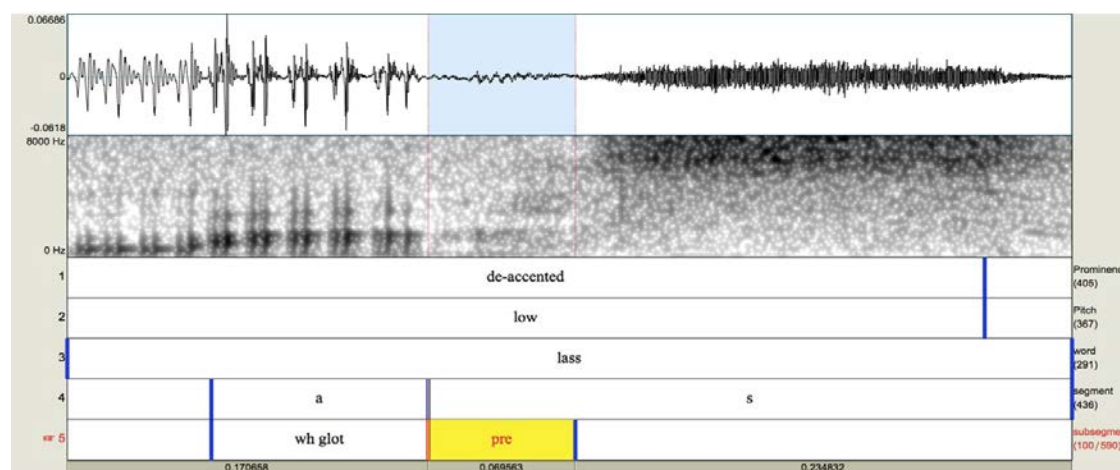


Figure 1. Segmentation of pre-aspiration in the fricative context (illustrated with Elly).

<sup>6</sup> Due to various aspects of breathiness in our data which make the phenomenon fairly complex this paper excludes discussions of breathiness.

### Flapping/tapping

Following Bohn and Caudery (2017, p. 79), and unlike Ladefoged (1968), we do not distinguish flapping and tapping. The term flapping is used in the rest of the paper. The phenomenon was identified on the basis of absence of the plosive burst and the presence of voicing, as illustrated in Figure 2 below.

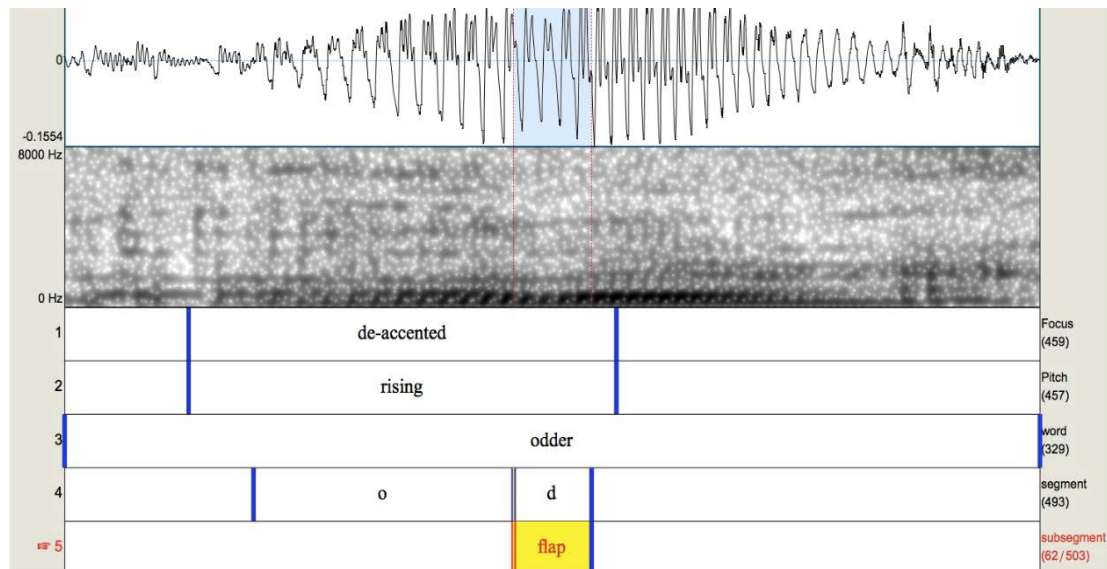


Figure 2. Identification of flapping (illustrated with Mark).

The vast majority of the potential cases of flaps at hand clearly met these criteria and were therefore unambiguously labelled as flaps. Similarly, the vast majority of the phonological plosives met the criteria used to identify voiced plosives. Nevertheless, it is noteworthy (but not surprising) that there is variation in the overall duration of voiced plosives in the data, resulting in what seems to be a durational continuum between a flap and a voiced plosive, where it can be solely the presence of a burst that distinguishes a flap from a plosive. This is found not only across the individual speakers but also within the individuals. For the purposes of this study, as long as a burst was clearly identifiable, the obstruent was classified as a plosive rather than a flap.

### Affrication

Affrication was identified by the presence of higher intensity energy in higher frequencies, which indicates oral rather than glottal friction (Figure 3). Affricated plosives were occasionally post-aspirated as well.

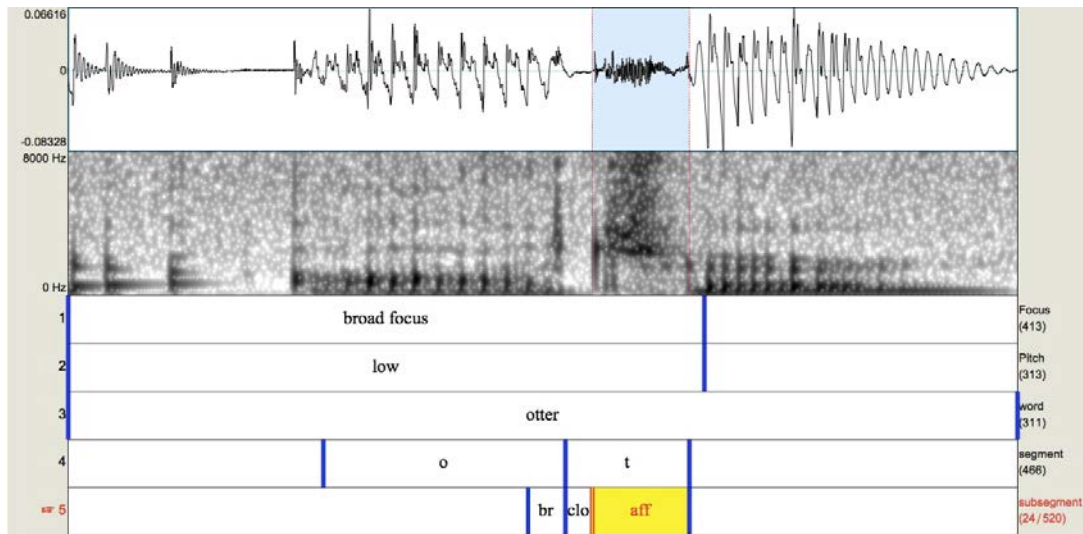


Figure 3. Identification of affrication (illustrated with Stephen).

### Spirantisation

Spirantisation of plosives was identified on the basis of the absence of the burst of the plosive. This could result either in a semi-spirantised plosive (semi-fricative; see Stevens & Hajek, 2005) or in a fully spirantised plosive, i.e. in a fricative (Figure 4 for full spirantisation). Interestingly, we found cases of fully spirantised plosives which were also pre- and/or post-aspirated.

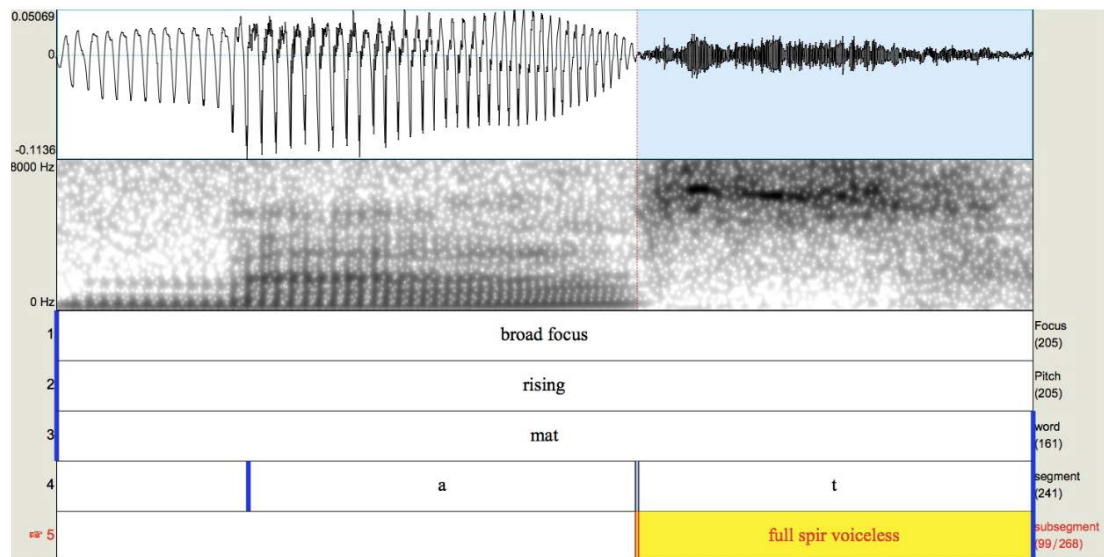


Figure 4. Identification of full spirantisation (illustrated with Antoinette).

### 2.3. Speakers

5 native speakers of English were recorded reading the words in the appropriate carrier sentences. These speakers were Ocke Bohn's colleagues, all of whom worked at the Department of English, Aarhus University, at the time of the recording (spring 2018). The five bravehearts include Antoinette Fage-Butler, Elly McCausland, Mark Eaton, Stephen Joyce, and Sophia Kier-Byfield. Since the speakers were born and raised in different regions, and have had variable life histories due to their profession as academics, their dialectological background is not uniform. Antoinette and Stephen represent Irish English (1 female speaker, 1 male speaker); Elly represents Standard Southern British English (1 female speaker); Mark represents Canadian English (1 male speaker); and Sophia represents Welsh English (1 female speaker), although Sophia's background and dialect history are more complex than that of the other participants.<sup>7</sup> Apart from regional and sex differences, the five departmental members also differ in other dimensions, most notably that of age. Whilst we acknowledge that this most likely accounts for much of the variation found in the dataset, these individual differences do not present a problem for our research questions.

### 2.4. Data processing

Statistical analyses were done through linear and logistic mixed effects and random forest modelling in R (R Core Team, 2018) and RStudio (RStudio Team, 2015), using the packages *lmer* (Kuznetsova et al., 2017), *lme4* (Bates et al., 2015), *effects* (Fox, 2003), *ranger* (Wright & Ziegler, 2017), and *lattice* (Sakar, 2008). Regression models were compared hierarchically using the *aov()* function. Post-hoc Tukey tests were performed through the *TukeyHSD()* function.

## 3. Results

In this section we give broad overviews of the main correlates of the fortislenis contrast as found across the speakers, but, given the highly variable speaker sample, also closely inspect potential individual patterns. We then examine the ways in which prosodic variation affects these realisations. We consider other correlates of the contrast, and then investigate the effects of focus and contour type on these lesser-investigated correlates as produced by five members of the AU Department of English.

---

<sup>7</sup> She spent most of her childhood in Wales and her parents are not speakers of a British English variety.



### 3.1. Preceding vowel duration, vowel-consonant ratio, and voicing of the obstruent

Firstly, the preceding vowel duration distinguishes the fortis-lenis obstruent contrast in all five department members (Figure 5): the duration of the pre-obstruent vowel is longer in the lenis than in the fortis obstruents, although this difference is not significant ( $F(3, 23)=1.21$ ,  $SS=2993.9$ ,  $p=.33$ ).<sup>8</sup> /t/ is associated with the shortest preceding vowel duration in our sample ( $\bar{x}=129\text{ms}$ ,  $SD=40.9\text{ms}$ ), as compared to /d/ ( $\bar{x}=155.7\text{ms}$ ,  $SD=60\text{ms}$ ) and /s/ ( $\bar{x}=139.5\text{ms}$ ,  $SD=42.9\text{ms}$ ).

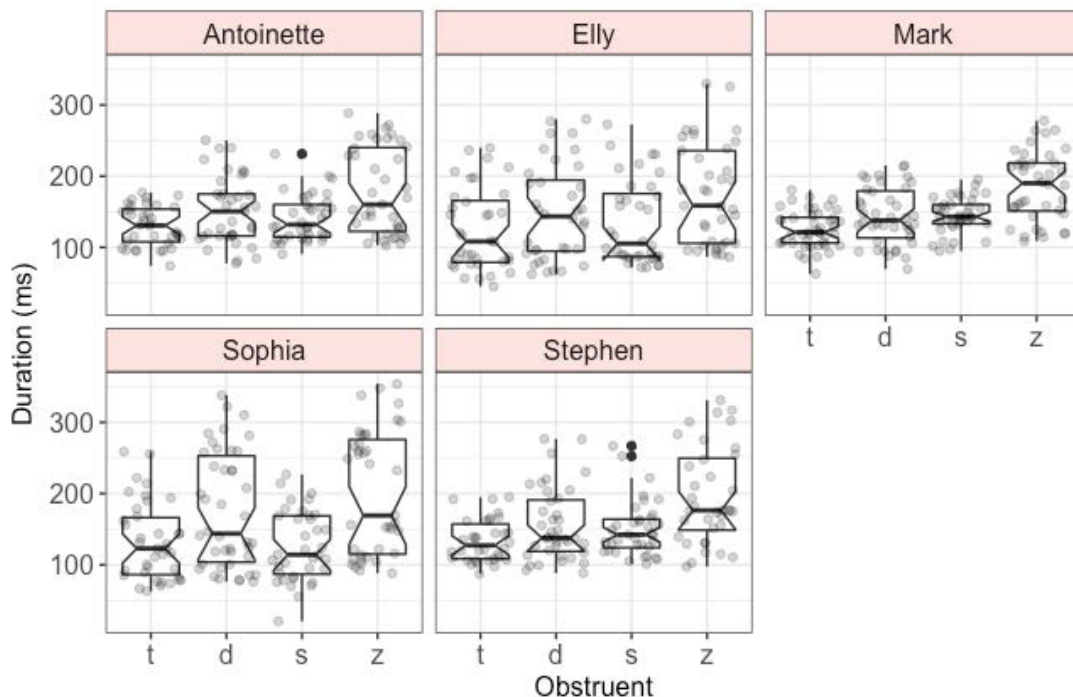


Figure 5. Preceding vowel duration (ms) and the fortis-lenis contrast by individual.

Tukey post-hoc tests revealed that the only statistically significant difference between individual obstruents is between /t/ and /z/ ( $p<.01$ ). Furthermore, we also see individual variation in the magnitude of these differences.<sup>9</sup> As shown in Figure 6, when we consider the V/C ratio and its role in the fortis-lenis contrast implementation, the duration of the preceding vowel is always longer than that of the obstruent in the lenis

<sup>8</sup> Final model:  $\text{lmer}(\text{Vdur} \sim \text{consonant} + (1|\text{speaker}) + (1|\text{word}), \text{data}=\text{data})$ .

<sup>9</sup> This individual variation is not due to potential differences in speaking rate: the same results are obtained with normalised vowel durations (as a percentage of the overall word duration).

series: /t/ ( $\bar{x}$ =1.31ms,  $SD$ =0.94ms) versus /d/ ( $\bar{x}$ =2.43ms,  $SD$ =1.1ms), and /s/ ( $\bar{x}$ =0.84ms,  $SD$ =0.24ms) versus /z/ ( $\bar{x}$ =1.6ms,  $SD$ =0.47ms). Consonant as a factor is thus a highly significant predictor of V/C ratio ( $F(3, 23)=20.457$ ,  $SS=24.54$ ,  $p<0.0001$ )<sup>10</sup>, and Tukey tests show that the fortis/lenis members in each pair are significantly different from each other (/t/-/d/:  $p<.0001$ ; /s/-/z/:  $p<.0001$ ). With respect to individual differences, Mark shows an individual trend with a considerable overlap in the values for the /t/ and /d/ contrast. This is most likely due to Mark's position as a /t/ and /d/ flapper, which makes him partially neutralise the /t/-/d/ contrast, as described in the literature on North American English (Braver, 2011; Derrick & Gick, 2011; Zue & Laferriere, 1979 for American English). Finally, as shown in Figure 7, all five department members use voicing as a correlate of the fortis-lenis contrast.

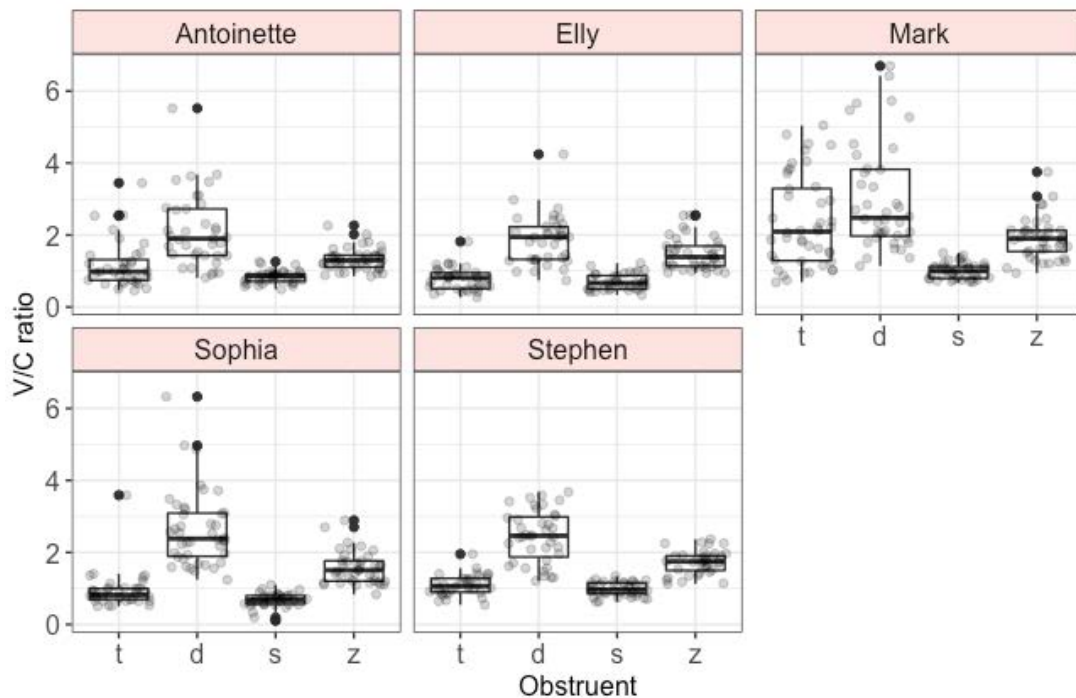


Figure 6. V/C ratio and the fortis-lenis contrast by individual.

Overall, the statistical analyses confirm that consonant type is a significant predictor of the presence of voicing ( $F(3, 23)=13.412$ ,  $SS=17304$ ,  $p<.0001$ ).<sup>11</sup> Post-hoc Tukey tests show that the fortis-lenis pairs are strongly significant against each other (/t/-/d/:  $p<.0001$ ; /s/-/z/:

<sup>10</sup> Final model: `lmer(V/C proportion ~ consonant + (1|speaker) + (1|word), data=data)`.

<sup>11</sup> Final model: `glmer(voicing ~ consonant + (1|speaker) + (1|word), family="binomial", data=data)`.

$p < .0001$ ) while /t/-/s/ are weakly significant against each other ( $p = .09$ ), and /d/-/z/ are not ( $p = .13$ ). However, there is clear individual variation in the specific amounts of voicing produced to maintain this distinction:

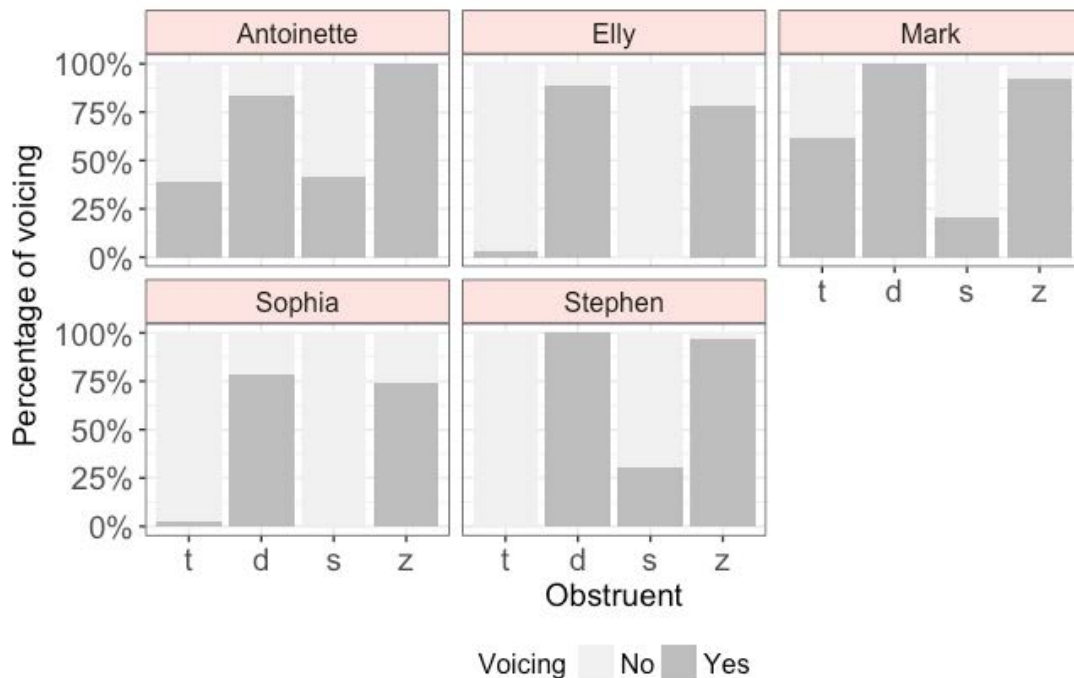


Figure 7. Presence of voicing (%) for each obstruent by individual.

Sophia and Elly are very consistent in that they never voice their /s/, whilst they produce voicing in 74% and 78% of their /z/ productions, respectively. Stephen is very consistent in that he never voices his /t/ and always voices his /d/. On the other hand, Stephen's fricatives show a less straightforward situation: whilst he voices both /s/ and /z/, he voices his /z/ more or less obligatorily (97%), whereas he voices his /s/ in 31%. Furthermore, the duration of the voicing in Stephen's /z/ reaches 47% of the overall consonantal duration on average, in contrast to that of 8% in his phonetically voiced /s/'s. Although Antoinette shows voicing in all of her obstruents, it is less frequent in her /t/ (39%) than /d/ (84%), and in her /s/ (42%) than /z/ (100%). Additionally, the duration of Antoinette's voicing is always shorter in the fortis obstruents (/t/: 18%; /d/: 39%; /s/: 7%; /z/: 21%). Mark presents a more complicated picture. His /d/ is categorically voiced and his /z/ is practically always voiced as well (93%); he voices his /s/ in 21% of the times and his /t/ in 61% of the times, which still renders him a speaker who utilises voicing to distinguish the fortis-lenis contrast in his production, albeit somewhat differently from the other department members.

### 3.2. Is the implementation of the contrast affected by prosodic variation?

This section highlights the influence of focus condition and contour type on the patterns shown in section 3.1. As shown in Figure 8, the duration of the pre-obstruent vowel in the implementation of the fortis-lenis contrast correlates with the type of intonation contour produced as follows: level contours (high, low) are associated with the shortest vowel durations; complex contours (rise-fall, fall-rise) are associated with the longest vowel durations and simple contours fall in between. This effect is highly significant ( $F(5, 715)=29.83$ ,  $SS=92191$ ,  $p<.0001$ ). We found no other effects of contour type on the three correlates of the contrast.

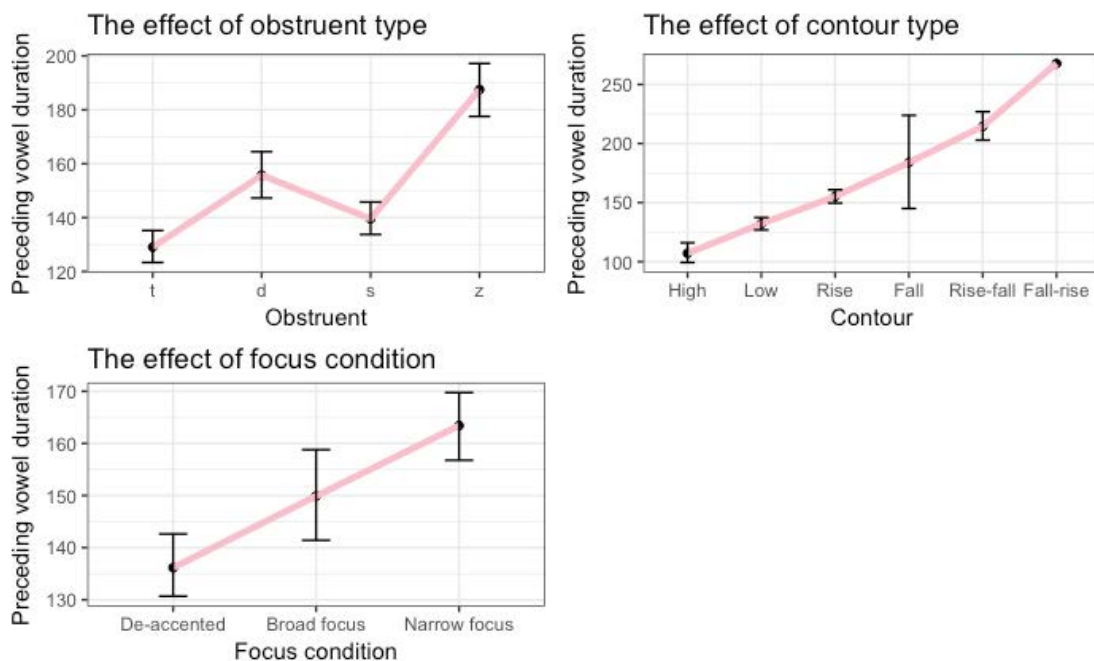


Figure 8. Effects of obstruent, contour and focus on pre-obstruent vowel duration (ms).

The role of the duration of the pre-obstruent vowel in the phonetic implementation of the fortis-lenis contrast is strongly affected by focus condition, with the shortest vowel durations being associated with the de-accented condition ( $\bar{x}=136.16\text{ms}$ ,  $SD=44.62$ ), the longest with the narrow focus condition ( $\bar{x}=163.42\text{ms}$ ,  $SD=63.96$ ), and broad focus in between ( $\bar{x}=149.89\text{ms}$ ,  $SD=55.54$ ). However, focus condition did not improve the model as a main effect. Its interaction with the type of consonant, on the other hand, was a significant predictor of the duration of the preceding vowel and was kept as part of the final model ( $F(6,$

715)=12.875,  $SS=10662$ ,  $p=.0089$ )<sup>12</sup>. Figure 9 illustrates this interaction. It can be seen from this graph that all four obstruents adhere to the trend for longer durations of the preceding vowel in the narrow focus condition described above, but that the main drivers of the trend are the voiced pair of obstruents, /d/ and /z/. Post-hoc Tukey tests reveal that the difference in length of the preceding vowel between de-accented and narrow focus is highly significant with /d/ ( $p=.0009$ ) and /z/ ( $p=.008$ ), while it is not significant with /t/ ( $p=.93$ ) or /s/ ( $p=.98$ ).

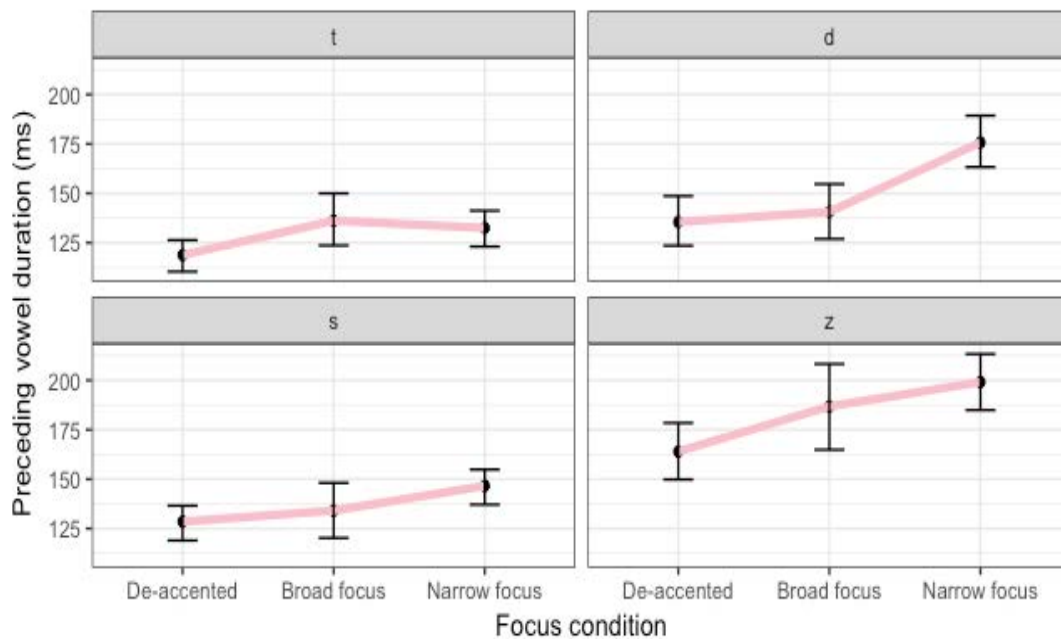


Figure 9. The interaction between obstruent and focus on pre-obstruent vowel duration (ms).

The regression model investigating V/C ratio and focus contains a significant interaction with consonant type ( $F(6,724)=3.6932$ ,  $SS=8.4122$ ,  $p=.0013$ )<sup>13</sup>. Tukey tests of the interaction show a magnification of the effect in the narrow focus condition (/t/-/d/:  $p<.0001$ ; /s/-/z/:  $p<.0001$ ) compared to the de-accented condition (/t/-/d/:  $p=.0002$ ; /s/-/z/:  $p=.0015$ ). These results are summarised in Figure 10.

<sup>12</sup> Final model:  $\text{lmer}(\text{vowel duration} \sim \text{consonant} * \text{focus condition} + \text{intonational contour} + (1|\text{speaker}) + (1|\text{word}), \text{data}=\text{data})$ .

<sup>13</sup> Final model:  $\text{lmer}(\text{V/C ratio} \sim \text{consonant} + \text{focus condition} + \text{intonation contour} + \text{consonant} * \text{focus condition} + (1|\text{speaker}) + (1|\text{word}), \text{data}=\text{data})$ .

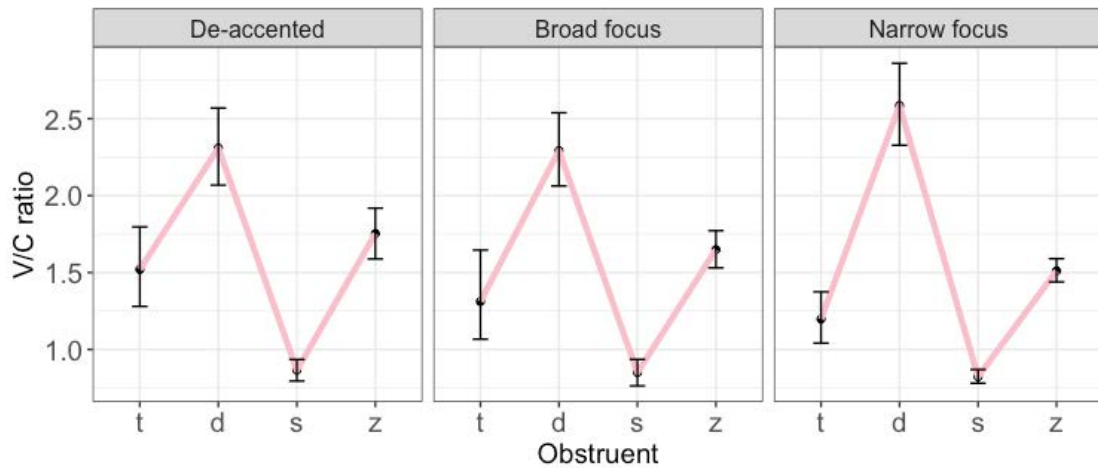


Figure 10. Focus effects on the preceding vowel duration (ms)

Regarding voicing, focus condition has a small, but non-significant, effect on its presence (Figure 11):

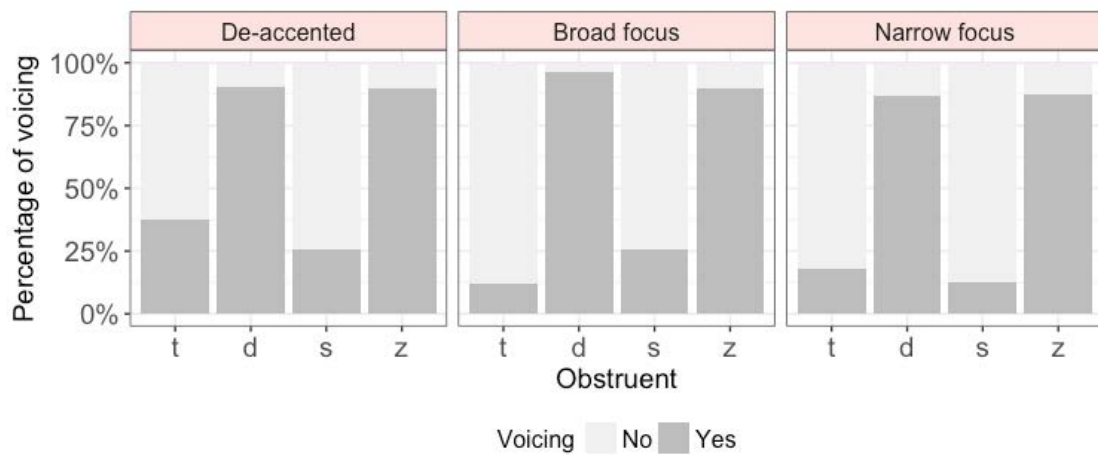


Figure 11. Presence of voicing (%) and the obstruent by focus condition.

On the whole, the fortis-lenis contrast distinction by voicing is more obvious in the broad focus condition as opposed to the de-accented condition, but this trend only approaches significance ( $SS=4148$ ,  $p=.078$ ) and does not improve the model. Interestingly, there is a significant interaction between obstruent and focus condition: the /t/-/d/ contrast is distinguished by the presence of voicing significantly more so in broad

and narrow focus than in the de-accented condition ( $F(6,709)=2.3937$ ,  $SS=6048.3$ ,  $p=.0269$ ).<sup>14</sup>

In addition to the analyses presented so far, we also carried out a random forest analysis. This analysis involved the presence of voicing, the duration of the preceding vowel, and the V/C ratio as predictors. It revealed that, when it comes to determining the identity of the obstruent in question, V/C ratio is more important than the presence of voicing, which is in turn more important than the duration of the preceding vowel. Considering the /s-/z/ contrast, the factor most important for the identity of the obstruent is V/C ratio, followed by the presence of voicing, followed by vowel duration (voicing: 51.037; vowel duration: 24.414; V/C ratio: 102.141).<sup>15</sup> Regarding the /t-/d/ contrast, the same tendency is observed: the most important factor determining the identity of the obstruent is V/C ratio, followed by the presence of voicing, followed by vowel duration (voicing: 59.923; vowel duration: 33.701; V/C ratio: 72.435). These results do not differ depending on whether the duration of the preceding vowel is normalised or raw.

### 3.3. Other correlates of the contrast

In this section, we offer a qualitative discussion based on descriptive statistics for potential correlates of the fortis-lenis contrast other than the preceding vowel duration, presence of voicing, and vowel/consonant durational ratio. Apart from the usually discussed correlates of the fortis-lenis contrast in English (as described in 3.1.-2.), we also report a range of other correlates of the contrast displayed by the individual department members.

Firstly, pre-aspiration is found consistently with all speakers only in their fortis obstruents<sup>16</sup>, and always more frequently in /s/ than /t/ (Figure 12). Stephen and Mark are somewhat shy pre-aspirators.

---

<sup>14</sup> Final model:  $\text{lmer}(\text{presence of voicing} \sim \text{consonant} + \text{focus condition} + \text{intonation contour} + \text{consonant} * \text{focus condition} + \text{consonant} * \text{intonation contour} + (1 | \text{speaker}) + (1 | \text{word}), \text{data} = \text{data})$ .

<sup>15</sup> These numbers reflect the Gini importance index, which is used to refer to node impurity in the models (Louppe, Wehenkel, Suter, & Geurts, 2013). The higher the number, the higher the misclassification of the obstruent is (as fortis vs lenis) if the variable of interest is left out. The more important variables will therefore show higher numbers.

<sup>16</sup> Sophia is a bit of an outlier in that she pre-aspirates her /d/ once and also her /z/ once.



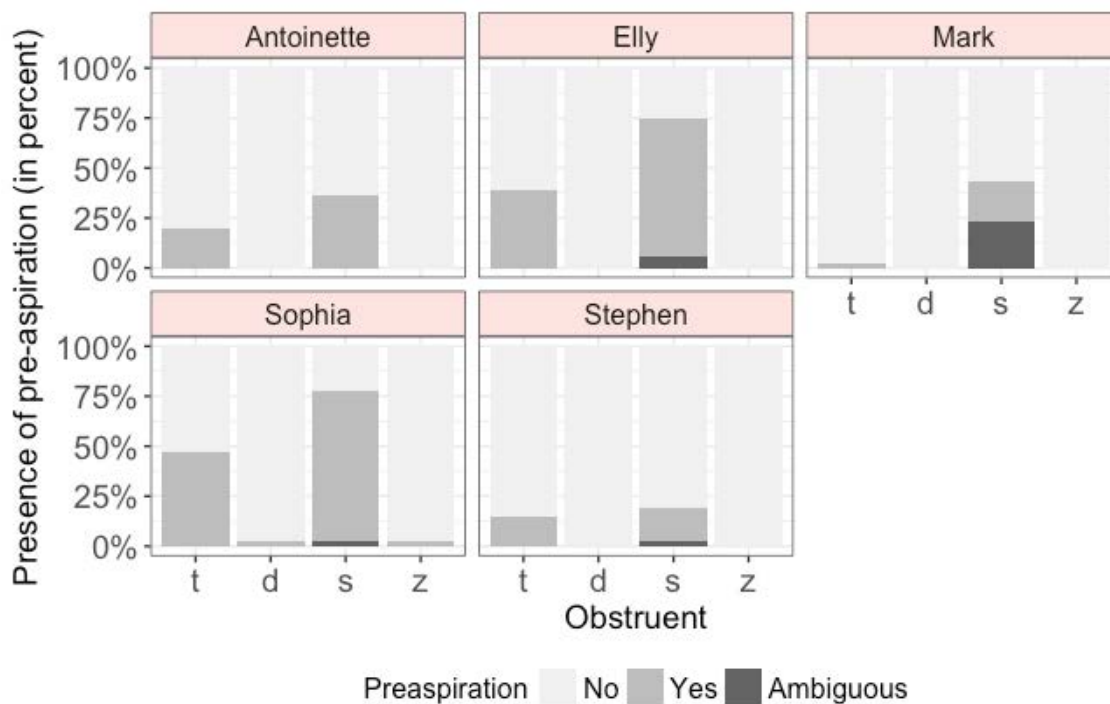


Figure 12. Presence of pre-aspiration (%) and the obstruent by individual.

As shown in Figure 13, affrication is clearly used as a correlate of the fortis-lenis /t-/d/ contrast by Stephen, who never affricates his /d/'s, but affricates 94% of his /t/'s. Mark follows suit in that he never affricates his /d/'s, but his /t/'s are not as frequently affricated as Stephen's, reaching only 34%. Whilst Sophia still gets to join the club of fortis plosive affricators in that she affricates her /t/'s more frequently than her /d/'s, the difference is far from clearcut in her case (/t/: 63%; /d/: 50%). Antoinette, on the other hand, shows a pattern unlike that of the three predominant /t/-affricators. Antoinette does affricate, but her affrication is higher in the context of /d/ than that of /t/ (/t/: 22%; /d/: 59%), which makes her a predominant /d/-affricator. As discussed further below, this is because her /t/'s are frequently spirantised. Finally, Elly also shows a unique behaviour by simply affricating both /t/ and /d/ more or less obligatorily, thus not utilising the feature as a correlate of the contrast at all.



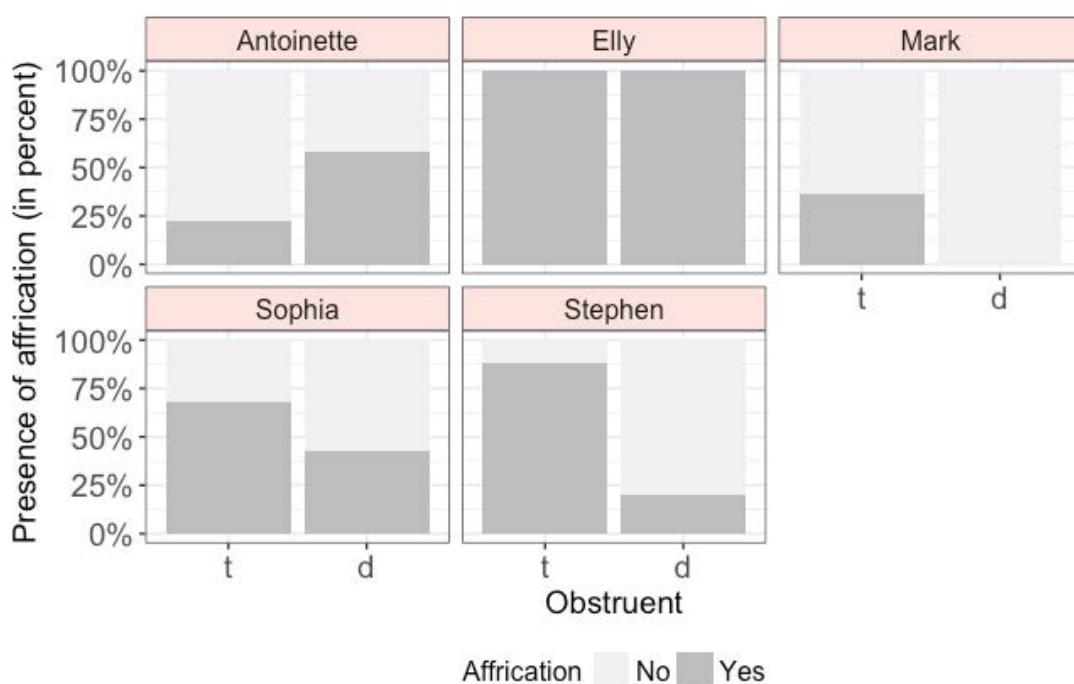


Figure 13. Presence of affrication (%) and the obstruent by individual.

If we consider the traditional descriptions of English as a post-aspirating language, looking into the use of post-aspiration as a potential correlate of the /t-/d/ contrast in our five speakers presents us with a surprise: post-aspiration is fairly marginal in the dataset. Moreover, it does not consistently occur only in /t/, as illustrated in Figure 14, but also in /d/. Although Antoinette is the only “well-behaved” department member in that she post-aspirates only her /t/, it needs to be noted that even Antoinette does not quite present us with the canonical post-aspirated /t/ because, in her case, the post-aspiration is often found following a fully spirantised /t/ (see further below). To make the situation even more variable, Stephen presents us with some plosive releases which are ambiguous as to their being post-aspirated or unaspirated. Stephen and Mark are the only flappers in our dataset (Figure 15). Stephen is only a sporadic /d/-flapper (8%), while Mark flaps both his /t/’s (25%) and his /d/’s (28%), at a rate of application fairly comparably across his /t/ and /d/ categories. In contrast, the female members of the department do not flap. On the whole then, we can say that flapping is not a strategy used to distinguish the /t-/d/ contrast by the five department members.

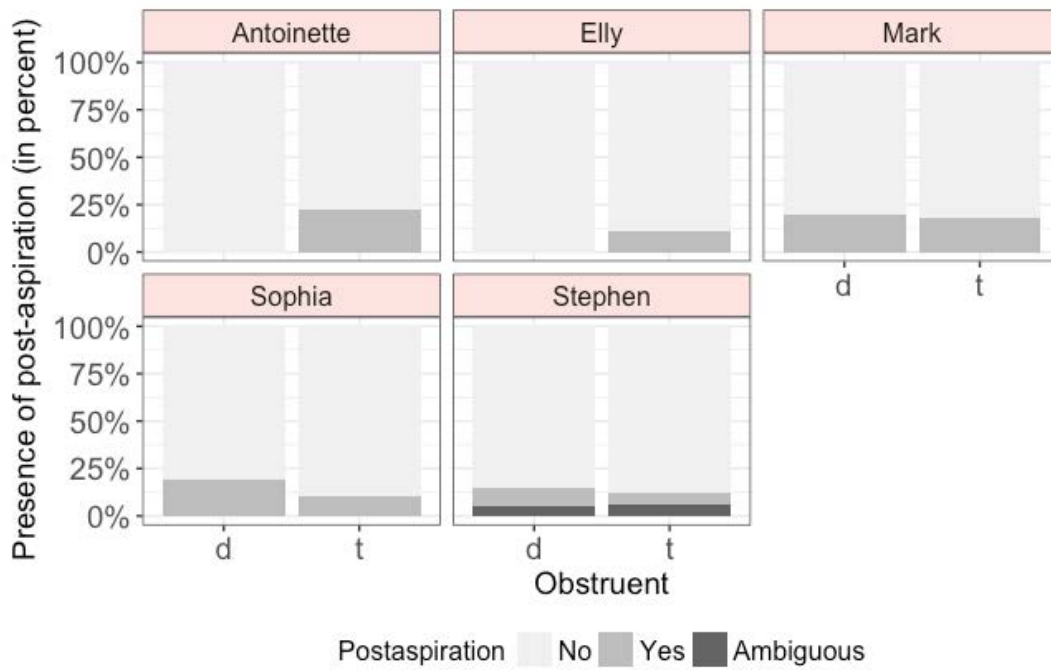


Figure 14. Presence of post-aspiration (%) and the obstruent by individual.

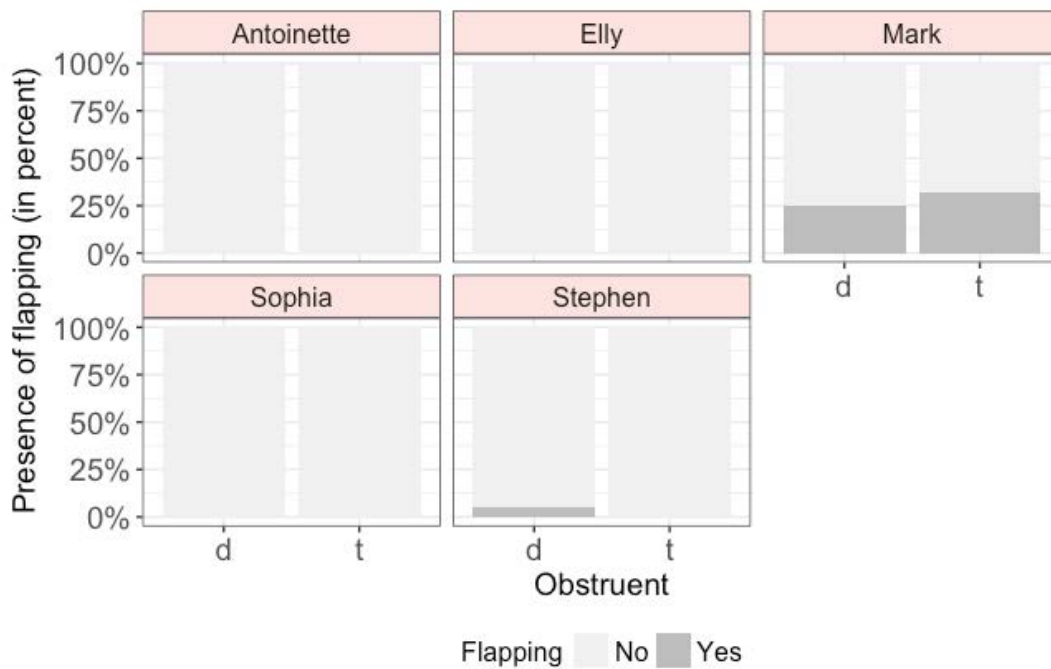


Figure 15. Presence of flapping (%) and the obstruent by individual.

As shown in Figure 16, Stephen is in the lead of the team of non-spirantisers, contrary to what has been claimed about /t/ spirantisation in southern Irish English dialects (Hickey, 2004). The rest of his closure-preserving team, Elly and Mark, do spirantise but only marginally so. Antoinette, on the other hand, very frequently spirantises both /t/ (78%) and /d/ (41%). Sophia also enjoys spirantisation in both obstruents (/t/: 37%; /d/: 10%). Importantly, both Antoinette and Sophia spirantise their /t/'s more frequently than their /d/'s. In addition, for both Antoinette and Sophia, full spirantisation is more frequently associated with their /t/'s and semi-spirantisation with their /d/'s.

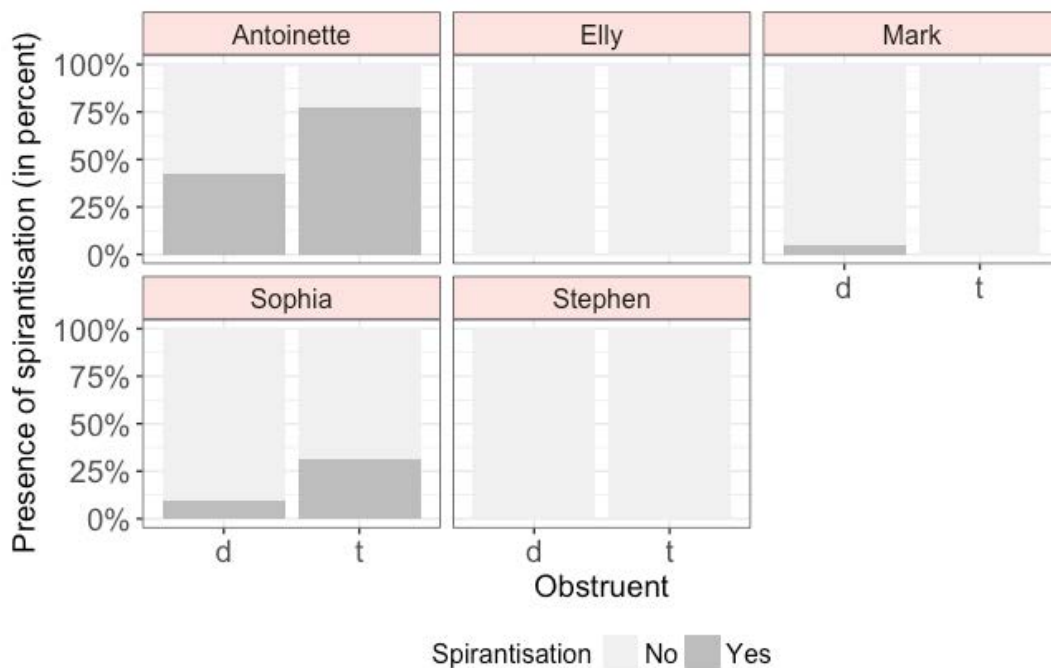


Figure 16. Presence of spirantisation (%) and the obstruent by individual.

### 3.4. Other correlates of the contrast and prosodic effects

As shown above, pre-aspiration contributes to the phonetic implementation of the fortis-lenis contrast in all five department members, albeit marginally so in some cases. Furthermore, Sophia, Stephen, Mark, and Antoinette use affrication as one of the correlates of the /t/-/d/ contrast. Finally, Sophia and Antoinette also employ spirantisation to distinguish /t/ and /d/. In this section, we look into whether these three features in the relevant speakers are subject to effects of focus and, if so, to what extent.

Firstly, as Figure 17 shows, pre-aspiration is the most frequent in the narrow focus condition, and this applies to all four obstruents. More specifically, /s/ is pre-aspirated 50% of the times in the focus condition, but only 34% of the times in both the broad focus and the de-accented conditions. In the same vein, /t/ is pre-aspirated 28% of the times in the narrow focus condition, which is higher than in the broad focus (21%) and the de-accented (18%) conditions.

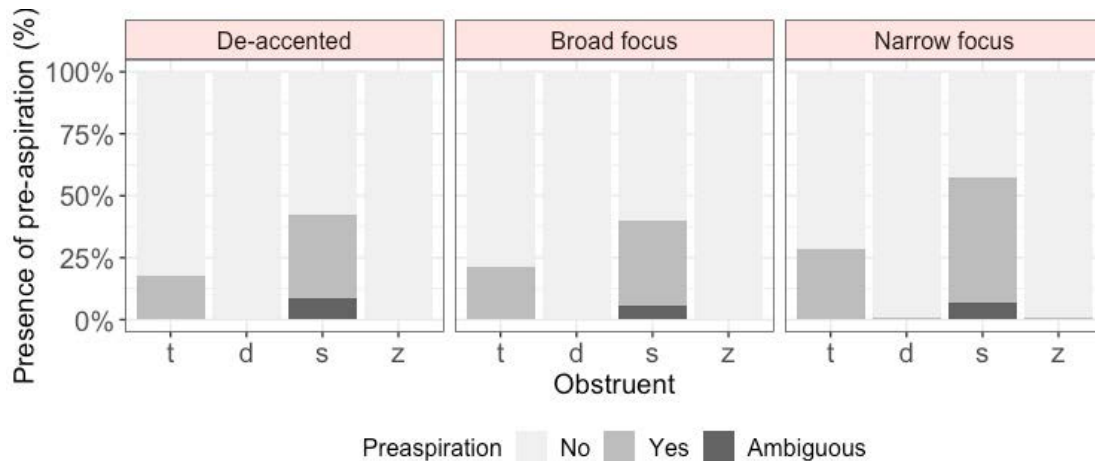


Figure 17. Presence of pre-aspiration (%) and the obstruent by focus condition.

Interestingly, the two cases of pre-aspirated lenis obstruents<sup>17</sup> are both found in the focus condition. On the whole, pre-aspiration application is a somewhat better correlate of the fortis-lenis contrast in the focus condition than; on the other hand, its marginal occurrence in the lenis series in this prosodic condition undermines this finding. The presence of affrication and that of spirantisation are not affected by focus condition in the relevant speakers.

#### 4. Discussion and conclusions

This study has examined the effects of different levels of prosodic focus on the realisation of the fortis-lenis contrast, and on obstruent realisation more generally. Furthermore, we have taken a brief look at the contribution of  $f_0$  variation on the three main correlates of the fortis-lenis contrast under investigation: the duration of the preceding vowel, the voicing frequency, and the vowel/consonant ratio.

<sup>17</sup> As we have seen earlier, Sophia is the lenis obstruent pre-aspiration culprit.

Investigating these correlates on their own, we have found that V/C ratio is the most important of the three correlates in our data, followed by the presence of voicing. Our analyses of the duration of the pre-obstruent vowel did not yield significant results, although post-hoc tests revealed a significant pairwise difference between the duration of the vowel preceding /t/ as opposed to that preceding /z/, by which the duration was shortest before /t/ and longest before /z/. With respect to V/C ratio, there was an effect of consonant type whereby lenis obstruents had higher V/C ratios, i.e. the preceding vowel was proportionally longer than the consonant more so prior to lenis than fortis obstruents. In addition, the presence of voicing in the obstruent also serves as a correlate of the contrast.

When we added focus and the type of intonation contour to the models we saw that these have predictive effects on several of the correlates of the implementation of the fortis-lenis contrast. In formulating our first research question, we hypothesised that increased levels of focal prominence would correlate with higher reliability and availability of the consonantal correlates (MacWhinney, 2001, 2012), with a decrease as the focal prominence decreases. After considering the effect of focus on pre-obstruent vowel duration, we found a trend for shorter vowel durations in the de-accented condition, through mid-level durations in broad focus and with the longest durations in the narrow focus condition. This trend was not significant as a main effect; however, an interaction between focus and the type of consonant *was* significant. Closer scrutiny showed that the lenis set of obstruents was the main driver of the correlation between focus condition and vowel length. A similar significant interaction between focus and consonant type was reported for the V/C ratio correlate, and a non-significant interaction was also reported for the third correlate, namely the presence of voicing. In both these cases, the fortis/lenis distinction was magnified in the narrow focus condition as compared to the two other conditions. Our prediction is thus partly confirmed: different levels of focus as elicited in this study correlate at least partially with stronger manifestations of the fortis-lenis contrast with all three main correlates measured, although these correlations were not always statistically significant.

Considering other potential correlates of the contrast (pre-aspiration, post-aspiration, affrication, spirantisation, flapping, glottalling, ejectives), we found that, firstly, all of the speakers use pre-aspiration to contribute to the fortis-lenis distinction (although a bit more strongly for the fricative rather than the plosive pair). Secondly,

affrication seems to contribute to the /t/-/d/ distinction for four of the speakers, with /t/ being affricated more frequently for three of these. Aspiration is only marginally employed to distinguish /t/ from /d/, and only in one speaker. Finally, spirantisation is higher for /t/ than /d/ in the two speakers who spirantise.

With regard to obstruent realisation, in connection with our second research question we predicted that increased levels of prosodic prominence would correlate with increasing amounts of variation in obstruent realisation, especially in the case of pre-aspiration. In our data, only pre-aspiration is sensitive to the different focus conditions in that it applies more frequently in the narrow focus condition than in the broad focus and the de-accented positions. This indicates that it is a somewhat more robust correlate of the contrast in the narrow focus condition. The fact that we find two cases of pre-aspirated lenis obstruents in the narrow focus condition may suggest that pre-aspiration as such is more likely to innovate in a more prosodically prominent condition. This fits in well with the overall findings of pre-aspiration being a laryngeal aspect of stressed syllables (Hejná, 2015, for an overview), and constitutes a partial confirmation of our hypotheses.

Finally, we briefly investigated the effects of different nuclear contours on obstruent realisation as formulated in our third research question. Our hypotheses include a tendency for increased contrast strength with high or rising intonational movements, as well as a correlation between shorter duration and level or simple contours (and between longer durations and complex contours). This has often been referred to in the literature (since English is a compression rather than a truncation language – Grabe, 1998; e.g. Ohala, 1978), but has not to our knowledge been attested through experimental work. We did not find any significant correlates between the type of nuclear contour and the presence of voicing or V/C ratio, but we did find a correlation with the duration of the preceding vowel whereby level contours (high, low) were associated with shorter vowel durations than simple contours (falls, rises), which were again associated with shorter vowel durations than complex contours (rise-falls, fall-rises). Our hypotheses stemming from the final research question are thus also partially fulfilled.

On the whole, then, there is a tendency for the correlates of the fortis-lenis obstruent contrast to be noticeably affected by prosodic focus and the complexity and directions of  $f_0$  movements. This suggests that prosody-segment interactions are a worthwhile avenue to pursue.

## 5. Implications for studies of focus, pitch contours and the fortis-lenis contrast

But how *exactly* should this avenue be explored? Firstly, the methodology used deserves further elaboration. As briefly reported in the results section, the finding that increased levels of focus affected the implementation of the fortis-lenis contrast to some extent, but did not yield many clear effects, may stem from difficulties in getting our department members to produce the intended levels of focus, from differences in the manifestations of narrow focus (and de-accentuation) produced by individual speakers, or from exaggerated or performed speech styles. While previous work has not tended to report on such difficulties, in our study different degrees of focus were generally realised in gradual rather than overtly categorical terms, and were highly speaker-specific. While we feel this probably reflects patterns found in spontaneous speech to some degree (something that needs to be confirmed in further research), our findings on focus need to be replicated and extended by further work.

Secondly, the general lack of effects of intonation contours on the realisation of consonants could in part be due to our lack of detailed acoustic measurements. The tendency of rising  $f_0$  to correlate with higher centres of gravity across fricatives (Niebuhr, 2012) indicates that such  $f_0$  movements may have effects on other aspects of frication, such as the intensity of the aspiration and the affrication noise components in obstruents and potentially other acoustic characteristics of their realisation as well. Another explanation might be that English is not a truncating language, unlike German, a fact which may limit the need to express the information carried by truncated intonational contours through consonantal means (as is the case in German). Further work could address these potential gaps by taking additional acoustic measures, such as spectral moments.

In addition, whilst the finding that vowel duration and the complexity of the intonational contour are correlated is in line with common wisdom in phonetics and phonology (see e.g. Ohala, 1978), to the best of our knowledge this has not in fact been demonstrated by experimental work for non-tone languages (and even with tone languages the picture is unclear, cf. Köhnlein, 2015, p. 232). Further work may therefore want to provide additional evidence for this finding using more controlled speech samples. In addition, high degrees of between-speaker differences suggest potential dialectal variation which could be fruitful

for future sociolinguistic work: there is more to consonantal variation than first meets the eye, and we conclude that it is not necessarily the case that vowels do in fact exhibit more variation than consonants.<sup>18</sup>

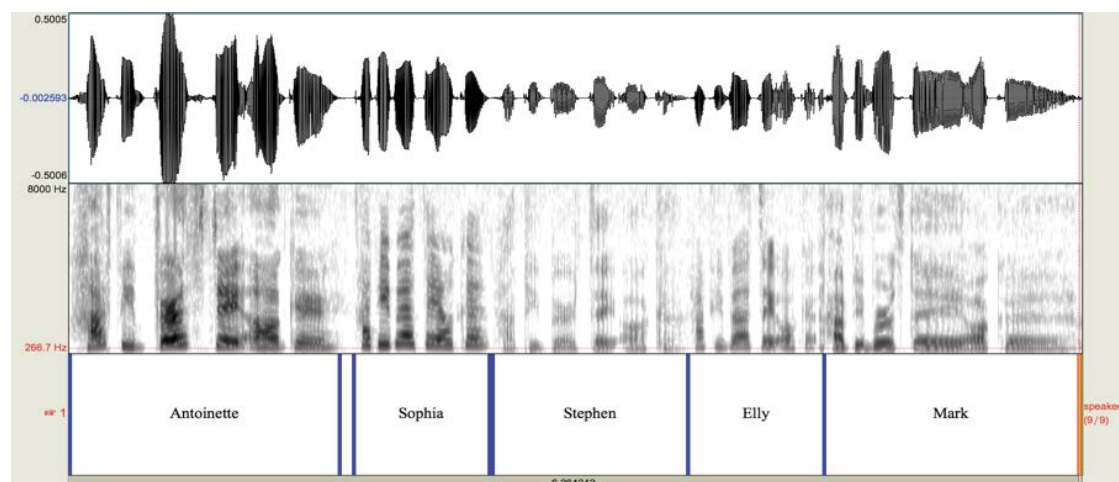
Finally, we note that it is not post-aspiration that distinguishes /t/ and /d/ foot-medially and -finally in our sample. For some speakers, affrication is employed as a correlate of the contrast, but this correlate is only a marginal strategy. Although we did not measure VOT, our observation of the data from the annotation of the other potential correlates suggests that VOT is an important correlate. If this is indeed the case, it is the release duration of the plosive (traditionally quantified as VOT) which is more important than whether the plosive is post-aspirated, affricated, both, or neither. This is in line with the findings available for Aberystwyth English (Hejná, 2016b).

### Acknowledgements

We would like to thank our five departmental guinea pigs as well as Oliver Niebuhr for his very useful comments on an earlier draft of this paper.

### Comments

Dearest Ocke, let your life be forever full of vowels, consonants, and prosodic phenomena. “Happy birthday, Ocke!” from us, but also from our five lovely participants: Antoinette, Sophia, Stephen, Elly, and Mark!



<sup>18</sup> However, it would be rather difficult to actually test this claim.



## References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Baumann, S., Becker, J., Grice, M. & Mücke, D. (2007). Tonal and articulatory marking of focus in German. *Proceedings of ICPHS 2007*, Saarbrücken, 1029-1032.
- Boersma, P., & Weenink, D. (1992-2017). *Praat: doing phonetics by computer*. Version 6.0.25.
- Baltazani, M., & Jun, S.-A. (1999). Focus and topic intonation in Greek. *Proceedings of ICPHS 1999*, San Francisco, 1305-1308.
- Bohn, O.-S., & Caudery, T. (2017). *The sounds of English. An activity-based course in English phonetics and phonology*. Aarhus University Press, Aarhus.
- Braver, A. (2011). Incomplete neutralization in American English flapping: production study. *University of Pennsylvania Working Papers in Linguistics / Proceedings of the 34th Annual Penn Linguistics Colloquium*, 17(1), 29-40.
- Cox, F., & Palethorpe, S. (2007). Australian English. *Journal of the International Phonetic Association*, 37(3), 341-350.
- Cruttenden, A. (1996) *Intonation*. Cambridge: Cambridge University Press.
- Derrick, D., & Gick, B. (2011). Individual variation in English flaps and taps: a case of categorical phonetics. *The Canadian Journal of Linguistics / La Revue Canadienne de Linguistique*, 56(3), 307-319.
- D'Imperio, M. (2001). Focus and tonal structure in Neapolitan Italian. *Speech Communication*, 33, 339-356.
- Docherty, G. (1992). *The timing of voicing in British English obstruents*. New York: Foris Publications.
- Dohen, M., Loevenbruck, H., & Hill, H. (2006). Visual Correlates of Prosodic Contrastive Focus in French: Description and Inter-Speaker Variabilities, *Proceedings of Speech Prosody 2006*, Dresden, 221-224.
- Fox, J. (2003). Effect displays in R for Generalised Linear Models. *Journal of Statistical Software*, 8(15), 1-27.
- Fox, J., & Palethorpe, S. (2007). Illustrations of the IPA: Australian English. *Journal of the International Phonetic Association*, 37(3), 341-350.
- Gordeeva, O., & Scobbie, J. (2013). A phonetically versatile contrast: pulmonic and glottalic voicelessness in Scottish English obstruents and voice quality. *Journal of the International Phonetic Association*, 43, 249-271.
- Grabe, E.. (1998). *Comparative Intonational Phonology: English and German*. MPI Series in Psycholinguistics 7, Wageningen, Ponsen en Looien.
- Görs, K., & O. Niebuhr. (2012). Hocus Focus – How prosodic profiles of contrastive focus emerge and change in different elicitation contexts. *Proceedings of Speech Prosody*, Shanghai, China, 262-265.
- Hejná, M. (2015). *Pre-aspiration in Welsh English: a case study of Aberystwyth* (PhD thesis), University of Manchester, Manchester.

- Hejná, M. (2016a). Pre-aspiration: manual on acoustic analyses 1.1. Manuscript released on LingBuzz.
- Hejná, M. (2016b). Multiplicity of the acoustic correlates of the fortis-lenis contrast: plosives in Aberystwyth English. *Proceedings of Interspeech 2016*, San Francisco, 3147-3151.
- Hejná, M., & Kimper, W. (Submitted). Pre-closure laryngeal properties as cues to the fortis-lenis plosive contrast in British English. *Yearbook of the Poznań Linguistic Meeting 2017*.
- Hickey, R. (2004). The phonology of Irish English. In B. Kortmann & E. W. Schneider (Eds.), *Handbook of varieties of English* (Vol. 1, *Phonology*, pp. 68-97). Berlin: Mouton de Gruyter.
- Hua, C., Li, B., & Wayland, R. (this volume). *Native and non-native English speakers' assessment of nuclear stress produced by Chinese Learners of English*.
- Iverson, G. K., & Salmons, J. C. (2006). On the typology of final laryngeal neutralisation: evolutionary phonology and laryngeal realism. *Theoretical Linguistics*, 32(2), 205-216.
- Jansen, W. (2004). *Laryngeal contrast and phonetic voicing: a laboratory phonological approach to English, Hungarian, and Dutch*. PhD thesis, University of Groningen, Groningen.
- de Jong, K. J. (1995). The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, 97(1), 491-504.
- Kingston, J. (1986). Are f<sub>0</sub> differences after stops accidental or deliberate?. *111th Meeting of the Acoustical Society of America*, S27.
- Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: the [voice] contrast. *Journal of Phonetics*, 36, 28-54.
- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59(5), 1208-1220.
- Kohler, K. (1984). Phonetic explanation in phonology. *Phonetica*, 41, 150-171.
- Kohler, K. (2011). Communicative functions integrate segments in prosodies and prosodies in segments. *Phonetica*, 68, 26-56.
- Köhnlein, B. Thee complex durational relationship of contour tones and level tones. *Diachronica*, 32(2), 231-267.
- Kügler, F. (2008). The role of duration as a phonetic correlate of focus. *Proceedings of Speech Prosody 2008*, Campinas, Brazil, 591-594.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1-26.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

- Ladefoged, P. (1968). *A phonetic study of West African languages*. (2nd edition). Cambridge: Cambridge University Press.
- Leemann, A., Kolly, M.-J., Li, Y., Chan, R., Kwek, G., & Jespersen, A. (2016). Towards a typology of prominence perception: the role of duration. *Proceedings of Speech Prosody 2016*, Boston, MA.
- Loupe, G., Wehenkel, L., Sutura, A., & Geurts, P. (2013). Understanding variable importances in forests and randomized trees. *NIPS' 13 Proceedings of the 26th International Conference on Neural Information Processing Systems*, vol. 1, Lake Tahoe, Nevada.
- MacWhinney, B. (2001). The Competition Model: the input, the context, and the brain. In P. Robinson (Ed.), *Cognition and Second Language Instruction*, (pp. 69-90). Cambridge: Cambridge University Press.
- MacWhinney, B. (2013). The logic of the unified model. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 211-227). London/New York: Routledge.
- Mennen, I. (2006). Phonetic and phonological influences in non-native intonation: an overview for language teachers. *Working paper WP-9*. Edinburgh.
- Mücke, D., & Grice M. (2014). The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation? *Journal of Phonetics*, 44, 47-61.
- Niebuhr, O. (2008). Coding of intonational meanings beyond F0: evidence from utterance-final /t/ aspiration in German. *Journal of the Acoustical Society of America*, 124(2), 1252-63.
- Niebuhr, O. (2012). At the Edge of Intonation: The interplay of utterance-final F0 movements and voiceless fricative sounds. *Phonetica*, 69(1-2), 7-27.
- Niebuhr, O. (2013). Coding of intonational meanings beyond F0: evidence from utterance-final /t/ aspiration in German. *Journal of the Acoustical Society of America*, 124, 1252.
- Niebuhr, O. (this volume). Pitch accents as prosodic constructions – Evidence from pitch-accent specific micro-rhythms in German.
- Niebuhr, O. & Michaud, A. (2015) Speech data acquisition - The underestimated challenge. *Kieler Arbeiten in Linguistik und Phonetik (KALIPH0)*, 3, 1-42.
- Niebuhr, O., Lill, C., & Neuschulz, J. (2011). At the segment-prosody divide: the interplay of intonation, sibilant pitch and sibilant assimilation. *Proceedings of the 17th ICPHS*, Hong Kong, 1478-1481.
- Niebuhr, O., & Pfitzinger, H. (2010) On pitch-accent identification – The role of syllable duration and intensity. *Proceedings of Speech Prosody 2010*, Chicago, US, 1-4.
- Nolan, F. (2006). Intonation. In B. Aarts & A. McMahon (Eds.), *Handbook of English linguistics* (pp. 433-457). Oxford: Blackwell.
- Norcliffe, E., & Jaeger, T. F. (2005). Accent-free prosodic phrases? Accents and phrasing in the post-nuclear domain. *Proceedings of Interspeech 2005*.

- Ohala, J. J. (1978). Production of tone. In V. A. Fromkin (Ed.), *Tone: a linguistic survey*, (pp. 5-39). New York: Academic Press.
- Penney, J., Cox, F., Miles, K., & Palethorpe, S. (2018). Glottalisation as a cue to coda consonant voicing in Australian English. *Journal of Phonetics*, 66, 161-184.
- R Core Team (2018). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (<https://www.R-project.org/>).
- RStudio Team (2015). *RStudio: integrated development for R*. RStudio, Inc., Boston, MA (<http://www.rstudio.com/>).
- Sarkar, D. (2008). *Lattice: multivariate data visualization with R*. Springer, New York.
- Scobbie, J. (2005). Interspeaker variation among Shetland Islanders as the long term outcome of dialectally varied input: speech production evidence for fine-grained linguistic plasticity. *QMUC Speech Science Research Centre Working Paper WP 2*.
- Smith, C. (1997). The devoicing of /z/ in American English: effects of local and prosodic context. *Journal of Phonetics*, 25(4), 471-500.
- Smith, J. L. (2002). *Phonological augmentation in prominent positions* (PhD thesis). University of Massachusetts Amherst, Massachusetts.
- Steriade, D. (1998). Alternatives to syllable-based accounts of consonantal phonotactics. *Proceedings of LP '98*, 205-245.
- Stevens, M., & Hajek, J. (2005). Spirantization of /p t k/ in Sieneese Italian and so-called semi-fricatives. *Proceedings of Interspeech 2005*, Lisbon, 2893-2897.
- Su, V. W. Y. (2007). *The gender variable in Australian English stop consonant production* (BA thesis). The University of Melbourne, Melbourne.
- Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, 34(3), 434-464.
- Trousdale, G. (2010). *An introduction to English sociolinguistics*. Edinburgh University Press, Edinburgh.
- Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35, 445-457.
- Wagner, P., Origlia, A., Avesani, C., Christodoulides, G., Cutugno, F., D'Imperio, M., Escudero Mancebo, D., Gili Fivela, B., Lacheret, A., Ludusan, B., Moniz, H., Ní Chasaide, A., Niebuhr, O., Rousier-Vercruyssen, L., Simon, A.-C., Šimko, J., Tesser, F., & Vainio, M. (2015) Different parts of the same elephant: a roadmap to disentangle and connect different perspectives on prosodic prominence. *Proceedings of the 18th International Congress of Phonetic Sciences*. University of Glasgow, Glasgow.

- Wells, J. C. (1990). Syllabification and allophony. In S. Ramsaran (Ed.), *Studies in the pronunciation of English, a commemorative volume in honour of A. C. Gimson*, (pp. 76-86), New York: Routledge.
- Wright, M. N., & Ziegler, A. (2017). ranger: a fast implementation of Random Forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1), 1-17.
- Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33, 159-197.
- Zue, V. W., & Laferriere, M. (1979). Acoustic study of medial /t, d/ in American English. *Journal of the Acoustical Society of America*, 66, 1039-1050.

## **Production and Perception of Korean Word-level Prominence by Older and Younger Korean Speakers**

Goun Lee  
Sungkyunkwan University

Allard Jongman  
University of Kansas

### **Abstract**

Prominence refers to the relative emphasis that may be given to a syllable in a word (word-level prominence) or to one or more words in a phrase (phrase-level prominence). Korean has been claimed to have both word-level (Ko, 2013) and phrase-level (Jun, 1996) prominence, with the former realized mainly with duration and the latter with F0 height. However, given the claim that younger Korean speakers have lost duration as the main cue expressing word-level prominence (Kim & Han, 1998; Magen & Blumstein, 1993), it is not clear if and how younger Korean speakers produce word-level prominence. Thus, the current study aims to investigate whether Korean still has word-level prominence. In the acoustic study of the production of Korean word-level prominence (Experiment 1), measurements of duration, intensity, and F0 on (so-called) Korean stress minimal pairs by older and younger Korean speakers revealed that only duration distinguishes Korean word-level prominence. A perception study on word-level prominence in Korean (Experiment 2) revealed that both older and younger Korean listeners weighted the duration cue most heavily in identifying minimal pairs when two of the suprasegmental cues were orthogonally manipulated in each syllable. Interestingly, this perceptual weighting was only observed in the first syllable: none of the listeners changed their perception when cues were signaling second-syllable stress. Based on these acoustic and perceptual findings we conclude that Korean does not have word-level prominence, but only has a phonemic vowel length distinction.

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 271-301). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.

## **1. Introduction**

Linguistic prominence is comprised of two levels of prosodic cues – word-level prominence and phrasal-level prominence. Word-level prominence creates lexical contrasts based on the acoustic manifestation of at least one cue, while phrasal-level prominence is conveyed by F0 peaks or valleys that express context-dependent pitch accents, which distinguish a prosodic boundary between words (e.g., Beckman, 1986; Cooper, Eady, & Mueller, 1985; Fry, 1958; Shport & Redford, 2014). For languages with word-level prominence, lexical prosody expresses whether certain syllables are more prominent than neighboring syllables within the same word. The prominence can be realized by multiple suprasegmental cues such as duration, pitch, and intensity. However, there is no absolute value that determines a prominent syllable: rather, the concept of strong-weak is abstract and relative to the adjacent syllables.

Languages selectively pick and choose which cues to use in expressing prominence. In stress languages like English, primary stress may be expressed by more than one acoustic cue, including increased F0, longer duration, and higher intensity compared to unstressed syllables (Fry, 1955; Fry, 1958; Gay, 1978). Segmental cues like vowel reduction can also express lexical stress (Gay, 1978; Koopmans-Van Beinum, 1980). For languages with phrasal-level prominence, pitch is used to group prosodic structure together that is determined by the domain of the accentual phrase. For example, in Japanese, which has both word-level prominence as well as phrasal-level prominence, a low boundary tone occurs at the beginning of every utterance and at the AP (Accentual Phrase)-final boundary. Thus, when this low tone occurs within a sentence, listeners interpret it as belonging to the phrasal boundary (Beckman & Pierrehumbert, 1986).

However, distinguishing the cues that are used to mark word-level prominence from those used to express phrasal-level prominence might be difficult, because the same acoustic correlates that are used to indicate word-level prominence in stress languages – F0 and duration – are utilized to indicate prosodic prominence as well. Cross-linguistically, syllables in sentence-final position are lengthened, and pitch is raised at the non-sentence-final phrasal boundary (Beckman, 1986; Tyler & Cutler, 2009; see Japanese for a low boundary tone at non-sentence-final phrasal-boundary position).

Regarding Korean, it is sometimes claimed that Korean has both word-level prominence (i.e., stress) and phrasal-level prominence. Previous research has claimed that Korean stress is mainly realized in terms

of duration (e.g., Ko, 2013), and phrasal-level prominence realized with pitch (at Accentual Phrase (AP)-level), intensity (at AP-initial position), and duration (at Intonational Phrase (IP)-final position) (Jun, 1993; 1998). While duration is claimed to be a cue to stress, the same cue can also mark the phrasal boundary in expressing phrasal-level prominence in Korean. In the following section, we will first briefly review Korean phrasal-level prominence and word-level prominence, and then discuss the cues that are used to indicate prominence at different levels.

### **1.1. Phrasal-level prominence in Korean**

Phrasal-level prominence plays a crucial role in speech segmentation and production by marking the phrasal boundary in terms of F0 or duration (e.g., Beckman, 1986). Phrasal boundary tones are marked with a raised F0 (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980), and phrase-final position is marked with an increased duration (Klatt, 1975; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). Listeners use these higher-level prosodic cues in segmenting ambiguous segmental information both in L1 and L2 speech (Cho et al., 2007; Christophe et al., 2004; Coughlin & Tremblay, 2012; Kim, 2004; Kim & Cho, 2009; Tremblay, Coughlin, Bahler, & Gaillard, 2012).

The Accentual Phrase (AP), an intonationally defined unit, can mark a phrasal boundary in Korean. The hierarchical structure of prosodic boundaries consists of the syllable, the Phonological Word (PW), the Accentual Phrase (AP), and the Intonational Phrase (IP). The edge of the larger unit always coincides with the edge of the smaller unit: the edge of IP always coincides with the edge of AP, and the edge of AP always coincides with the edge of PW (Selkirk, 1984). Jun (1993; 1998) proposed in her Accentual Phrase (AP) theory that Korean has intonationally defined units (AP) that pattern independently from the word-level prosody.

In the Korean AP system, the initial boundary of the prosodic domain is always delimited with a low tone and the final boundary with a high tone (i.e., #LHLH#; # refers to an AP boundary; each syllable is associated with a tone; Jun 1993; 1998). This LHLH tone pattern occurs when at least 4 syllables exist in one AP domain. When there are less than 4 syllables in an AP, 2 or 3 surface tone patterns appear by undershooting the initial two tones. For example, when an AP has 3 syllables, two different tone patterns can appear: a #LH# (or #HH#) pattern when the first two syllables are undershot, and a #LHH# (or #HHH#) pattern when



only the first syllable is undershot. When an AP has 2 syllables, only the #LH# (or #HH#) pattern can appear. When the domain-initial syllable is either aspirated or tense, the pitch is raised on the first syllable, bearing #HHLH# intonational pattern (Jun, 2000; Kim, 2004; Kim & Cho, 2009).

The IP-final boundary is also characterized by different tonal patterns such as L%, H%, LH%, and HL% (% refers to an IP boundary). When the AP boundary coincides with the IP boundary, the AP-final tone (L#) is overridden by the IP-boundary tone. At the IP-boundary, final lengthening also occurs along with the IP-boundary tone.

In addition to F<sub>0</sub>, previous studies have found that other cues, such as duration and amplitude, can also characterize the phrasal-level prominence. With respect to duration, phrase-final lengthening can mark IP boundaries in Korean. Jun (1993) and Chung et al., (1996) found that final lengthening does not occur at the AP level, but at the IP level. However, Cho and Keating (2001) and Oh (1998) found a small but significant AP-final lengthening effect compared to non-AP-final words. Although these studies are not consistent regarding AP-final lengthening in Korean, there is a strong consensus at least that phrase-final lengthening exists in IP-final position (Cho & Keating, 2001; Chung et al., 1996; Jun, 1993; 2000).

Amplitude can also mark both the AP-initial and -final boundary in Korean. Jun (1995) found that the amplitude of the first syllable was greater than that of the second syllable when a trisyllabic reiterative word like ‘mamama’ was embedded in sentence-medial position. The amplitude of the first syllable was comparable with that of the third syllable, but the third syllable was also marked with low F<sub>0</sub> because it was in AP-final position.

## **1.2. Word-level prominence in Korean**

Historically, Korean has been claimed to have a vowel length distinction. The long vowels only appear in the first syllable (Heo, 1965) and are realized with a rising tone. Although it had been widely accepted that Korean had a vowel length distinction for pairs with identical vowel quality (see the IPA manual, 1999, p. 44), most younger speakers have lost this distinction (Kim, 2001; Kim & Han, 1998; Magen & Blumstein, 1993) and only speakers from a few dialects like Chonnam (Ko, 2013) and North Kyungsang (Kenstowicz & Park, 2006) preserve the distinction.

This vowel length distinction has been argued to influence lexical stress<sup>1</sup> in Korean, which is realized as rhythmic shortening or lengthening. The traditional vowel shortening rule takes the long vowel as the underlying form, and posits that the long vowel undergoes vowel shortening, since the realization of the long vowel is only limited to the first syllable. When a monosyllabic word with a short vowel is combined with a monosyllabic word with a long vowel, the long vowel in the second syllable is shortened in the compound word. This vowel shortening also occurs when the long vowel is attached to vowel-initial suffixes (Kim-Renaud, 1974; B.-G. Lee, 1978), but optionally occurs when being attached to consonant-initial suffixes (Ko, 2002, 2013).

Based on the optional vowel shortening, Ko (2002, 2010, 2013) proposed that vowel shortening occurs to avoid accent clash.<sup>2</sup> Ko (2002) defines ‘stress’ to refer to “the metrical head physically realized on the surface”, and ‘accent’ to refer to “the underlying specification for prominence on a syllable” (Ko, 2002, p. 81, lines 27-29). In other words, ‘stress’ is the actual location at which physical correlates of word-level prominence are realized with acoustic features such as duration, F0, and amplitude, while ‘accent’ is the potential location of stress. Ko (2013) claimed that the long vowel is realized with the stress on the syllable, and when a syllable is not realized with stress, the vowel remains as a short vowel. Therefore, when a suffix that is carrying an accent is attached to a monosyllabic long vowel verb stem, the stem vowel is shortened in order to avoid the accent clash. However, when the accent-triggering suffix is attached to a disyllabic verb stem, the long vowel does not need to undergo vowel shortening. The accent of these suffixes is never realized because stress in Korean needs to fall on the first syllable.

### **1.3. Phonetic evidence for lexical stress in Korean**

Ko (2013) examined whether other acoustic correlates of lexical stress are realized along with vowel duration in two dialects of Korean. Since, unlike Seoul Korean, the Chonnam Korean dialect still preserves the vowel length distinction, Ko (2013) hypothesized that the Chonnam dialect might be more conservative in preserving lexical stress and therefore, the

---

<sup>1</sup> As lexical stress realized by a vowel length distinction is argued to be word-level prominence in Korea, we use ‘lexical stress’ and ‘word-level prominence’ interchangeably in this study.

<sup>2</sup> In her analysis, both ‘stress’ and ‘accent’ are used in reference to word-level prominence.

manifestation of the four acoustic correlates of stress will be more apparent compared to Seoul Korean. Two different age groups across two dialects (younger Chonnam speakers vs. older Seoul speakers) were chosen, because both the current Chonnam dialect and the traditional Seoul dialect (older Seoul speakers) still preserve the vowel length distinction. Four young Chonnam speakers (1 male, mean age = 34) and four old Seoul Korean speakers (2 males, mean age = 69) recorded 17 stress minimal pairs (e.g., *sákwa* ‘apology’ vs. *sakwá* ‘apple’) embedded in a contextually related sentence (e.g., As for **apples**, Taegu is famous for it) and in a contextually neutral sentence (e.g., ‘Please pronounce apple clearly’).<sup>3</sup> Three acoustic parameters – duration (ms), intensity (dB), and F0 (semitone) – were examined for the first and second vowel of the target word from the neutral sentence. Measurements averaged across the entire vowel from the stressed syllables (*sá* from *sákwa* ‘apology’) were compared to those of the unstressed syllables (*sa* from *sakwá* ‘apple’).

A series of paired t-tests found that young Chonnam speakers use duration, intensity, and F0 to distinguish the vowel-length minimal pairs. The vowel in the ‘stressed first syllables’ was 77.7 ms longer, 0.64 semitone higher, and 2.6 dB more intense than that in the ‘unstressed first syllables’. The vowel in the stressed second syllables was 35.72 ms longer, 1.51 semitone higher, and 1.8 dB more intense than that in the ‘unstressed second syllables’. On the other hand, older Seoul Korean speakers only used vowel duration in the first syllable and intensity in the second syllable to distinguish vowel-length minimal pairs. For the older Seoul speakers, the vowel in stressed first syllables was 96.71 ms longer than that in unstressed first syllables, and the vowel in stressed second syllables was 1.02 dB more intense than that in unstressed second syllables.

Additionally, results of two separate models of a mixed effect logistic regression indicated that, in the Chonnam dialect, all three correlates showed a significant effect on predicting stress on the first syllable, whereas only vowel duration and F0 showed a significant effect on predicting stress on the second syllable. On the other hand, in the Seoul dialect, vowel duration was found to be the only factor to predict stress on the first syllable, whereas both vowel duration and F0 showed a significant effect in predicting the stress on the second syllable.

<sup>3</sup> The claims regarding word-level prominence in Korean were made based on cases where the phonological process of vowel shortening occurs, whereas Ko (2013) used words with phonemically long and short vowels for the acoustic analysis.

Based on these results, Ko (2013) concluded that Chonnam uses vowel duration, F0, and intensity to express lexical stress, while the Seoul dialect is exhibiting a diachronic change from a stress language to a phrasal-accent language, as supported by the limited expression of word-level prominence. Ko (2013) argued that Seoul Korean had a “duration-based prominence system based on a very limited window of initial syllable” (p. 108, line 10), but the stress has eventually been lost in contemporary Seoul Korean. This raises the question how diachronic change in Seoul speakers’ word-level prominence has affected the production and perception of lexical prominence.

#### **1.4. A few potential problems**

There are a number of issues in the design and interpretation of the Ko (2013) study that warrant a closer look at the notion of stress/word-level prominence in Korean. First, the participants in Ko (2013)’s study were limited to four older Seoul Korean speakers who came to the USA almost 40 years ago. Their exposure to English for this long period could have affected their production of word-level prominence in Korean, and also, their production in L1 might not reflect contemporary Seoul Korean, especially with respect to the ongoing language changes. It has been found that the use of VOT and F0 in indicating the three-way laryngeal distinction among stops in Korean has changed (Kang & Guion, 2008; Lee & Jongman, 2015; Lee, Politzer-Ahles, & Jongman, 2013; Perkins & Lee, 2010; Silva, 2006; Wright, 2007) and more importantly, the vowel length distinction has been claimed to have disappeared among younger Korean speakers (Kim, 2001; Kim & Han, 1998; Magen & Blumstein, 1993). Considering that, while living in the USA, Ko’s speakers did not get as much L1 input as Korean residents, their productions might not be representative of Seoul Korean speakers.

Second, in order to determine the effect of the phrasal boundary on lexical stress, we need to examine productions in two different contexts. Ko (2013) recorded the tokens produced in a carrier sentence, where the target words were embedded in sentence-medial position. However, the carrier sentence that Ko used is unnatural due to the absence of the case marker after the target word. Ko first used a contextually-related sentence in order to prompt the intended word, and then asked the participants to read the target word embedded in a contextually neutral sentence, ‘clearly *apple* pronounce’ [t’o.bak. t’o.bak. **sa.gwa**. par.ɪm.ha.se.jo.]. However, she

deliberately omitted the case marker after the target word, *apple*, since the case marker is an allomorph which will appear as two syllables following an open syllable (e.g., [sa.gwa. ra.go.]) and as three syllables following a closed syllable (e.g., [si.ɕʌŋ. i.ra.go.]). However, the sentence without a case marker sounds extremely unnatural, which might make participants produce the words with unnatural F0 patterns. In fact, Ko explained in her earlier study that the same sentence can be used to express two different prosodic frames, depending on how the sentence is parsed, as illustrated below. (Ko, 2002; p 144). Thus, Ko's stimuli had the potential to attract prosodic focus, introducing another level of prominence.

(1) Two possible ways of phrasing the frame sentence from Ko (2002)

a. Two independent prosodic domains

{t'obak t'obak} {sa:gwa} {parimhasejo}  
 'clearly apple say'

b. A single prosodic domain from the VP

{t'obak t'obak} {sa:gwa parimhasejo}  
 'clearly apple say'

Moreover, Ko instructed her speakers to produce the sentences with a falling intonation, which could also result in unnatural prosody. Ko (2013) explained that this was done in order to prompt the speakers to read the target words in a citation form and also to avoid a list effect. However, Ko did not provide a clear motivation, or references, to clarify how this procedure would achieve natural speech.

Also, Ko only measured raw values for each syllable and compared the difference between the values from the stressed syllables and unstressed syllables in their respective positions. However, the obtained difference might be misleading, because the same difference can also be found from vowel length minimal pairs. Moreover, a direct comparison between the first syllables of the stress minimal pairs does not provide insight into the relative differences between the syllables within a word. In addition, if the speakers claimed that they pronounced the stress pairs as homophones, Ko eliminated those tokens from the analysis. Thus, it is unclear whether the difference found in Ko (2013) is a fair representation of lexical stress, given the fact that the recording procedure was problematic and the data was subjectively selected.

Lastly, Ko (2013) made claims about the use of perceptual cues on the basis of her acoustic findings. Without any direct perception data, it is hard to conclude which cue(s) Korean listeners use in their perception. To our knowledge, no study has been conducted examining cue weighting for Korean word-level prominence. Thus, by conducting an acoustic study as well as a perception study, the present research aims to provide evidence regarding whether Korean indeed has lexical stress or simply has a vowel length distinction. In the acoustic study (Experiment 1), we examine whether we can replicate the findings from Ko in two contexts (i.e., at the sentence level and in words in isolation) with speakers of two generations. In the perception study (Experiment 2), we examine which acoustic cue(s) Korean listeners weight in identifying Korean stress pairs.

Given that it is still unclear whether contemporary Seoul Korean has word-level prominence, the present study aims to investigate whether younger Korean speakers produce word-level prominence. If there is word-level prominence in Korean and younger speakers produce it, will younger speakers also perceive it? If not, do they transfer the use of higher-level prosodic cues to the perception of word-level prominence?

## **2. Experiment 1: Production experiment**

### **2.1. Methods**

#### **2.1.1. Participants**

We recorded 21 male native speakers of Korean (10 older and 11 younger speakers). All subjects were born and raised in Seoul or Suwon, Kyunggi area where the standard Korean dialect is spoken. The mean age was 71.9 years ( $SD = 1.52$ ) for the older speakers and 23.5 years ( $SD = 3$ ) for the younger speakers. None of the subjects lived in any other region where a different dialect is spoken, except for the older Korean speakers during the Korean war from 1951-1953. All subjects were literate in Korean, and none of the subjects reported any speech or hearing disorder.

#### **2.1.2. Stimuli**

Seventeen minimal pairs that were used by Ko (2013) were adopted for the production study. These word pairs are traditionally considered as minimal pairs contrasting in vowel length; Ko (2013) treated them as minimal pairs in terms of stress. For example, for the minimal pair /sa:kwa/ ‘apology’ and /sakwa/ ‘apple’, Ko (2013) treated /sa:kwa/ ‘apology’ as having first-

syllable stress and /sakwa/ ‘apple’ as having second-syllable stress. These word pairs were first embedded in contextually related sentences in order to cue the semantic meaning of the target word to the participants, and then presented in a contextually neutral sentence as well as in isolation. The number of syllables of the contextually related sentences was balanced. Examples of semantically-related sentences and neutral sentences for the word pair /sakwa/ are as follows:

(1) Examples for ‘apology’

- a. Semantically-related sentence for ‘*apology*’  
 [ɕʌl.mo.sil. ha.mjən. **sa:gwa**.ha.go. mʌn.ɕʌ. jon.sa.lil. pin.da.] (16 syllables)  
 ‘If you do wrong, you should give an **apology** first and ask for forgiveness.’
- b. Semantically neutral sentence for ‘*apology*’  
 [i. dan.ʌ.nin. **sa:gwa**. im.ni.da.]  
 ‘This word is **apology**’

(2) Examples for ‘apple’

- c. Semantically-related sentence for ‘*apple*’  
 [ɕʌ.sa. gwa.il.lo. **sa:gwa**.wa. pɛ.ga. ɕʌ.ɕu. sa.jon.dwen.da.]  
 (16 syllables)  
 ‘For fruits to use at ancestor veneration ceremonies, **apples** and pears are often used.’
- d. Semantically neutral sentence for ‘*apple*’  
 [i. dan.ʌ.nin. **sa:gwa**. im.ni.da.]  
 ‘This word is **apple**’

Only the tokens that were produced in neutral sentences (e.g., critical words produced in AP-initial position) and in isolation were examined. All stimuli were presented in Korean orthography in a randomized order, without any indication of the vowel length or stress location. In total, 714 tokens were recorded in a contextually neutral sentence (17 pairs x 2 repetitions x 21 speakers), and 357 tokens in word isolation (17 pairs x 21 speakers).

### **2.1.3. Procedure**

The recordings were conducted in Suwon and Seoul. For the older Korean speakers, the recording was made in a quiet room in a local hotel, and for the younger Korean speakers, the recording was made in a seminar room at Sungkyunkwan University in Seoul, Korea. A Marantz Digital Recorder (PMD 671) and a Shure head-mounted microphone were used for the recording of both groups. The subjects were asked to read the stimulus sentences where the target words were embedded in different carrier sentences. First, the subjects read the target words embedded in contextually related sentences. Immediately after that, the subjects read the same target words embedded in the contextually neutral sentences with two repetitions. Then, the speakers read the same target word in isolation with one repetition. The sampling rate of the recording was 22,050 Hz and these recordings were analyzed using the speech analysis program Praat (version 5.4.03) (Boersma & Weenink, 2018).

### **2.1.4. Measurements**

Duration, intensity, and F0 values were measured for each vowel from the first and second syllable of the target words. Duration was measured from the onset of F1 to the offset of F2 of each syllable.<sup>4</sup> Tokens that exhibited devoicing of high vowels between two voiceless consonants were eliminated from the analysis. This applied to both members of the minimal pairs. A total of 102 tokens (47 from the older speakers' productions) were eliminated from the productions recorded at the sentence level, and 58 tokens (28 from older speakers' productions) were eliminated from the productions recorded in isolation. The intensity values were averaged over each vowel. For F0, the F0 values from 20% to 80% of the duration of each vowel were averaged to avoid perturbation effects from the preceding consonant.<sup>5</sup> Thus, a total of 3,060 measurements (612 tokens x 5 measurements) were taken from the tokens produced in contextually neutral sentences, and 1,495 measurements (299 tokens x 5 measurements) were taken from the tokens produced in isolation.

To control for differences across speakers in terms of duration, F0, and amplitude of the target syllable, second-to-first syllable ratios for these measurements were used, following Beckman (1986)'s formulas:

---

<sup>4</sup> We used duration ratio in order to control for any variations in speech rate.

<sup>5</sup> We used average F0 values to be able to directly compare our measurements to those of previous studies.



F0 ratio (in semitones) =  $17.31 \ln[\text{Hz}(S2)/\text{Hz}(S1)]$

Average intensity ratio = dB (S2) - dB (S1)

Log duration ratio =  $\ln[\text{ms}(S2)/\text{ms}(S1)]$ .

Thus, it is expected that first-syllable stressed words (e.g., [sa:gwa] ‘apology’) have a negative value of each ratio, and second-syllable stressed words (e.g., [sagwa] ‘apple’) will result in a positive value.

### 2.1.5. Data analysis

Factorial repeated measures ANOVAs were conducted with the second-to-first-syllable ratio values for intensity, duration, and F0 as dependent variables, and Stress (first syllable vs. second syllable), Age Group (older vs. younger), and Phrasal Condition (sentence vs. isolation) as independent variables.

## 2.2. Results

### 2.2.1. Duration

When examining second-to-first syllable duration ratios, the results showed main effects of Stress [ $F(1, 19)=11.92, p=.003$ ], Age Group [ $F(1, 19)=4.51, p=.005$ ], and Phrasal Condition [ $F(1, 19)=10.60, p=.004$ ]. These results indicate that first-syllable stressed words had smaller duration ratio values (0.21) than second-syllable stressed words (0.34), and older Korean speakers produced stress pairs with smaller ratio values (0.19) than the younger Korean speakers (0.37). Also, speakers produced the stress pairs with smaller duration ratio values at the sentence level (0.19) than in isolation (0.36).

We also found a marginally significant three-way interaction among Stress, Age Group, and Phrasal Condition [ $F(1, 19)=3.78, p=.067$ ], indicating that the duration ratio by stress between the two speaker groups was marginally affected by Phrasal Condition. The duration ratio difference between the first- and second-syllable stressed words as a function of Phrasal Condition was greater for the older speakers (sentence: 0.24, isolation: 0.1), than for the younger speakers (sentence: 0.07, isolation: 0.08).

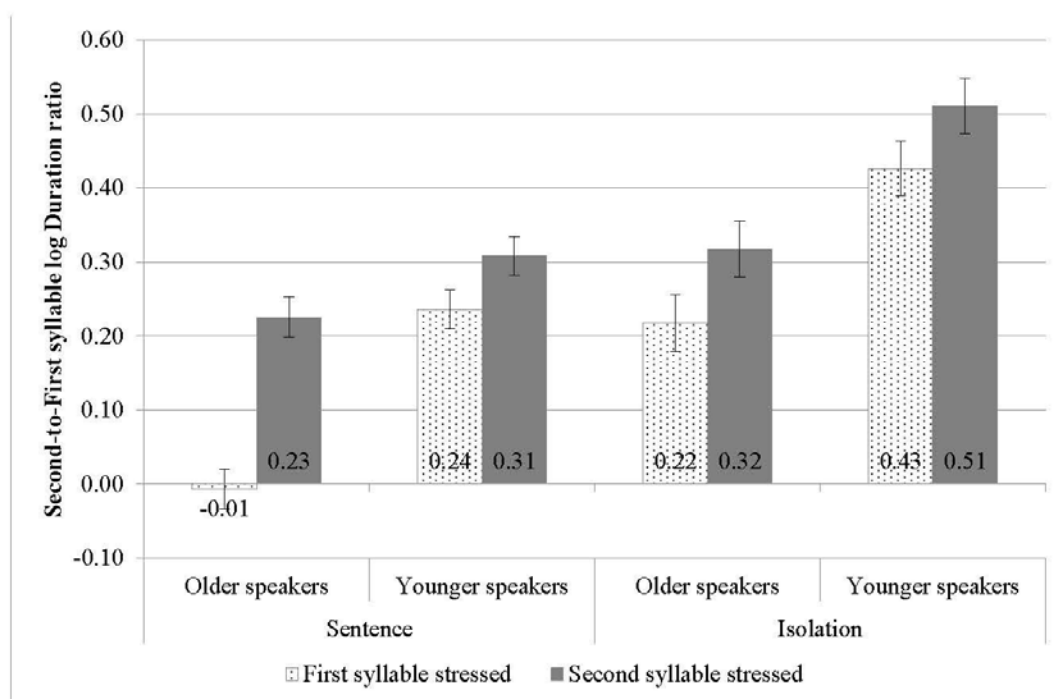


Figure 1. Second-to-first syllable log duration ratio values of the first- and second-syllable stressed words in two different contexts between two speaker groups.

Figure 1 illustrates the duration ratio values of the first- and second-syllable stressed words in two different contexts between the two speaker groups.

### 2.2.2. Intensity

When examining second-to-first syllable intensity ratios, the results showed main effects of Stress [ $F(1, 19)=10.64, p<.001$ ] and Phrasal Condition [ $F(1, 19)=46.50, p<.001$ ]. These results indicate that the first-syllable stressed words had a smaller intensity ratio (-0.19 dB) than the second-syllable stressed words (0.37 dB), and the intensity ratio values were greater for the productions from the sentence level (2.10 dB) than those from isolation (-1.92 dB). We also found a two-way interaction between Stress and Age Group [ $F(1, 19)=4.85, p<.001$ ], indicating that the intensity ratio difference between the stress pairs was greater for the older speakers than the younger speakers. Figure 2 illustrates the second-to-first syllable intensity ratio values of the first- and second-syllable stressed words in two different contexts between the two speaker groups.

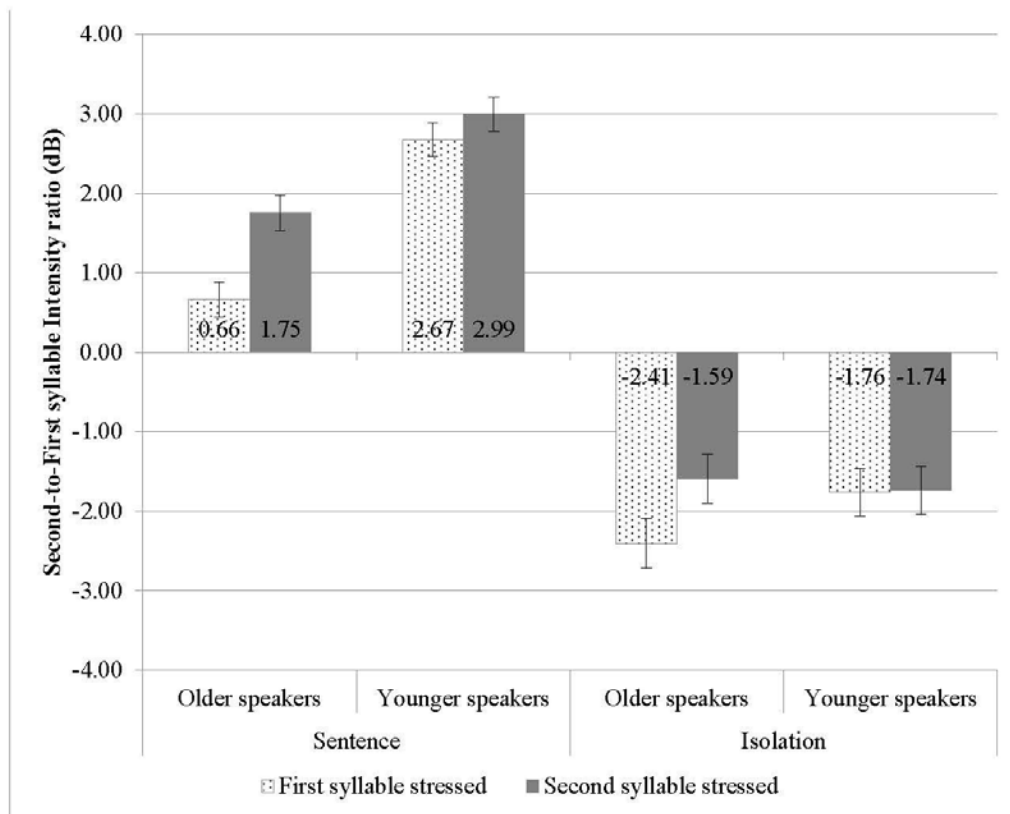


Figure 2. Second-to-first syllable intensity ratio values of the first- and second-syllable stressed words in two different contexts between two speaker groups.

### 2.2.3. F0

When examining second-to-first syllable F0 ratios, only a main effect of Phrasal Condition [ $F(1, 19)=81.50, p<.001$ ] was found, indicating that the F0 ratio was greater in the productions at the sentence level (0.98) than in the productions in isolation (-2.21). We also found a significant interaction between Phrasal Condition and Age Group [ $F(1, 19)=6.23, p<.001$ ], indicating that the F0 ratio difference between the two contexts was greater for the younger speakers (1.85) than the older speakers (1.62). Figure 3 illustrates the second-to-first syllable F0 ratio values of the first- and second-syllable stressed words in two different contexts between the two speaker groups.

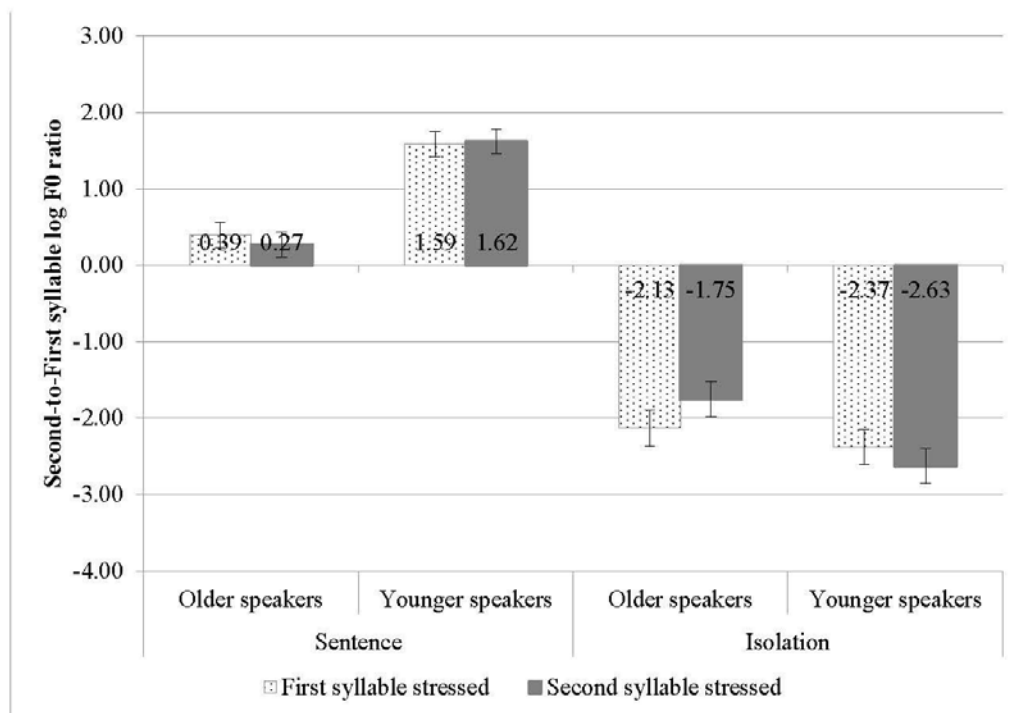


Figure 3. Second-to-first syllable F0 ratio values of the first- and second-syllable stressed words in two different contexts between two speaker groups.

### 2.3. Discussion and conclusion

A comparison of the ratio values in the two different contexts revealed several interesting facts. First, we found that both older and younger Korean speakers use durational difference in their production of Korean stress, as supported by the main effect of Stress on the second-to-first syllable duration ratio. We also observed a generational difference in the way the two groups of speakers used duration. Younger speakers always produced a second syllable that was longer than the first syllable, regardless of the stress position. In contrast, older speakers produced a longer first syllable at the sentence level, as shown by the negative values of the second-to-first syllable duration ratio. We also found that different phrasal levels affect the durational difference, as shown by the main effect of Phrasal Condition, suggesting that the final lengthening effect on the second syllable in isolation partially neutralized the effect of the vowel length distinction in the first syllable. This final-lengthening effect on stress seems to influence older speakers more than younger speakers. This is not surprising, considering

that previous studies have claimed that younger speakers have been losing the durational differences in contemporary Seoul Korean (Kim, 2001; Kim & Han, 1998; Magen & Blumstein, 1993).

Second, intensity is used as a cue to stress only by older speakers, as supported by a two-way interaction between Stress and Age Group. The ratio analysis for each condition revealed that intensity ratio only cues stress for older speakers at the sentence level. However, how intensity expresses prominence within a word varied as a function of context: intensity ratio values in isolation consistently showed negative values, while intensity ratio values at the sentence level consistently had positive values. This indicates that at the sentence level, the second syllable had higher intensity values than the first syllable, whereas in isolation, the first syllable had higher intensity values than the second syllable. The different intensity pattern between the two syllables across different contexts suggests that intensity is strongly affected by the phrasal-boundary effect, resulting in higher intensity at the boundary-initial position in isolation than at the boundary-initial position in sentence-medial position.

With respect to F0, no effect of stress was found for either speaker group, indicating that F0 is not a reliable parameter to indicate Korean stress. The two-way interaction between Phrasal Condition and Age Group indicates that F0 clearly marks the phrasal domain, especially in younger Korean speakers' productions.

The question then arises whether duration was ever used as a primary cue to lexical stress. Although we predicted that the duration effect would be weakened in isolation due to final lengthening, if Korean indeed had lexical stress, multiple cues in the first syllable should still indicate lexical stress. While both duration and intensity were significant cues to stress, inconsistent ratio patterns were observed as a function of Phrasal Condition. If Korean had lexical stress expressed with duration and intensity, as found by Ko (2013), both cues should pattern similarly in the two contexts. However, such a pattern was not found. Note that while the duration ratio values showed positive values in both contexts except for older speakers' productions at the sentence level, intensity ratios were either positive (at the sentence) or negative (in isolation) for both speaker groups. Therefore, the absence of consistent results across two contexts between the two speaker groups seems to suggest that duration could not be a cue to stress, and the effect of intensity was due to the effect of phrasal-level prominence.

However, there is still a possibility that lexical stress has diminished only in production but still exists in perception. Thus, the next issue to investigate is whether Korean speakers will only be sensitive to duration in differentiating stress in perception, or whether they will use other cues as well. For example, will older Korean speakers use intensity in addition to duration in distinguishing Korean stress, as we found in their production? Will younger listeners be only sensitive to duration or will they also be sensitive to other cues like intensity and F0? Considering that the second-to-first syllable duration ratio that we found in Experiment 1 was rather small, albeit significant, younger listeners may not rely on duration at all in perception. These issues will be addressed in the following perception study.

### **3. Experiment 2: Perception**

Stress is realized by multiple cues (e.g., Ladefoged, Draper, & Whitteridge, 1958; Lehiste & Peterson, 1959; Lieberman, 1960) and the intrinsic characteristics of stressed syllables – longer duration, higher F0, and greater intensity – are strongly correlated with the perception of stress. When one of these parameters is not in the predicted direction, there is always a trade-off effect with other cues (e.g., Lieberman, 1960). That is, when one acoustic parameter (e.g., F0) does not contribute to the perception of stress, other cues (e.g., amplitude) may compensate and take on a greater role. This pattern was found not only in free-stress languages, such as English and Dutch, but also in fixed-stress languages such as Spanish (Llisterri, Machuca, de la Mota, Riera, & Ríos, 2003) and Arabic (de Jong & Zawaydeh, 1999), suggesting that stress is conveyed by multiple cues, and different languages use these acoustic cues with different degrees of saliency in indicating stress.

Based on this, we will use a perception study to re-examine whether Korean has a truly lexical stress that is mainly realized with duration or whether Korean only has a phonemic vowel length distinction. If Korean has both lexical stress and phonemic vowel length, Korean listeners should be sensitive to duration in the first syllable and weight intensity in the second syllable in distinguishing Korean word-level prominence. If Korean has only phonemic vowel length, but not lexical stress, then Korean listeners may only be sensitive to duration in the first syllable.

In addition, we investigate if older and younger Korean listeners differ in their use of acoustic cues in processing stress contrasts in Korean.

Even if younger Korean speakers have lost the vowel length contrast (or lexical stress) in their production, they might still have a perceptual distinction, since they are exposed to the duration distinction in the speech of their elders. Moreover, we also aim to investigate whether in their perception Korean listeners use the same cue(s) they used in their production to distinguish the stress pairs.

### 3.1. Methods

#### 3.1.1. Participants

The same 10 older Korean speakers and 12 younger Korean speakers who participated in Experiment 1 took part in the perception study on Korean stress pairs.

#### 3.1.2. Stimuli

##### 3.1.2.1. Original base token

The Korean minimal stress pair ‘sakwa’ was chosen as the stimulus because both members of the pair had a similar frequency of occurrence (frequency of first-syllable stressed /sa:kwa/ ‘apology’: 48 per 3 million words; frequency of second-syllable stressed /sakwa/ ‘apple’: 63 per 3 million) as provided by the National Institute of Korean Corpus (2002). These tokens were produced by an older Korean male speaker (Speaker A, age 68) who did not participate in the production study. We selected the production of the first-syllable stressed word, /sa:kwa/ ‘apology’, as the baseline token in order to preserve possible acoustic information in the long vowel and also to minimize any possibility of losing acoustic information by lengthening the short vowel to a long vowel. The manipulation range was based on the minimum and maximum value of three acoustic parameters – duration, F0, and intensity – of both younger and older speakers’ productions of the /sakwa/ pair from Experiment 1.

##### 3.1.2.2. Stimulus manipulation

All stimuli were manipulated from the single token /sa:kwa/, so that we could control any unintended changes in phonation type or vowel quality. The stimuli were first produced in semantically-related sentences (e.g., ‘If you do wrong, you should give an **apology** first and ask for forgiveness.’), and then produced in a semantically neutral carrier sentence (i.e., [i. dan.Λ.nin. **sa:.gwa.** im.ni.da.] ‘This word is apology’). The token that was produced in the neutral sentence was used as the baseline

token. For the stimulus manipulation, the maximum and minimum values of F0, intensity, and duration for the two sets of stimuli across all the older and younger speakers were used as endpoints. For each condition, two parameters (e.g., duration x F0) were orthogonally manipulated to signal the stress pattern while the other cue (e.g., intensity) was controlled to be ambiguous (step 3 in Table 1). Each cue had 5 steps from unstressed to stressed syllable based on the acoustic data collected in Experiment 1. We also made sure the step size was greater than the Just Noticeable Difference (JND) (Flanagan, 1955; Flanagan & Saslow, 1958; Klatt, 1973; Fujisaki, Nakamura, & Imoto, 1975; Klatt & Cooper, 1975; Nishinuma, Di Cristo, & Espesser, 1983; Turk & Sawusch, 1996) for each cue, so that the listeners could perceive the difference between successive steps. When the first syllable was manipulated, the second syllable was controlled to be ambiguous (at step 3) between a stressed and an unstressed syllable, and the first syllable was controlled to be ambiguous (at step 3) when the second syllable was manipulated. These manipulated tokens were then embedded in a semantically neutral carrier sentence (e.g., This word is \_\_\_\_ . [i. dan.ʌ.nin. \_\_\_\_\_ im.ni.da.]) produced by Speaker A, and presented as the auditory stimuli in the perception experiment.

Duration, F0, and intensity were all manipulated using Praat (Boersma & Weenink, 2018). Table 1 represents the manipulation values of the five steps for the first and second syllables.

	First syllable					Second syllable				
	Unstressed		Stressed			Unstressed		Stressed		
Step	1	2	3	4	5	1	2	3	4	5
Duration (ms)	56	85	114	143	172	87	89.25	91.5	93.75	96
F0 (Hz)	98.95	103.32	107.69	112.06	116.43	112.63	117.02	121.41	128.80	130.19
Intensity (dB)	60.00	62.75	65.5	68.25	71.00	60.00	61.25	62.50	63.75	65.00

Table 1. Five steps of manipulation values of first and second syllable for duration, F0, and intensity.



Thus, 25 tokens were created for each manipulation condition (e.g., 5 steps of duration x 5 steps of intensity; 5 steps of intensity by 5 steps of F0; 5 steps of F0 by 5 steps of duration) for each syllable. Three pairs of cues were manipulated: F0 & intensity; F0 & duration; duration & intensity. The 5 steps of both cues in each pair were crossed to form 25 stimuli, while the third cue was controlled to be neutral at step 3. For example, we created 25 stimuli by manipulating F0 and duration, while keeping intensity at step 3. In all, we defined 75 stimuli in this way. However, this procedure resulted in repeating conditions with two cues at level 3 twice, and three cues at level 3 three times for a total of 14 repetitions. Therefore, 61 unique stimuli were created with this procedure at each syllable level for a total of 122 stimuli. These were each repeated 3 times for a total of 366 tokens for each subject.

### **3.1.3. Procedure**

Before the stress perception test, we examined the older Korean listeners' hearing sensitivity using a pure tone threshold test. Using an up-5dB, down-10dB procedure, 6 octaves were tested in total: 250, 500, 1000, 2000, 4000, and 8000 Hz. Normal-hearing thresholds were defined as thresholds which are better than 20 dB. All older listeners passed the threshold of the hearing acuity test.

Next, a word identification task was employed to examine which suprasegmental cue(s) Korean listeners are sensitive to in perceiving stress contrasts, and whether there is a generational difference between older and younger Korean listeners. First, participants were presented with two pictures of objects denoting the target words on a computer screen, associated with either number key [1] (first-syllable stressed word, /sa:kwa/) or [0] (second-syllable stressed word, /sakwa/) on the keyboard. Next, they were asked to identify the auditorily presented word by clicking either the [1] or [0] key. The position of the pictures and the numbers associated with them were counterbalanced across participants.

The experiment was conducted with three different blocks of 122 trials in randomized order. The intertrial interval (ITI) was 1500 ms. A practice session with 12 trials was conducted before the main experiment to ensure that the participants were familiar with the task. The subjects were allowed to take a short break between blocks.

### **3.1.4. Data analysis**

We conducted a binomial logistic regression to examine the effect of three acoustic parameters (duration, intensity, and F0) on the perception of the first and second syllable between the two listener groups, using the lme4 package (Bates, 2005; Bates & Maechler, 2010) in the R statistical environment (R development Core Team, 2012, Version 3.1.2). The model had Response (/sa:kwa/ ‘apology’ vs. /sakwa/ ‘apple’) as a dependent variable, and Age Group (younger vs. older), Syllable (first vs. second), Intensity manipulation (steps 1-5), Duration manipulation (steps 1-5), and F0 manipulation (steps 1-5) as fixed effects and Participants as random effect. The model tested main effects of the independent variables, two-way interactions between the cues (Duration by Intensity, Intensity by F0, F0 by Duration), two-way interactions between Age Group and Syllable, and three-way interactions among Age Group and two of the cues (e.g., Age Group by Duration by Intensity). When there was a significant interaction between the independent variables, we stratified the data by Syllable and Age Group to probe the interaction between the variables. The older speaker group was used as the baseline to determine whether younger Korean listeners are less sensitive to Korean stress. Thus, the baseline in the model was the older group’s performance on words with second syllable prominence (e.g., /sakwa/ ‘apple’) with step 1 values of intensity, duration, and F0.

### **3.2. Results**

A linear mixed-effects model fitted on all participants’ responses in identifying the Korean stress pairs. A series of fitted mixed-effects regression models were tested in a stepwise analysis to find the most parsimonious model. Table 2 presents the result of the logistic regression on both syllables. Only significant results are reported here. The results revealed significant main effects of Age Group ( $p=.05$ ), Syllable ( $p<.01$ ), and Duration ( $p<.01$ ). These results indicate that older listeners’ responses were more biased toward the first-syllable stressed word /sa:kwa/ (68 %) than those of younger listeners (57 %), and the listeners’ response was biased toward the first-syllable stressed word (74%) when the second syllable was manipulated, as compared to the tokens for which the first syllable was manipulated (50%). Also, the probability that listeners perceived the second syllable as stressed was 53 % when the duration step was at 1, and decreased to 45%, 37%, 30%, and 27% when the duration step was at 2, 3,

4, and 5, respectively. We also found significant interactions between Age Group and Duration ( $p=.05$ ), and Syllable and Duration ( $p<.01$ ). In order to better understand these interactions, we stratified the data by Syllable, and then ran two separate models at each syllable level.

Variable	Estimate (SE)	Z	p
(intercept)	7.02 (1.82)	3.85	<.01
Age Group_young	-4.54 (2.29)	-1.98	=.05
Syllable	-4.03 (1.15)	-3.50	<.01
Duration	-2.25 (0.51)	-4.43	<.01
Age Group_young:Duration	1.24 (0.63)	1.98	=.05
Syllable:Duration	1.18 (0.32)	3.73	<.01

Table 2. Summary of results of the optimal model from the logistic regression examining responses at both syllable levels.

In the *post-hoc* analysis, we ran separate linear mixed-effects models examining all participants' responses to the tokens for which the first syllable was manipulated. The main effects of Age Group ( $p<.01$ ) and Duration ( $p<.01$ ) indicate that older listeners gave more first-syllable stressed responses (54 %) than younger listeners (47 %), and as the duration of the first syllable increased, listeners' responses shifted from words with second-syllable prominence to first-syllable prominence. The second-syllable stressed response rate was 78 % when the duration step was at 1, and decreased to 65%, 48%, 34%, and 27% when the duration step was at 2, 3, 4, and 5, respectively. Table 3 presents a summary of results of the model at the level of the first syllable.

Variable	Estimate (SE)	Z	p
(intercept)	2.80 (0.28)	9.85	<.01
Age Group_young	-1.40 (0.37)	-3.76	<.01
Duration	-1.00 (0.57)	-17.83	<.01
Age Group_young:Duration	0.59 (0.07)	8.50	<.01

Table 3. Summary of results of the logistic regression examining responses at the first syllable level

Figure 4 presents the probability of second-syllable stressed responses between the two groups for the tokens for which three acoustic cues of the first syllable were manipulated, showing the different use of the duration cue (red lines) between the two listener groups (old: dashed lines; younger: solid lines) in perceiving Korean stress pairs.

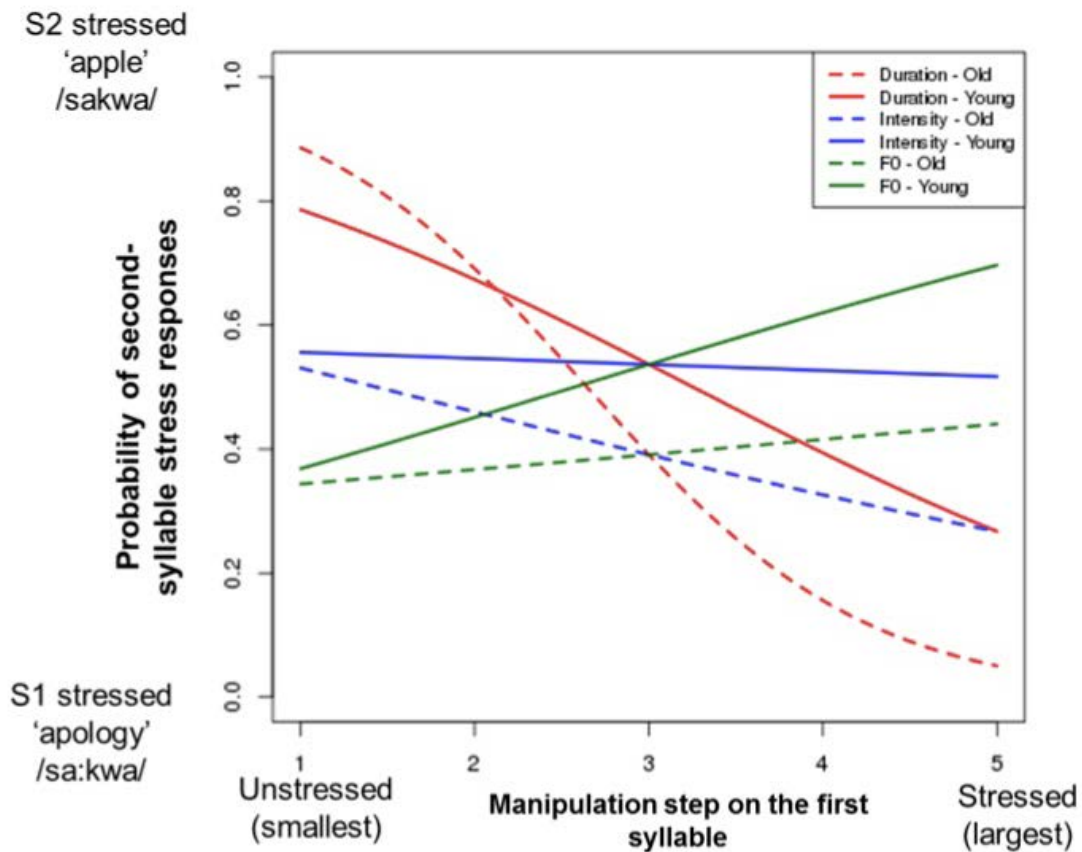


Figure 4. Probability of second-syllable stressed responses, /sakwa/, between the two listener groups. X-axis indicates the manipulated steps of each cue. 1 indicates that duration values were at the minimum endpoint, expressing first-syllable unstressed; and 5 indicates that duration values were at the maximum endpoint, expressing first-syllable stressed. Dotted lines indicate older Korean listeners' responses, and solid lines indicate younger Korean listeners' responses. Listeners' responses for each cue are illustrated with different colors: red, blue, green lines indicate listeners' responses for duration, intensity, and F0, respectively.

In order to explore the interactions between Age Group and Duration, we further conducted a separate linear mixed-effects model examining participants' responses to the tokens for which the first syllable was manipulated as a function of the listener groups (see Table 4).

Variable	Estimate ( <i>SE</i> )	<i>Z</i>	<i>p</i>
(intercept)	2.81 (0.31)	8.94	< .01
Duration	-1.01 (0.06)	-17.81	< .01

Table 4. Summary of results of the logistic regression examining responses of older listeners at the level of the first syllable.

For the older listeners, we found a main effect of Duration ( $p < .01$ ), indicating that older listeners' second-syllable stressed responses increased as duration decreased. The second-syllable stressed response rate was 52 % when the duration step was at 1, and decreased to 41 %, 31 %, 22 %, and 17 % when the duration step was at 2, 3, 4, and 5, respectively. For younger listeners, we also found a main effect of Duration ( $p < .01$ ), indicating that the probability of the second-syllable stressed responses increased as the duration decreased (see Table 5). The second-syllable stressed response probability was 54 % when the duration step was at 1, and decreased to 48 %, 42 %, 36 %, and 35 % when the duration step was at 2, 3, 4, and 5, respectively.

Variable	Estimate ( <i>SE</i> )	<i>Z</i>	<i>p</i>
(intercept)	1.39 (0.22)	6.46	<.01
Duration	-0.42 (0.04)	-10.72	<.01

Table 5. Summary of results of the logistic regression examining responses of younger listeners at the level of the first syllable

A separate linear mixed-effects model examining all participants' responses to the tokens for which the second syllable was manipulated, found no main effects or interactions, indicating that none of the listener groups were using the duration cue in perceiving Korean stress pairs. Figure 5 represents older and younger listeners' responses as a function of manipulated steps of three cues, showing the lack of effect of acoustic cues on the perception of the Korean stress pairs on the second syllable.

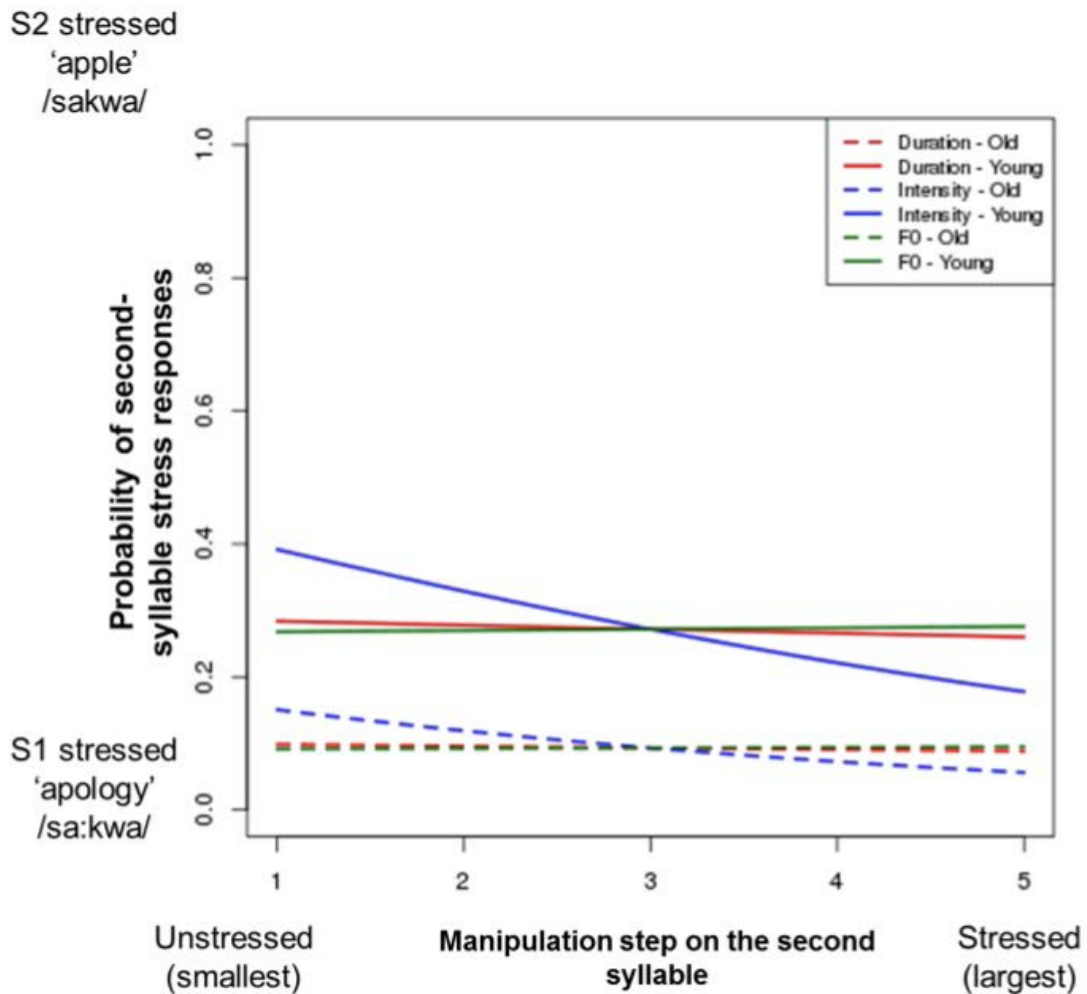


Figure 5. Probability of second-syllable stressed responses, /sakwa/, between the two listener groups. X-axis indicates the manipulated steps of each cue. 1 indicates that duration values were at the minimum endpoint, expressing first-syllable unstressed; and 5 indicates that duration values were at the maximum endpoint, expressing first-syllable stressed. Dotted lines indicate older Korean listeners' responses, and solid lines indicate younger Korean listeners' responses. Listeners' responses for each cue are illustrated with different colors: red, blue, green lines indicate listeners' responses for duration, intensity, and F0, respectively.

### 3.3. Discussion and Conclusion

Experiment 2 revealed several interesting facts. First, we found that neither older nor younger Korean listeners use intensity and F0 independently from duration in identifying prominence in Korean. Both older and younger Korean listeners used only the duration cue in identifying Korean stress pairs. In addition, we also found that younger Korean listeners still weight duration in their identification of phonemic vowel length, despite

the fact that contemporary Seoul Korean has almost completely lost the vowel length distinction. According to exemplar theory (Goldinger, 1996; Johnson, 1997), listeners mentally store variant details of speech sounds in their episodic memory, while mapping similar tokens into a single abstract category as a large cloud of exemplars. Highly similar tokens are tightly clustered and organized within a category, while dissimilar tokens are far apart and mapped onto two different categories. However, an exemplar-based model predicts that the production of categories may be deviant from perception, since the lexical entries that each listener stores are gathered from different speakers (Bybee, 2001; Johnson, 1997; Pierrehumbert, 2000, 2001). In the present case, the younger Korean speakers may have stored both long and short vowels as poor exemplars of a single category, reflecting the loss of the vowel length distinction in their production. However, the variations in the categorized percept (e.g., older speakers' contrastive production of the vowel length distinction) may have influenced younger listeners' perceptual sensitivity in identifying the vowel length contrasts. In other words, although younger listeners do not have two distinct vowel length categories, since younger listeners have collected both long vowels as in some of exemplars of /sa:kwa/ 'apology' and short vowels for /sakwa/ 'apple', the younger listeners may recognize this pair based on the duration of the vowel. This view, then, is compatible with our finding that younger Korean listeners still are sensitive to phonemic vowel length although the vowel length distinction has all but disappeared in their production.

Another interesting finding is that only cue manipulation in the first syllable influenced listeners' responses; a perceptual shift from the first-syllable stressed word to the second-syllable stressed word was not found when the second syllable was manipulated in terms of duration, intensity, or F<sub>0</sub>. Given that stress is defined as the relative difference in prominence between two syllables (Pierrehumbert, 1979; Beckman & Pierrehumbert, 1986), the fact that the cue manipulation in the second syllable did not trigger a change in the perception of the stress location indicates that Korean listeners only put perceptual weight on the first syllable. If Korean had stress, the changes in prominence in the second syllable should also induce the perceptual shift; however, the current study did not find such a pattern. Therefore, taking into account these two pieces of evidence provided by the current perception study, in addition to the acoustic evidence from Experiment 1, we conclude that Korean does not employ lexical stress, and that what has been claimed as stress pairs are actually vowel length contrasts.



Overall, the findings of the current study revealed that Korean does not employ word-level prominence, but that (so-called) Korean lexical stress pairs are only differentiated in terms of vowel length in the initial syllable.

## References

- Bates, D. (2005). Fitting linear mixed models in R. *R News*, 5, 27-30.
- Bates, D., & Maechler, M. (2010). Matrix: sparse and dense matrix classes and methods. *R package version 0.999375-43*, URL <http://cran.r-project.org/package=Matrix>.
- Beckman, M. E. (1986). *Stress and non-stress accent*. Vol. 7. Dordrecht: Foris..
- Beckman, M. E. & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3(1), 255–309.
- Boersma, P. & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.41, retrieved 6 August 2018 from <http://www.praat.org/>
- Bybee, J. (2001). *Phonology and Language Use. Language*. Cambridge, UK: Cambridge University Press.
- Cho, T. & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, 29(2), 155-190.
- Cho, T., McQueen, J. M. & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35(2), 210-243.
- Christophe, A., Peperkamp, S., Pallier, C., Block, E. & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language*, 51(4), 523-547.
- Chung, K., Chang, S., Choi, J., Nam, S., Lee, M., Chung, S., & Jee, S. (1996). *A study of Korean prosody and discourse for the development of speech synthesis/recognition system*. Daejeon, Korea: KAIST Artificial Intelligence Research Center.
- Cooper, W. E., Eady, S. J. & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *The Journal of the Acoustical Society of America*, 77(6), 2142-2156.
- Coughlin, C. E. & Tremblay, A. (2012). Non-native listeners' delayed use of prosodic cues in speech segmentation. In Cox, F., Demuth, K., Lin, S., Miles, K., Palethorpe, S., Shaw, J. & Yuen, I. (Eds.), *Proceedings of the 14th Australasian Conference on Speech Science and Technology* (pp. 189-192). Macquarie University, Sydney, Australia.
- de Jong, K., & Zawaydeh, B. A. (1999). Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics*, 27(1), 3-22.

- Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 26(1), 138.
- Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1(126), 126-152. Retrieved from <http://las.sagepub.com/content/1/2/126.short>
- Gay, T. (1978). Physiological and acoustic correlates of perceived stress. *Language and Speech*, 21(4), 347-353.
- Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1166-1183.
- Heo, W. (1965). *Kwuke Umwoonhak* [Korean phonology]. Seoul: cengumsa.
- International Phonetic Association. 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge, U.K.: Cambridge University Press.
- Johnson, K. (1997). Speech perception without speaker normalization: an exemplar model. In Johnson, K. & Mullennix, J. W. (Eds.), *Talker variability in speech processing* (pp. 143-166). New York, NY: Academic Press.
- Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody*. Ph.D. dissertation, The Ohio State University.
- Jun, S.-A. (1995). A phonetic study of stress in Korean. *The Journal of the Acoustical Society of America*, 98(5), 2893.
- Jun, S.-A. (1996). *The phonetics and phonology of Korean prosody: intonational phonology and prosodic structure*. New York: Garland Press.
- Jun, S.-A. (1998). The Accentual Phrase in the Korean prosodic hierarchy. *Phonology*, 15(2), 189-226.
- Jun, S.-A. (2000). K-ToBI (Korean ToBI) Labelling Conventions - Version 3.1. *The Korean Journal of Speech Science*, 7(1), 143-169.
- Kang, K.-H. & Guion, S. G. (2008). Clear speech production of Korean stops: changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America*, 124(6), 3909-3917.
- Kenstowicz, M. & Park, C. (2006). Laryngeal features and tone in Kyungsang Korean: A phonetic study. *Studies in Phonetics, Phonology and Morphology*, 12(2), 247-264.
- Kim-Renaud, Y.-K. (1974). *Korean consonantal phonology*. Ph.D. dissertation, University of Hawaii.
- Kim, H.-S., & Han, J.-I. (1998). Vowel length in Modern Korean: An acoustic analysis. In Park, B.-S. & Yoon, J. H. (Eds.), *Proceedings of the 14th International Conference on Korean Linguistics* (pp. 412-418). Seoul: Hanshin.
- Kim, S.-H. (2001). *Hyenday kwuke-euy eumcang* [The Vowel Length of Modern Korean]. Seoul: Yeok-Lak.
- Kim, S. (2004). *The role of prosodic phrasing in Korean word segmentation*. Ph.D. dissertation, University of California Los Angeles.

- Kim, S. & Cho, T. (2009). The use of phrase-level prosodic information in lexical segmentation: evidence from word-spotting experiments in Korean. *The Journal of the Acoustical Society of America*, 125(5), 3373-3386.
- Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, 3(3), 129-140.
- Ko, E.-S. (2002). *The phonology and phonetics of word-level prosody and its interaction with phrase-level prosody: A study of Korean in comparison to English*. Ph.D. dissertation, University of Pennsylvania.
- Ko E.-S. (2010). Stress and Long Vowels in Korean: Chicken or Egg First? *Japanese Korean Linguistics*, 17, 377-390.
- Ko, E.-S. (2013). A metrical theory of Korean word prosody. *Linguistic Review*, 30(1), 79-115.
- Koopmans-Van Beinum, F. J. (1980). *Vowel contrast reduction: an acoustic and perceptual study of Dutch vowels in various speech conditions*. Ph.D. dissertation, Universiteit van Amsterdam.
- Ladefoged, P., Draper, M. H. & Whitteridge, D. (1958). Syllables and stress. *Miscellanea. Phonetica*, 3, 1-14.
- Lee, B.-G. (1978). *Kukeuy cangmoumhwa-wa posangseng* [Vowel lengthening and compensation in Korean]. Seoul: Kukehak 3. (Reprinted from Umwun Hyensange Isseseye Ceyak [Constraints in phonological phenomena], pp. 23-57, 1979, Seoul: Hanshin).
- Lee, H. & Jongman, A. (2015). Acoustic evidence for diachronic sound change in Korean prosody: A comparative study of the Seoul and South Kyungsang dialects. *Journal of Phonetics*, 50, 15-33.
- Lee, H., Politzer-Ahles, S. & Jongman, A. (2013). Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of Phonetics*, 41(2), 117-132.
- Lehiste, I., & Peterson, G. E. (1959). Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America*, 31(4), 428-435.
- Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, 32(4), 451-454.
- Llisterri, J., Machuca, M., de la Mota, C., Riera, M. & Río, A. (2003). The perception of lexical stress in Spanish. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2023-2026).
- Magen, H. S. & Blumstein, S. E. (1993). Effects of speaking rate on the vowel length distinction in Korean. *Journal of Phonetics*, 21, 387-410.
- Oh, M. (1998). The prosodic analysis of intervocalic tense consonant lengthening in Korean. *Japanese Korean Linguistics*, 8, 317-330.
- Perkins, J. & Lee, S. J. (2010). Korean affricates and consonant - tone interaction. In *Proceedings of the 6th workshop on Altaic formal linguistics*. MIT Working Paper of Linguistics (pp. 277-286).

- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Pierrehumbert, J. B. (2000). What people know about sounds of language. *Studies in Linguistic Sciences*, 29(2), 111-120.
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J. & Hopper, P. (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137-158). Amsterdam: John Benjamins.
- Selkirk, E. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, Massachusetts: MIT Press.
- Shport, I. A., & Redford, M. A. (2014). Lexical and phrasal prominence patterns in school-aged children's speech. *Journal of Child Language*, 41(04), 890-912.
- Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, 23(2), 287-308.
- Tremblay, A., Coughlin, C. E., Bahler, C. & Gaillard, S. (2012). Differential contribution of prosodic cues in the native and non-native segmentation of French speech. *Journal of Laboratory Phonology*, 3(2), 385-423.
- Tyler, M. D. & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America*, 126(1), 367-376.
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, 91(3), 1707-1717.
- Wright, J. D. (2007). *The phonetic contrast of Korean obstruents*. Ph.D, dissertation, University of Pennsylvania.



## Comparing Monosyllabic and Disyllabic Training in Perceptual Learning of Mandarin Tone

Yingjie Li  
University of Kansas

Goun Lee  
Sungkyunkwan University

Joan A. Sereno  
University of Kansas

### Abstract

Although computer-assisted auditory perceptual training has been shown to be effective in learning Mandarin Chinese tones in monosyllabic words, tone learning has not been systematically investigated in disyllabic words. In the current study, seventeen native English-speaking beginning learners of Chinese were trained using a high variability phonetic training paradigm. Two perceptual training groups, a monosyllabic training group and a disyllabic training group, were compared and accuracy in identifying the tonal contrasts in naturally produced monosyllabic and disyllabic words (produced by native Mandarin Chinese speakers) was evaluated. Results showed that after only four training sessions in a two-week period, beginning learners of Chinese significantly increased their tonal identification accuracy from the pretest (72%) to posttest (80%). The current findings overall show significant differences between the monosyllabic perceptual training group and disyllabic perceptual training group. Although native English-speaking learners in both training groups made improvements in their tonal identification performance in general, when examining learning for the two types of stimuli (monosyllabic and disyllabic stimuli), the results showed distinct patterns in learners' performance. While both training groups improved tonal perception, training with disyllabic stimuli (disyllabic training group) was much more effective (especially for the disyllabic

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 303-319). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.

stimuli) and significantly helped native English-speaking participants to acquire the tones. These results illustrate limitations of the current tone teaching based solely on monosyllabic words. Instead, the current results advocate for incorporating more common and variable disyllabic words in the classroom in order to achieve native-like tone acquisition.

## **1. Introduction**

While it is important for language learners to acquire the correct pronunciation of a target language (Jenkins, 2004), it is especially crucial to acquire native-like pronunciation of tone for language learners of Chinese since Mandarin Chinese is a tonal language in which tone is a key component of the lexicon used to distinguish word meaning. Accurately perceiving and correctly producing tones is of critical importance for Chinese language learners to communicate successfully in the language. In the present study, American learners of Chinese were trained using a high variability phonetic training paradigm, in which two training groups were contrasted: a group trained with monosyllabic stimuli and a group trained with disyllabic stimuli. Accuracy in identifying tonal contrasts before and after training in naturally produced monosyllabic and disyllabic words (produced by native Mandarin Chinese speakers) was evaluated.

## **2. Perceptual training**

Native English learners of Chinese have difficulty perceiving and producing tones in Mandarin Chinese since the phonemic tone feature is not in part of their native language system (Miracle, 1989; Shen, 1989; Shen & Lin, 1991; Sun, 1998; Jongman, Wang, Moore, & Sereno, 2006; Lee, Tao, & Bond, 2010; He, 2010; He & Wayland, 2010, 2013; Chang, 2011; Hao, 2012). These studies have analyzed native English learners' perception of Mandarin tone in isolation and found that American listeners have particular difficulty differentiating Tone 2 (T2) and Tone 3 (T3), attributing the confusion to American listeners assigning more weight to F0 height than F0 direction when perceiving Mandarin T2 and T3 in isolation (Gandour, 1983; Gottfried and Suiter, 1997; Lee, Tao, & Bond, 2009).

While it is vital to understand tones of monosyllabic words in an isolated environment, tones in disyllabic words are equally, or even more significant, since disyllabic words are dominant in the vocabulary of modern Mandarin Chinese (Duanmu, 1999). Disyllabic words and their connected tones are used most often in Chinese people's daily life rather than monosyllabic words, with disyllabic tones mirroring the tones

perceived and produced at the sentence level more than isolated tones. Few studies have examined tones in disyllabic words and tones at the sentential level (Sun, 1998; Guo & Tao, 2008; He, 2010; He & Wayland, 2010, 2013). These researchers found, as expected, that across learning experience and proficiency level, American learners did significantly better at identifying tones in monosyllabic words than in disyllabic words. Moreover, native English learners' accuracy rate of tone perception was systematically improved according to their learning experience: the higher the proficiency level or the longer they studied Mandarin Chinese, the better their accuracy was. When examining learners' identification performance of the four phonemic tones across both monosyllabic and disyllabic words, Sun (1998) found that T2 and T3 were identified significantly poorer than T1 and T4 across all four proficiency level groups. Similarly, He (2010) found that for both monosyllabic and disyllabic tonal contrasts, T3 was the most difficult to identify, then T1, T2 and T4 by inexperienced learners while T2 was the most difficult to identify among the four tones by experienced learners.

Improving native English learners' tonal categories in Mandarin Chinese from the onset of learning the language is clearly important. Moreover, learning not only should pay attention to monosyllabic tones but also should focus on disyllabic tone practice, including tone alternation and coarticulation among the two adjacent tones. These coarticulated tones regularly occur in Mandarin Chinese natural speech contexts, and by examining disyllabic words, English speakers may be able to improve their comprehension and pronunciation of Mandarin.

Current in-class pedagogical approaches to teach Mandarin Chinese tones often use traditional methods, such as listen-and-repeat, minimal-pair drills, and reading aloud tasks. However, short-term auditory training methods in various languages have proved to be effective in assisting learners to acquire new phonetic contrasts that do not exist in their native phonological system (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Francis, Ciocca, Ma, & Fenn, 2008; Herd, Jongman, & Sereno, 2013; Kingston, 2003; Lively, Logan, & Pisoni, 1993; Logan, Lively & Pisoni, 1991; Wang, Spence, Jongman, & Sereno, 1999).

In these high variability training procedures, language learners listen to a large variety of stimuli produced naturally by multiple native speakers of the target language. Within a short period, the learners' perception of non-native language contrasts is improved through the exposure to the target language. Furthermore, this perceptual improvement was successfully extended to the learners' production, as shown by Japanese



learners of English learning /ɪ/ and /I/ (Bradlow, et al., 1997, Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Logan, et al., 1991; Lively, et al., 1993).

High variability phonetic training is not only effective at the segmental level but also at the suprasegmental level (Wang, Jongman & Sereno, 2003; Wang et al., 1999). Wang et al. (1999) found that American learners of Mandarin Chinese showed improved tone perception after training, from a pretest accuracy rate of 69% to a posttest accuracy rate of 90%, a significant improvement (21%). Wang et al. (1999) used eight 40-minute training sessions and showed improved perception of Mandarin monosyllable tonal categories. Furthermore, they found that trainees also showed generalization of the learning to new words and new speakers (Wang et al., 1999). This improvement was also retained six months after training. Wang et al. (2003) extended this perceptual improvement to Mandarin tone production. Using perceptual training techniques, the production data showed that learners' pitch contours better approximated native norms after training. Additionally, identification of trainees' post-test tone productions (compared to their pre-test productions) improved by 18%. These results indicate improved tone identification accuracy and better productions after a short perceptual training period.

Wang et al. (1999, 2003) found that through a short high variability phonetic training using monosyllabic tones in Mandarin Chinese, American beginning learners of Mandarin Chinese all improved significantly in their tonal perception and production of the four Mandarin Chinese tones in monosyllable words. But their study did not address whether the monosyllabic tone training would help learners identify tones in disyllabic words, that is words that are most often encountered in sentences and daily conversations and reflect tonal contexts more accurately. Would learners' tonal perception improve through perceptual training on disyllabic words just as they did with Wang et al.'s training on monosyllabic ones?

Therefore, the current study examined whether perceptual training can effectively be used to train native English-speaking listeners to accurately perceive common and naturalistic (involving tonal coarticulation) disyllabic words. Monosyllabic and disyllabic training will be compared in order to determine the amount of improvement in tone identification. In addition, both monosyllabic and disyllabic stimuli will be examined to determine which type of training material is most effective in helping native English learners to shape tonal categories that do not exist in their phonological inventory.

### **3. Current study**

The purpose of the current study is to examine if beginning English-speaking learners' perception of Chinese Mandarin tones in both monosyllabic words and disyllabic words will be improved after perceptual training as learners gain greater proficiency in Mandarin Chinese.

Three research questions are addressed. First, will disyllabic perceptual training be more or less effective compared to monosyllabic perceptual training in helping English-speaking learners shape their tonal categories and improve their tone perception of Mandarin Chinese? Second, when contrasting these two types of training materials, monosyllabic stimuli and disyllabic stimuli, which will be more effective in learning monosyllabic tones and which will be more effective in learning disyllabic tones? And finally, will training using monosyllabic material transfer to disyllabic tone identification, and will training using disyllabic material transfer to monosyllabic tone identification? The goal is to determine which perceptual training (monosyllabic or disyllabic) will help native English learners of Chinese improve their perception of Chinese words (monosyllabic and disyllabic stimuli), and to examine if there is transfer effect between two types of training in learning tones in Mandarin Chinese.

### **4. Method**

Three phases were included in the Mandarin tone experiment: a pretest, a training session (either monosyllabic or disyllabic training), and a posttest. All Mandarin Chinese beginning learners participated in identical pretests and posttests, with a forced-choice identification (ID) task used. For the pretest and the posttest, both monosyllabic stimuli and disyllabic stimuli were used. For the training sessions, training (either monosyllabic training or disyllabic training) consisted of four perceptual training sessions. The monosyllabic training group was trained exclusively with monosyllabic stimuli while the disyllabic training group was trained exclusively with disyllabic stimuli. For all training sessions, immediate feedback was given after each response for both monosyllabic and disyllabic training groups.

The two training groups were compared across pretest and posttest to observe any improvement after the training. In addition, the performance for the two types of training material (monosyllabic and disyllabic training stimuli) were examined to determine which type of training material would show the most learning improvement.

## 5. Participants

Native English learners of Mandarin Chinese participated in a two-week training program. All participants were beginning learners of the Chinese language with less than two semesters (less than 7 months) of learning Mandarin. All were college students. Overall, seventeen native English learners of Mandarin Chinese participated. Nine learners participated in the monosyllabic training group, and eight learners participated in the disyllabic training group. Participants were randomly assigned to one of the two training groups. None of these seventeen learners had any history of hearing, speech, or language difficulties.

## 6. Stimuli

All the stimuli were recorded by six native Mandarin Chinese speakers, three males and three females, in order to ensure speaker variability. Two types of stimuli, monosyllabic stimuli and disyllabic stimuli, were used throughout the pretest, training, and posttest. All monosyllabic stimuli were adopted from Wang et al. (1999). These monosyllabic stimuli included all possible permissible combinations of various initial consonants and final vowels, and different syllabic structures in Mandarin Chinese (i.e. V, CV, CVNasal, VN, CGlideV, and CGVN). Each disyllabic stimulus was composed of two randomly combined syllables from the monosyllabic stimuli. Thus, every individual syllable used for the disyllabic stimuli was identical to those used in the monosyllabic stimuli. For example, the monosyllabic stimuli *-mǎ* (“horse”) and *-shāng* (“injury”) were combined to form a two-syllable word that served as a disyllabic stimulus, *-mǎ shāng*. All monosyllabic stimuli were real words in Mandarin Chinese; the randomly combined disyllabic stimuli were non-words with a decomposable meaning. To preserve the characteristics of disyllable words in connected speech, all six speakers were instructed to produce the stimuli as natural as possible, and to avoid producing any disyllable stimuli as two separate individual syllables. In total, there were 288 monosyllabic stimuli and 144 disyllabic stimuli in the current experiment.

## 7. Procedure

The present experiment consisted of three phases: pretest, training, and posttest. All the tests and training were conducted in the KU Phonetics and Psycholinguistics Laboratory. All stimuli were presented over headphones

using Paradigm software (Tagliaferri, 2008) and all learners' responses were recorded in Paradigm. Seventeen native English learners of Chinese participated in the two-week training program. Each learner participated for a total of six days (Pretest; Training1; Training2; Training3; Training4; Posttest), with three sessions the first week and three sessions the second week (see Figure 1). The pretest and posttest were 60 minutes long and each training session was 30 minutes long.

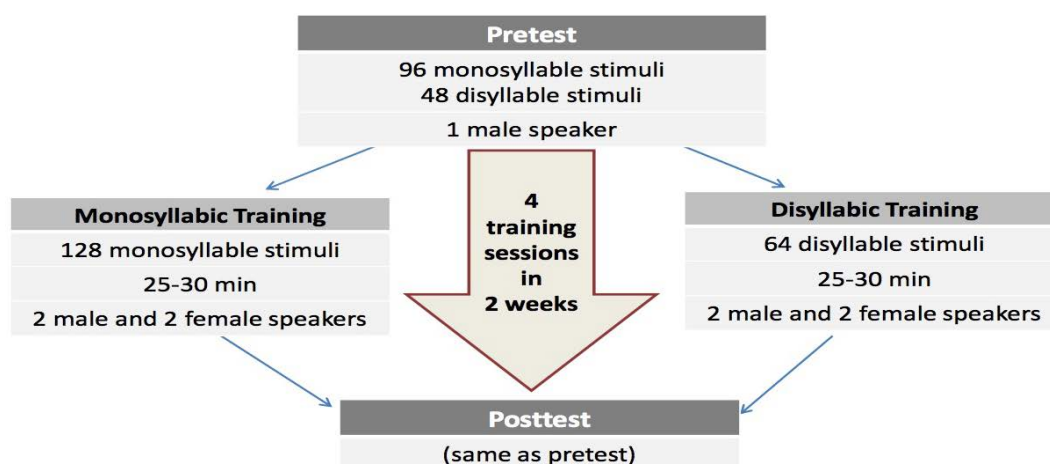


Figure 1. Experimental procedure.

The pretest consisted of two parts, monosyllable word identification and disyllable word identification. All stimuli were produced by a male native Chinese speaker (speaker 1). For both parts, learners indicated which Mandarin Chinese tones they heard. No feedback was provided. The pretest lasted about 60 minutes, approximately 30 minutes for each part.

For the monosyllable pretest, learners heard a monosyllabic stimulus and were instructed to give their tone identification response by pushing the corresponding button that represented one of the four tones (1=Tone1, 2=Tone2, 3=Tone3, and 4=Tone4). All tonal diacritics and numbers were labeled on the buttons on the keyboard. There were 96 monosyllabic stimuli in the pretest. Stimuli in the monosyllable pretest were the same 96 monosyllabic stimuli as in Wang et al. (1999) study. All monosyllabic stimuli were real words in Mandarin Chinese. There were 24 monosyllable words for each of the four phonemic Mandarin tones. All monosyllabic stimuli were presented with a 3 second inter-trial interval (ITI). Learners' accuracy during the identification task were recorded in Paradigm (Tagliaferri, 2008).

For the disyllable pretest, the learners heard a disyllabic stimulus and they were asked to indicate their tone identification response by pushing two corresponding buttons (one response followed by the other response) that represented the tone of the first syllable followed by the tone of the second syllable (1= Tone1, 2= Tone2, 3= Tone3, and 4= Tone4). All tonal diacritics and numbers were labeled on the buttons on the keyboard. There were 48 disyllabic stimuli in the pretest. Each disyllabic stimulus was composed of two randomly combined syllables from the monosyllabic stimuli. Thus, every individual syllable used for the disyllabic stimuli was identical to those used in the monosyllabic stimuli. The randomly combined disyllabic stimuli were non-words with a decomposable meaning. There were 3 disyllable words for each of the 16 (4 tones X 4 tones = 16 pairs) combinations. In order to directly compare identification of the disyllable and monosyllable stimuli, accuracy for each syllable of the disyllabic stimuli was tabulated. So if a T1 + T4 was presented and the response was T2 + T4, the first syllable was recorded as incorrect and the second syllable was recorded as correct. Also, due to a productive third tone sandhi rule in Mandarin, for one of the sixteen pairs (Tone3 + Tone3), the first Tone 3 syllable is systematically produced as a Tone 2 when followed by a Tone 3 syllable. For these stimuli, the correct identification was Tone 2 + Tone 3. As with the monosyllabic part, the ITI was 3 seconds, all disyllable tonal diacritics and numbers were labeled on the buttons, and no feedback was given. Learners' accuracy in the identification task were recorded in Paradigm (Tagliaferri, 2008).

## **8. Training sessions**

Both Monosyllabic and Disyllabic training consisted of four perceptual training sessions that lasted 30 minutes each. Learners participated in a forced-choice ID task and immediate feedback was given after each response for all training sessions to help learners focus their attention on the critical acoustic cues of the four tones.

### **8.1. Monosyllabic training**

The monosyllabic training group was trained exclusively with monosyllabic stimuli. There were 128 monosyllabic training stimuli, which consisted of 32 monosyllable words for each of the four tones. All were produced by four native Chinese speakers (speaker 2, speaker 3, speaker 4, speaker 5), including two male speakers and two female speakers. Overall, there were 512 stimuli in the monosyllabic training produced by the four native Chinese speakers.

For monosyllabic training, participants heard a stimulus, “*má*”, which contained a target tone (e.g., Tone 2) in a monosyllabic word, and they then indicated what they heard among four tones (1=T1, 2=T2, 3=T3, and 4=T4) by pushing the corresponding button on the keyboard. If the choice was correct, the participant would hear: “Correct! That was Tone 2, it is *má*.” The next stimulus was then presented. If the response was incorrect, the participant would hear: “Uh-oh! That was *má*, Tone 2. Let’s hear it again *má*”. In each of the four training sessions, the trainees were trained with the stimuli produced by only one speaker at a time.

## 8.2. Disyllabic training

The disyllabic training group was trained exclusively with disyllabic stimuli. The monosyllabic training stimuli were used to create the disyllabic stimuli, which shared all the same syllables as those in the monosyllabic training. There were 64 disyllabic training stimuli. The same four native Chinese speakers (speaker 2, speaker 3, speaker 4, and speaker 5) produced these 64 disyllabic stimuli. In each session, the learners heard stimuli only produced by one speaker. Overall, then, there were 256 disyllabic stimuli across the four training sessions.

For disyllabic training, participants heard a disyllabic stimulus, for example, “*mǎ shāng*”, which was a Tone 3 + Tone 1 combination. The learner would then make two responses by pushing two buttons sequentially on the keyboard. The accuracy of each syllable of the disyllable stimulus was counted. Immediate feedback was given just as in the monosyllabic training. For instance, if the choice was correct, the participant would hear: “Correct! That was Tone 3 and Tone 1, it is *mǎ shāng*.” The next stimulus was then presented. If either of the two responses was incorrect, the participant would hear: “Uh-oh! That was *mǎ shāng*, Tone 3 and Tone 1. Let’s hear it again *mǎ shāng*. ” After feedback, the next stimulus was presented.

## 8.3. Posttest

The posttest was identical to the pretest, including both monosyllabic stimuli and disyllabic stimuli. Learners indicated which Mandarin Chinese tones they heard by pushing the corresponding button for the four tones (1=Tone1, 2=Tone2, 3=Tone3, and 4=Tone4) and received no feedback. The posttest lasted about 60 minutes, approximately 30 minutes for each part.

## 9. Results

A binomial logistic regression was conducted to examine the effect of training type on the perception of Chinese tones, using the lme4 package (Bates, 2005; Bates & Maechler, 2010) in the R statistical environment (R development Core Team, 2012, Version 3.4.3). The model had Correct (1=Correct vs. 0=Incorrect) as a dependent variable, and Training Group (Monosyllabic Training vs. Disyllabic Training), Tested Stimuli (Monosyllabic Stimuli vs. Disyllabic Stimuli), and Test (Pretest vs. Posttest) as fixed effects. Subjects and Stimuli were entered as random factors. When there was a significant three-way interaction among the independent variables, we stratified the data by Tested Stimuli to probe the interaction.

The model showed a significant main effect of Test,  $c^2(1)= 51.16$ ,  $p<0.001$ , indicating that there was a significant improvement after the training from pretest (72%) to posttest (80%), an 8% improvement. There was also a significant main effect of Tested Stimuli,  $c^2(1)= 170.45$ ,  $p<0.001$ , indicating that the participants identified monosyllabic stimuli more accurately (90%) than the disyllabic stimuli (65%). We also found a significant two-way interaction between Tested Stimuli and Test ( $c^2(1)= 9.05$ ,  $p=0.002$ ), indicating that there was a greater improvement on disyllabic stimuli (9% improvement) than monosyllabic stimuli (7% improvement) after the training. We also found a significant interaction between Test and Training Group ( $c^2(1)= 6.38$ ,  $p=0.011$ ), indicating that the disyllabic training showed a greater improvement (11% improvement) than the monosyllabic training (6% improvement). A statistically significant three-way interaction among Tested Stimuli, Test, and Training Group was also found ( $c^2(1)= 6.45$ ,  $p=0.011$ ).

To further examine this three-way interaction, we stratified the data by Tested Stimuli, and ran two binomial logistic regressions for each stimuli type, including Training Group and Test as main effects and Subject and Stimuli as random effects. The model analyzing the monosyllabic stimuli showed a main effect of Test ( $c^2(1)= 44.35$ ,  $p<0.001$ ) only, indicating that there was a significant improvement on identifying monosyllabic stimuli after the training regardless of the training regime (8% improvement). For the monosyllabic test stimuli, the accuracy of pretest and posttest for the monosyllabic training group was 87% and 94%, and the accuracy of the pretest and posttest for the disyllabic training group was 82% and 90%, respectively. Figure 2 indicates the similar degree of improvement between the monosyllabic training group and disyllabic training group for the monosyllable tested stimuli.

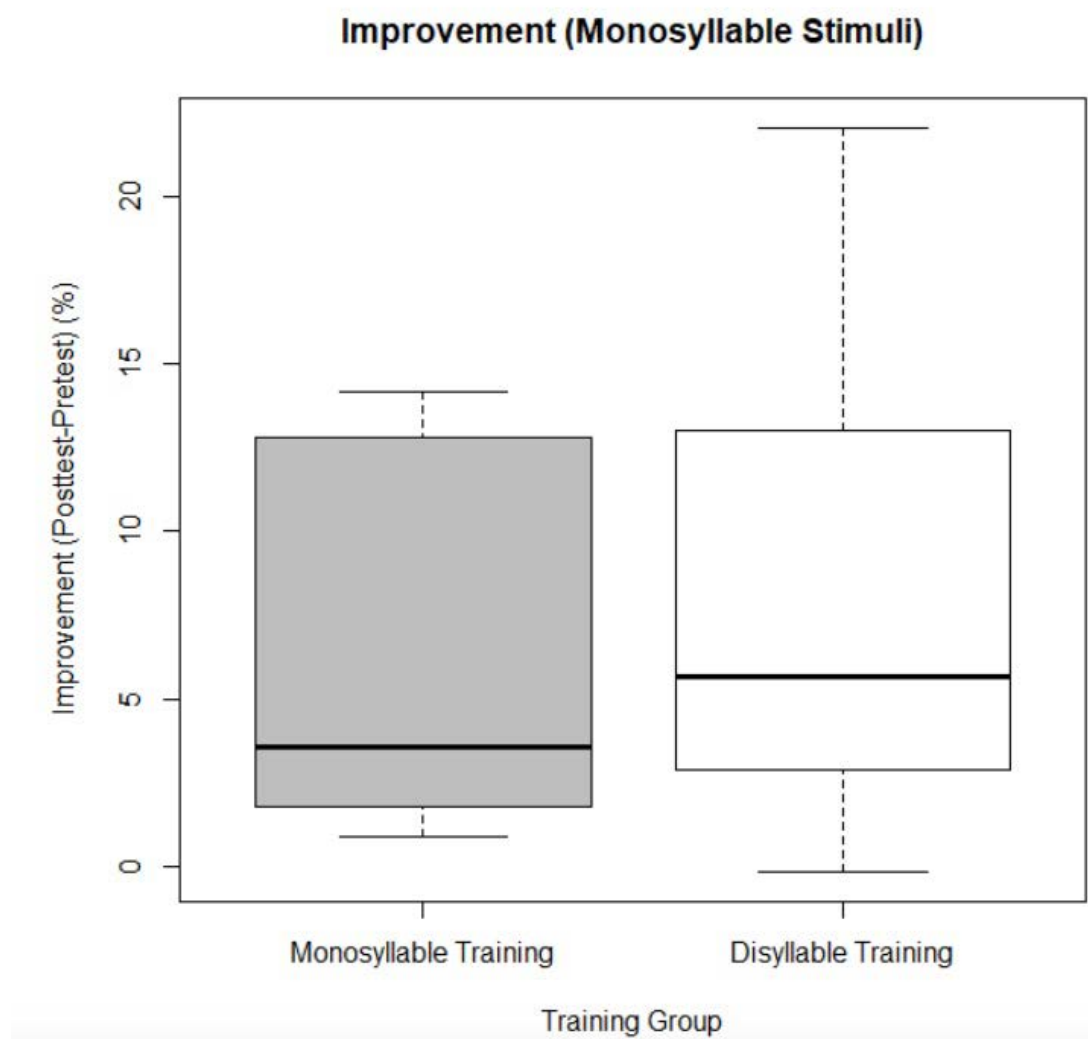


Figure 2. Percent improvement (%) (and standard error) for the monosyllabic stimuli from pretest to posttest by native English-speaking learners of Chinese in the monosyllabic and disyllabic training groups.

The model analyzing the tested disyllabic stimuli also showed a main effect of Test ( $c^2(1)=16.70$ ,  $p<0.001$ ), indicating that there was a significant improvement on identifying disyllabic stimuli after the training (9% improvement) regardless of the training regime. However, for the disyllabic stimuli, we also found a significant two-way interaction between Test and Training Group ( $c^2(1)= 11.86$ ,  $p<0.001$ ), indicating that the disyllabic training group improved more in identifying tones in disyllabic stimuli than the monosyllabic training group did. The accuracy of the pretest and posttest for the monosyllable group was 65% and 70%, while the accuracy of the pretest and posttest for the disyllabic group was



51% and 66%, respectively. Figure 3 indicates the different degree of improvement between the monosyllabic training group and the disyllabic training group for the disyllabic stimuli, with disyllabic training showing greater gains than monosyllabic training.

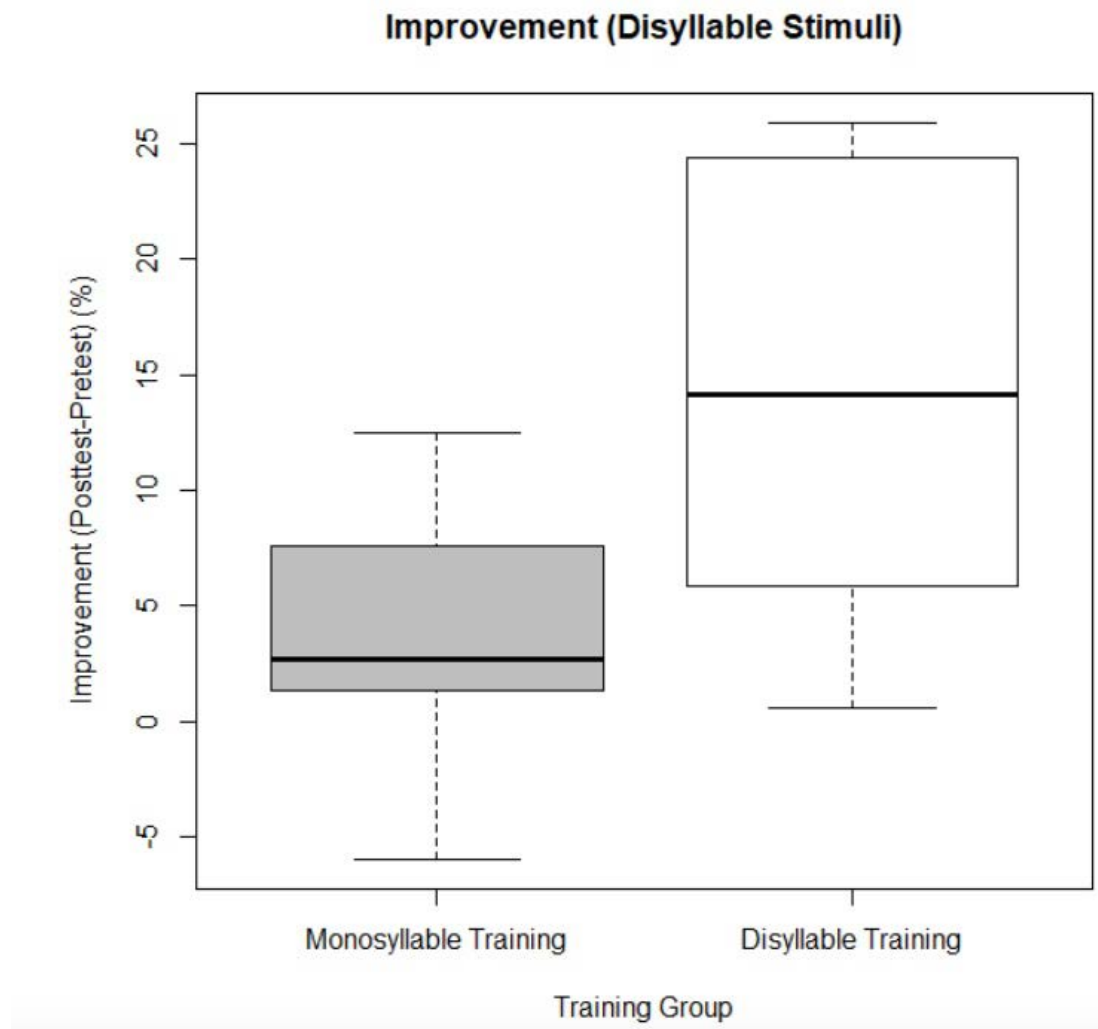


Figure 3. Percent improvement (%) (and standard error) for the disyllabic stimuli from pretest to posttest by native English-speaking learners of Chinese in the monosyllabic and disyllabic training groups.

## 10. Discussion

The results of the current study demonstrated that after high variability perceptual training, adult native English-speaking beginning learners of Chinese were able to significantly improve their tone perception in both monosyllabic and disyllabic stimuli in Mandarin Chinese. Participation in

a short (four 30-minute sessions) two-week training showed a significant 8% increase ( $p < 0.001$ ) from pretest 72% to posttest 80% in learners' overall tone perception accuracy. These data are similar to Wang et al. (1999) in which examined monosyllabic perceptual training, showing a sizable 21% increase. More substantial learning in their study was most likely due to the fact that more training sessions were used (8 sessions of 40 minutes each) and also due to the fact that training stimuli for Wang et al. (1999) were arranged pairwise, which allowed for a systematic increase in difficulty of tone contrasts as learning progressed. Interestingly, while Wang et al. (1999) only used monosyllabic stimuli with monosyllabic training, the current study showed that while identification of tones in disyllabic stimuli is more challenging, there was greater improvement on disyllabic stimuli (9% improvement) than monosyllabic stimuli (7% improvement) after training. These data suggest that inclusion of more complex and variable disyllabic stimuli will not harm the beneficial aspects of high variability training.

It should be noted that learners generally did significantly better ( $p < 0.001$ ) when identifying tones in monosyllabic stimuli, with an accuracy of 90%, as compared to disyllabic stimuli, with an accuracy of 65%. Such a substantial identification accuracy gap between the two types of stimuli was also observed by Sun (1998) and He (2010). Recall that in the current study, identification of the monosyllabic stimuli in pretest and posttest is based on an isolated syllable while, for the disyllabic stimuli, listeners heard a sequence of two syllables which they were asked to identify. Differences in overall monosyllabic and disyllabic identification accuracy are likely due to the tonal environment, with tones in monosyllabic stimuli occurring in isolated environments, such that these tones are preserved in their canonical forms, while tones in disyllabic stimuli were often coarticulated with the adjacent tones' pitch (Shen, 1990; Xu, 1994, 1997, 1998) or they undergo contextually-driven phonological processes (e.g. third tone sandhi). Despite overall accuracy differences between monosyllabic and disyllabic stimuli and the challenges of disyllabic tone identification, the current results show that there was greater improvement on disyllabic stimuli (9% improvement) than monosyllabic stimuli (7% improvement) after training. Given these data showing successful improvement using disyllabic stimuli, teachers, when teaching Mandarin Chinese tones, should not shy away from providing students with disyllabic stimuli that contain more contextual variability.

More importantly, the current findings also showed significant differences between the monosyllabic perceptual training group and the disyllabic perceptual training group from pretest to posttest. Critically, these differences due to training were observed regardless of the syllabic structure of the stimuli tested. When identifying tones in both types of stimuli (monosyllabic and disyllabic), the learners in the monosyllabic training group showed a significant 6% increase from pretest 76% to posttest 82% ( $p < 0.001$ ). Similarly, learners in the disyllabic training group also showed a significant improvement from the pretest 67% to the posttest 78%, with an 11% increase ( $p < 0.001$ ). While both monosyllabic and disyllabic perceptual training was beneficial for learners to aid in building robust tonal categories in Mandarin Chinese, those learners who had disyllabic training made nearly double the improvement (11%) on their tonal identification compared to the monosyllabic training group (6%). The disyllabic training for native English-speaking learners seemed to provide more effective learning of Mandarin Chinese tones in both monosyllabic and disyllabic stimuli than did the monosyllabic training.

Interestingly, transfer effects of training were also found in current study. Learners who received the monosyllabic training improved significantly when perceiving tones not only in monosyllabic stimuli (from pretest 87% to posttest 94%), but also in disyllabic stimuli (from pretest 65% to posttest 70%) ( $p < 0.001$ ). Moreover, learners who received the disyllabic training not only showed substantial improved tone identification when identifying tones in disyllabic stimuli (from pretest 51% to posttest 66%), but also in monosyllabic stimuli (from pretest 82% to posttest 90%) ( $p < 0.001$ ). These results show that both training regimes seem to improve tonal perception, with either monosyllabic training or disyllabic training being beneficial for learners to identify Mandarin Chinese tones in monosyllabic stimuli and disyllabic stimuli. But importantly, while listeners in the monosyllabic perceptual training group exhibited similar improvement for both monosyllabic and disyllabic test stimuli (7% and 5%, respectively), listeners in the disyllabic training group showed more improvement, as expected, in the disyllable test stimuli (15%), but also showed substantial improvement in the monosyllabic stimuli (8%). Thus, when teaching the language, it may be helpful for instructors to introduce tones in disyllable words since this exposure provides learners with more typical real-world contexts exhibiting more tonal variability, and, crucially, this encourages learners to develop more robust tonal categories.

## 11. Conclusion

This study investigated whether native speakers of English can be guided using a high variability phonetic training method to accurately perceive Mandarin Chinese tones in monosyllabic stimuli and disyllabic stimuli. The perception results clearly showed that learners improved their tone accuracy for both monosyllabic and disyllabic stimuli after a short period of perceptual training. Additionally, this research investigated which training group, the monosyllabic training group or the disyllabic training group, would be most helpful for native English-speaking learners to establish tonal categories in their speech system. Although both groups' identification performance improved, it was found that the learners in the disyllabic training group seemed to show more learning not only on disyllabic tones but also on monosyllabic tones when comparing to those in the monosyllabic training group. These data show that disyllabic tones with tonal variation and coarticulation can help learners. Mandarin Chinese classes should not solely focus on teaching tones in isolation, but should also include disyllabic stimuli, as a way to improve learning and better simulate natural learning environments.

## References

- Bates, D. (2005). Fitting linear mixed models in R. *R News*, 5(1), 27-30, <http://CRAN.R-project.org/doc/Rnews/>
- Bates, D., & Maechler, M. (2010). Matrix: sparse and dense matrix classes and methods. R package version 0.999375-43, <http://cran.r-project.org/package=Matrix>.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English /r/ and /l/: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101(4), 2299-2310.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61(5), 977-985.
- Chang, Y.-h. S. (2011). Distinction between Mandarin Tones 2 and 3 for L1 and L2 Listeners. In Z. Jing-Schmidt (Ed.), *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23)*. 1, pp. 84-96. Eugene: University of Oregon.

- Duanmu, S. (1999). Stress and the development of disyllabic words in Chinese. *Diachronica*, vol. 16 (1), 1-35.
- Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, 36(2), 268-294.
- Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11(2), 149-175.
- Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, 25(2), 207-231.
- Guo, L., & Tao, L. (2008, April). Tone production in Mandarin Chinese by American students: A case study. In *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)* (Vol. 1, pp. 123-138).
- Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, 40(2), 269-279.
- He, Y. (2010). Perception and production of isolated and coarticulated Mandarin Tones by American Learners. University of Florida, Gainesville, Ph.D. Dissertation.
- He, Y., & Wayland, R. (2010). The production of Mandarin coarticulated tones by inexperienced and experienced English speakers of Mandarin. In *Speech Prosody 2010-Fifth International Conference*, 123:1.4.
- He, Y., & Wayland, R. (2013). Identification of Mandarin coarticulated tones by inexperienced and experienced English learners of Mandarin. *Chinese as a Second Language Research*, 2(1), 1-21.
- Herd, W., Jongman, A., & Sereno, J. (2013). Perceptual and production training of intervocalic/d, r, r/in American English learners of Spanish. *Journal of the Acoustical Society of America*, 133(6), 4247-4255.
- Jenkins, J. (2004). Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics*, 24, 109-125.
- Jongman, A., Wang, Y., Moore, C. and Sereno, J. A. (2006). Perception and production of Mandarin tone. In E. Bates, L. Tan, and O. Tzeng (Eds.), *Handbook of East Asian Psycholinguistics*, Cambridge: Cambridge University Press, 209-217.
- Kingston, J. (2003). Learning foreign vowels. *Language and Speech*, 46(2-3), 295-348.
- Lee, C. Y., Tao, L., & Bond, Z. S. (2010a). Identification of multi-speaker Mandarin tones in noise by native and non-native listeners. *Speech Communication*, 52(11), 900-910.
- Lee, C. Y., Tao, L., & Bond, Z. S. (2010b). Identification of acoustically modified Mandarin tones by non-native listeners. *Language and Speech*, 53(2), 217-243.

- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, 94(3), 1242-1255.
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89(2), 874-886.
- Miracle, W. C. (1989). Tone production of American students of Chinese: A preliminary acoustic study. *Journal of the Chinese Language Teachers Association*, 24(3), 49-65.
- Orton, J. (2013). Developing Chinese oral skills-a research base for practice. *Research in Chinese as a Second Language*, 9, 3-26.
- Shen, X. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association*, 24(3), 27-47.
- Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech*, 34(2), 145-156.
- Sun, S. H. (1998). *The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese* (No. 16). University of Hawaii Press.
- Tagliaferri, B. (2008). Paradigm: Perception Research Systems [Computer Program]. Retrieved from <http://www.paradigmexperiments.com>.
- Wang, Y., Spence, M., Jongman, A., and Sereno, J.A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, 106, 3649-3658.
- Wang, Y., Jongman, A., and Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after training. *Journal of the Acoustical Society of America*, 113, 1033-1043.
- Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25(1), 61-83.



# Pitch Accents as Multiparametric Configurations of Prosodic Features – Evidence from Pitch-accent Specific Micro-rhythms in German

Oliver Niebuhr  
University of Southern Denmark

## Abstract

Pitch accents are typically described in terms of the alignment and shape of their F0 peaks. However, some studies suggest that pitch-accent peaks also create systematic duration and intensity changes in the triplet of pre-accented, accented, and post-accented syllable. The present study examines this phenomenon in detail for three rising-falling German pitch accents: the early, medial, and late peak. A production experiment with 4 speakers finds clear acoustic evidence for these systematic duration and intensity changes. In addition, these changes also manifest themselves in a parallel dataset of syllable-synchronous finger tapping. In combination, the changes of two prominence-relevant acoustic parameters, i.e. syllable duration and intensity, and the reflection of these changes in a rhythmical finger-tapping task suggest that nuclear pitch-accent categories in German are not purely intonational phenomena but multiparametric prosodic configurations (i.e. “prosodic constructions”) that include, besides their F0-peak characteristics, a pitch-accent-specific micro-rhythm in the triplet of pre-accented, accented, and post-accent syllable. The implications of this conclusion for intonational modeling are discussed.

## 1. Introduction

It is 30 years ago now that Kohler (1987) published his seminal paper on the categorical perception of F0-peak alignment. Kohler shifted a constant sharply rising-falling nuclear pitch-accent peak in 11 steps

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 321-351). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.



across the sentence “Sie hat ja gelogen” (She’s been lying, with the relevant nuclear pitch accent on [lo:] of “gelogen” [g̊ilo:g̊]). For each of the 11 equidistant 30-ms steps of the F0 peak-shift continuum, a stimulus was resynthesized. The resulting 11 stimuli were included in a serial discrimination test, a 2AFC AX pair discrimination test, and a 2AFC indirect identification test, in which listeners judged the stimulus sentence as either matching or not matching with a constant preceding context utterance (see also Nash & Mulac, 1980 further explanations on this test paradigm and the discussion of semantic tasks in Gussenhoven, 1999). Based on the integrated results of these experiment series, Kohler found a categorical change in perception for those stimuli whose F0 peak maximum was no longer located before but inside the accent vowel of [lo:] in “gelogen” and another, slightly weaker categorical change in perception as soon as the F0 peak was shifted with its maximum out of the accented vowel (Kohler, 1991a).

It is probably not an exaggeration to state that the finding of a categorically perceived F0-peak alignment continuum marked an important milestone for the development of phonological models of intonation and for the linguistic analysis of intonation in general. Categorical perception showed to researchers of those days that the gap between segmental elements and melodic elements (like F0 peaks) was not as big as had been commonly assumed, and that intonation would thus be open to linguistic approaches and phonetic analyses in a similar way as sound segments are. Other milestones were undoubtedly the works of Bruce (1977) and Pierrehumbert (1980), whose frameworks and conclusions also shaped Kohler’s expectation about the perceptual organization of the F0 peak-shift continuum in German and, thus, about the basic principles of Kohler’s Kiel Intonation Model, KIM. With reference to the perceptual tripartition of his F0 peak shift continuum and the alignment characteristics of each perceptual category relative to the boundaries of the accented vowel, Kohler (1987, 1991a) called the three identified pitch accents the early peak (i.e. the F0 maximum is before the vowel), the medial peak (i.e. the F0 maximum is inside the vowel), and the late peak (i.e. the F0 maximum is after the vowel). In terms of their communicative function, early peaks are used to mark a piece of information as being settled or unchangeable. Medial peaks signal new information and openness to discussing this new information with the interlocutor. Late peaks also signal new information, but additionally mark this new information as being in contrast to the speaker’s expectation (Kohler, 2005). Depending

on the context, the early-peak meaning can also express resignation. The late-peak meaning can express either positive surprise or indignation (Niebuhr, 2007a). In major autosegmental-metrical (AM) models of German intonation, such as GToBI, the three pitch-accent categories correspond to H+L\* (or H+!H\*), H\*, and L+\*H, see Grice et al. (2005). And since the accents consist of rising-falling intonation peaks, GToBI would also add a L-% to each label in phrase-final (nuclear) position, which is the position relevant to the present paper.

The historic experimental genesis of the early, medial, and late peak as well as their acoustic definitions by Kohler are probably well known among most intonation researchers. Perhaps less well known, however, is the fact that Kohler encountered slightly different results in a later replication of his peak-shift experiment (Kohler, 1991b). For example, for the same F<sub>0</sub> peak-shift continuum in the stimulus sentence “Sie hat ja gejedelt” (She’s been yodeling, the relevant nuclear accent was on the final word and its syllable [jo:]) the perceptual change from early-peak to medial-peak perception occurred significantly earlier than in the original “gelogen”-sentence. That is, replacing the accent syllable “-lo-” [lo:] by “-jo-” [jo:] had a decisive influence on the category boundary. Kohler (1991b) explains this different outcome by the less sharp segment boundary between the accented vowel and the preceding approximant in [jo:] as compared to [lo:]. Unlike in [lo:], the continuous movement of the tongue in [jo:] does not create an inherent articulatory and spectral discontinuity landmark. On this basis, Kohler argues that the listeners were unable to detect the segment boundary accurately.

Niebuhr (2006, 2007b) countered this argument by pointing out the fact that a blurred segment boundary would have also led to a blurred, i.e. less categorical change in perception from early to medial peak. However, there is no evidence for such a weaker category boundary in Kohler’s data. So, in order to find an alternative explanation for the fact that the pitch-accent boundary is closer to the accented-vowel onset in “Sie hat ja gejedelt” than in “Sie hat ja gelogen”, it was necessary to look for further aspects, in which “gejedelt” differed from “gelogen”. Niebuhr (2006, 2007b) focused on the intensity contour. Due to the approximant at the syllable onset of [jo:] in “gejedelt”, the intensity increase into the accented vowel begins earlier and at a higher level than in the case of the alveolar lateral in [lo:] of “gelogen”. Due to its higher level, the intensity increase is also shorter and culminates earlier in the accented vowel than in [lo:] of “gelogen”.

On this basis, Niebuhr (2006, 2007b) hypothesized that the key factor in the alignment characteristics of early, medial, and late peaks is not the position of the F0 peak maximum relative to the spectrally defined segment boundary of the accented vowel (the vowel onset in the transition from early to medial; and the vowel offset in transition from medial to late). Rather, Niebuhr assumed that the actual key factor in the transition from early to medial and from medial to late peak perception would be the positioning of F0 and intensity movements or their maxima relative to one another.

Niebuhr gained experimental-empirical evidence for this hypothesis in a perception experiment whose methodology is largely adopted from the seminal study of Kohler (1987). Two stimulus series were created. The first series resulted from shifting a pointed rising-falling F0 peak in 11 steps from an early to a medial position into the accented vowel of “Ma-“ in “Sie war mal Malerin” (She was once a painter, with the nuclear pitch accent being on [ma:] of “Malerin” [ma:ləʁɪn]). The second series was derived from the first series such that each stimulus showed exactly the same F0 and intensity patterns as in the first series. Only the segmental string was removed and replaced by a constant schwa-like sound (‘hum’ in PRAAT). The stimuli of the two series were judged by different groups of German native speakers. The judgments for the first stimulus series were made on the basis of an indirect-identification test. The stimuli of the second series were presented in AXB triplets, with A and B being the first or the last stimulus of the series. Similar to the indirect-identification test in which the speech stimuli were judged on a semantic basis as either matching or not matching with a given context utterance, the frame of A and B in the AXB triplets also provided a constant context frame against which the individual ‘hum’ stimuli (X) could be judged - on a melodic basis - as either matching or not matching (with A or B, respectively). In this sense, the listeners’ tasks in the two experiments were designed to be as closely related as it was possible for a comparison of speech and non-speech stimuli.

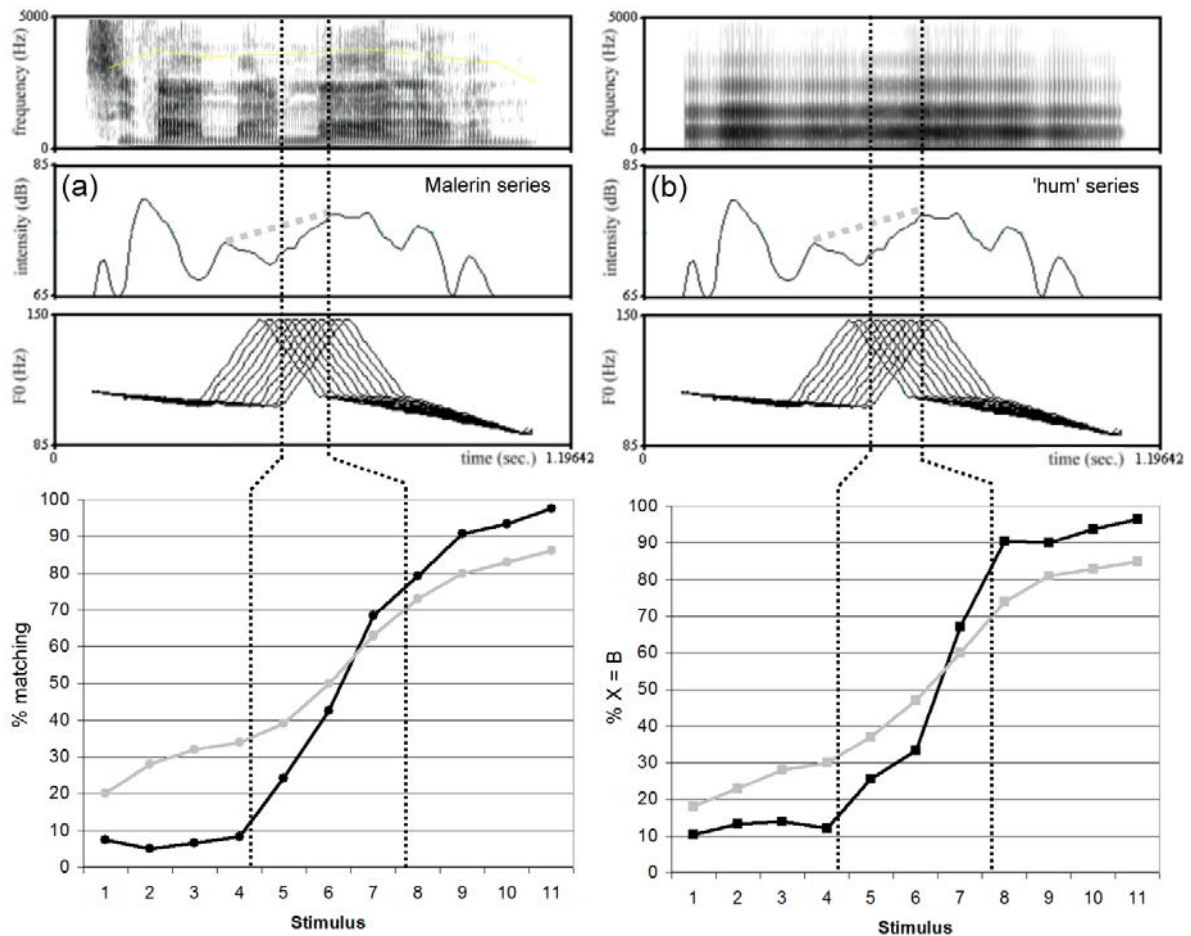


Figure 1. Top: the 11 stimuli of the ‘Malerin’ (a) and ‘hum’ (b) series. Bottom: percentages ( $n=140$ ) of medial-peak intonations in terms of ‘matching’ (a) or (b) ‘ $X=B$ ’ judgments; grey lines in top and bottom panels refer to the repeated experiments, in which the intensity increase into the accented vowel was more gradual.

As is shown in Figure 1, the first stimulus series yielded an abrupt change from the early-peak to the medial-peak category, just as in Kohler’s original “gelogen” series. The crucial new finding is, the second series (‘hum’) was able to replicate this perceptual change from early to medial peak perception solely on the basis of the F0 and intensity patterns of the first series. Moreover, the change from early to medial peak intonation exactly coincided with the intensity increase from the low level of the consonant to the high level of the vowel of the accented syllable. This coincidence made Niebuhr (2007b) repeat the experiment of Niebuhr (2006) with a single modification: the steep intensity increase across the CV boundary in the stimuli was turned into a more gradual one by means of the intensity-course manipulation procedure in Adobe Audition, cf.

the gray lines in Figure 1. As a result, the less dynamic intensity increase was clearly paralleled by a less dynamic perceptual transition from early to medial peak intonation across both the original and the delexicalized ‘hum’ stimulus series. Niebuhr (2007b) also applied the same experimental procedure to a F0 peak-shift continuum from medial to late and gained similar results. That is, the intonation judgments for the ‘hum’ stimuli were statistically identical to those of the original speech stimuli, and the perceptual change from medial to late was the more gradual the more gradual the intensity decrease was after the accented-vowel offset.

However, it is not just that the intensity contour characteristics underlying a F0-peak pattern influence its pitch-accent identification. The intensity characteristics are also systematically varied by speakers in the production of pitch accents. In selecting and creating base stimuli for his perception experiments, Kohler (1991c, p. 144) already noted a “natural parallelism” of the F0 and intensity patterns in the production of pitch accents. Moreover, own informal listening made Kohler assume that these “coupled time courses [of F0 and intensity] are expected by listeners” (Kohler 1991c, p. 188), because breaking this natural parallelism (e.g., by manipulating the F0-peak alignment) seems to have, in Kohler’s ears, negative consequences for the naturalness of the respective sentences and the clarity with which the pitch accents inside these sentences are perceived.

Niebuhr & Pfitzinger (2010) took up Kohler’s idea and investigated the production and perception of the F0 and intensity contours of pitch accents in more detail. An acoustic analysis of read-speech material showed, not surprisingly, that the accented syllable always had higher duration and intensity levels than the two surrounding syllables. However, in addition, the read-speech material revealed pitch-accent-specific intensity levels in the triplet of pre-accented, accented, and post-accented syllable. Moreover, the variation in the intensity patterns was linked with a variation in syllable duration, a phenomenon which was already briefly pointed out by Gartenberg & Panzlaff-Reuter (1991). As is displayed in Figure 2 (upper panel), the early peak was consistently produced with high duration and intensity levels in the pre-accented syllable, whereas the duration and intensity levels in the post-accented syllable were both relatively low. The late peak had an opposite effect on the duration and intensity levels in the pre- and post-accented syllables. That is, the post-accented syllable was consistently realized with higher duration and intensity levels than the pre-accented syllable. Compared to both the early and the late peak, the medial peak was characterized by a

roughly symmetrical duration and intensity pattern across the triplet of pre-accented, accented, and post-accented syllable. While the accented syllable clearly stood out in terms of duration and intensity, the pre-accented and post-accented syllables were both produced at similarly low duration and intensity levels relative to the accented one.

Based on these production patterns, Niebuhr & Pfitzinger (2010) conducted a perception experiment with a semantic differential. They used two types of stimuli: (1) PSOLA resyntheses of original “Eine Malerin” (a painter) utterances being produced with early-, medial- and late pitch-accent peaks on the nuclear-accent syllable [ma:] of “Malerin”; and (2) PSOLA resyntheses of these original “Eine Malerin” productions but with interchanged F0 contours. The condition-(1) stimuli were only resynthesized in order to impair their sound quality in the same way as for the condition-(2) stimuli. The interchanged F0 contours in the condition-(2) stimuli had the same proportional F0-peak alignments relative to the vowel boundaries as in the condition-(1) stimuli. Naturally produced differences in F0-peak shape between the early, medial, and late peak categories were also retained.

The stimuli were presented with multiple repetitions and in differently randomised orders to native speakers of German. However, the stimuli of condition (1) were judged in a separate block after those of condition (2). The results are perfectly in accord with Kohler’s (1991c) assumptions, see Figure 2, lower panel. Firstly, the stimuli with interchanged F0 contours sounded significantly more artificial than the original stimuli. Secondly, for the stimuli with interchanged F0 contours there was a separate effect of the original duration and intensity pattern. It biased judgments towards the semantic profile of that pitch-accent category with which the duration and intensity pattern was naturally produced. Thus, the findings suggest that listeners are sensitive to the duration and intensity patterns that co-occur with a certain pitch accent, and that duration and intensity make a separate contribution to identifying or conveying the communicative functions of that pitch accent. Later, follow-up experiments by Niebuhr (2011) suggest further that listeners have an internal representation of the pitch-accent-specific duration and intensity patterns shown in the upper panel of Figure 2. The “Eine Malerin” utterances were resynthesized with completely flattened F0 course and listeners were asked, in one experiment, to repeat the corresponding utterance in a more melodic fashion (and with stress on “Malerin”) or, in another experiment, to draw the speech melody that they consider appropriate for the corresponding stimulus utterance on a sheet of paper. Both

experiment tasks yielded a significant correlation between the originally produced but then flattened and hence removed pitch accent in the stimulus utterance and the pitch accent that was reinserted into the repeated stimulus utterance or drawn by the participant. The only possible acoustic cues that had been able to guide these performances were the retained duration and intensity patterns in the stimuli.

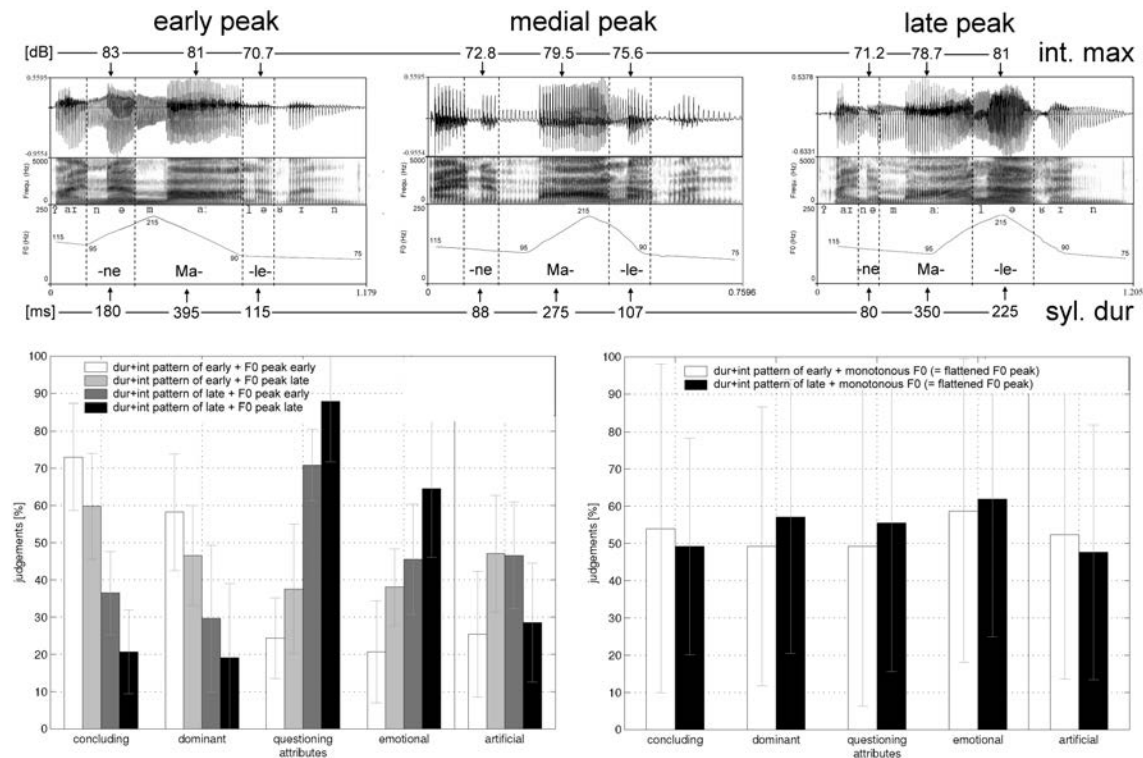


Figure 2. Top panel: Acoustic characteristics of pre-accented, accented, and post-accented syllables in “Eine Malerin” (a painter), produced with the early (left), medial (mid), and late (right) pitch accent on the accented syllable “Ma-“. Bottom panel: Listener judgments on the key meaning dimensions of early and late pitch accents when being presented in original and exchanged duration and intensity contexts (left) and with flattened F0 peaks (right).

Since both duration and intensity are also important triggers of perceptual prominence in German, Niebuhr & Pfitzinger (2010) decided to refer to these characteristic duration and intensity levels that co-occur with German pitch accents and shape the triplet of pre-accented, accented, and post-accented syllable as *pitch-accent-specific micro-rhythms*. The term “rhythm” was used because Niebuhr & Pfitzinger indeed noted, on an informal auditory basis, a characteristic tri-syllabic prominence sequence for the early, medial, and late pitch accent; and a rhythm is nothing else than a sequence of different syllable-based prominences.

However, Niebuhr & Pfitzinger also noted that this rhythm is relatively subtle in perception and embedded in a larger sequence of higher-level syllable prominences, determined by the lexical-stress realizations and pitch-accent distributions in utterances. Therefore, and with respect to the restricted three-syllable time domain in which these prominence differences occurred, Niebuhr & Pfitzinger used the term “micro-rhythm”.

Against the outlined body of empirical evidence on the role of duration and intensity patterns in the production and perception of German early, medial, and late pitch accents, the point of departure of the present study is as follows: Although Niebuhr & Pfitzinger coined the term pitch-accent-specific micro-rhythm for the duration and intensity patterns they found, there is no direct evidence as yet, that the characterization as “rhythm” is actually appropriate. That is, the question addressed here is whether the duration and intensity patterns across the triplet of pre-accented, accented, and post-accented syllable in fact represent a rhythm in the performance-oriented, prominence-related, and cognitive sense of the term.

In order to shed light on this issue, a formal speech-production experiment was performed. The experiment makes use of a method that has been successfully applied for “tapping into naïve listeners’ intuitions about speech rhythm” (Cumming, 2010, p. 158) for more than a century: syllable-based finger tapping (cf. the historic experiments by Brücke, 1871; Meyer, 1898; and Scripture, 1902).

Rhythm is a phenomenon that, as dancing shows impressively, can connect the beats or prominence structures that listeners perceive with their physical body movements. This cross-modal mechanism is used for the purpose of the present study. Note that it is controversial in phonetic studies to what extent and how exactly a participant’s individual finger tapping is time-aligned with the acoustic and perceptual boundaries of the syllables s/he perceives (see Wagner, 2008 and Cumming, 2010 for summaries). But, this potential problem is irrelevant to the present production experiment, because the present production experiment is not about *when* exactly relative to a syllable the participant’s finger hits the targeted object (such as the tabletop or the push-button of a technical device). Rather, the present experiment is about *how strongly* and *for how long* the participant’s finger hits the targeted object. Recent results of Samlowski & Wagner (2016) support the assumption underlying this method that there is a positive correlation between the perceived prominence of a syllable and the power and duration of the finger tapping



for that syllable (see also Parrell et al., 2014). Moreover, finger tapping (or “drumming” in the case of Samlowski & Wagner, 2016) proved to be a “a fast, intuitive and exact method” that yields fine-grained prominence patterns “for experienced and naive subjects alike” (Samlowski & Wagner, 2016, pp. 1,5). On this basis, we expect the following results to emerge for our production experiment.

- Irrespective of the pitch accent, the finger tapping for the accented syllable is always stronger and longer than for the two surrounding syllables.
- An early peak leads to an asymmetrical finger tapping pattern with an overall declining strength and duration. That is, the finger tapping is stronger and longer for the pre-accented syllable than for the post-accented syllable.
- A late peak results in an asymmetrical finger tapping pattern with an overall increasing strength and duration. That is, the finger tapping is weaker and shorter for the pre-accented syllable than for the post-accented syllable.
- A medial peak leads to a symmetrical finger tapping pattern. That is, the finger tapping is similarly weak and short for both the pre-accented and the post-accented syllable, and only strongly pronounced in terms of power and duration for the intervening accented syllable.

In addition, we analyze the finger-tapped sentences acoustically and expect that the results from Niebuhr & Pfitzinger (see Figure 2) will be replicated. This means that

- duration and intensity levels are higher for the accented syllable than for the two framing unaccented syllables.
- For the early peak, the duration and intensity levels of the pre-accented syllable are higher than those of the post-accented syllable.
- For the late peak, the duration and intensity levels of the pre-accent syllable are lower than those of the post-accented syllable.
- For the medial peak, the duration and intensity levels of the pre- and post-accented syllables are similarly low relative to those of the the accented syllable.

If the pitch-accent-specific micro-rhythm is not an integral part of the representation and production of pitch accents in German, but, for example, an epiphenomenon of another F<sub>0</sub>-related factor (perhaps of a different magnitude or onset of increased effort in the coordination of glottal and supraglottal gestures), then we still expect that the

speakers of the speech production experiment are able to tap the syllables of the utterances in parallel to their production. However, under these circumstances, we would expect the tapping to be either homogeneous in the relevant triplet of pre-accented, accented, and post-accented syllable, i.e. each of the three syllables should be tapped with the same duration and intensity; or we would expect that only the macro-rhythm of the utterances would manifest itself in the finger-tapping. In this case, the accent syllable would always be tapped stronger and longer than the two surrounding syllables, while the latter two do not differ, regardless of the category of the pitch accent on the accent syllable. Only if a pitch accent is represented, planned and executed as a sequence of specifically varying syllable prominences should this be reflected in a pitch-accent-specific finger tapping.

## **2. Method**

### **2.1 Participants**

The study is based on realizations of target sentences by four native German speakers. The number of speakers was small but deliberately chosen with respect to validity considerations. That is, instead of recruiting a large number of naive speakers and then trying to elicit the early, medial, and late peaks on target utterances in a consistent fashion by means of specifically tailored semantic-pragmatic context precursors (Niebuhr & Michaud, 2015; Kohler, 2017), we used experienced intonation researchers as participants who were well trained in the contrastive production and perception of early, medial, and late peaks. In pilot studies, this solution proved to be better for several reasons.

For example, it turned out to be problematic for naive speakers to produce target sentences with the intended (i.e. context-matching) intonation contours, while simultaneously tapping syllable by syllable the rhythm of these target sentences. Under this condition of double cognitive workload, naive participants produced, virtually independently of context precursors, the same nuclear pitch accent, namely the medial peak, which is considered the default pitch-accent category in German. Medial peaks account for 53 % of all nuclear pitch accents in the Kiel Corpus of Spontaneous Speech (Peters, 1999; Peters et al., 2005). In addition, the use of trained speakers prevented the sentences and, thus, the relevant pitch-accent patterns from being realized in an exaggerated enacted speech style, which typically occurs when naive speakers are asked to realize target sentences in expressive-emotional contexts (such

as those that would have been required for eliciting early and late peaks). The consequences of such a speech style for the external validity of the findings would have been difficult to estimate. Another reason that argued against the use of naive speakers was the internal validity of the data, more precisely the avoidance of circular reasoning. Previous studies showed that especially early and late peaks cannot be triggered and elicited with 100 % reliability by semantic-pragmatic contexts alone (Niebuhr, 2007c). However, it would have also been difficult to re-classify pitch-accent patterns with reference to acoustic or auditory criteria, because, as was pointed out by Pfitzinger & Niebuhr (2010), pitch-accent-specific micro-rhythms are in an acoustic and perceptual interplay with the alignment of F0 peaks. Thus, any post-hoc re-classification of pitch-accent patterns would have involved the risk that we either take into account this interplay and, in this way, make the actual object under investigation the criterion according to which we organize our sub-samples; or that we deliberately ignore this interplay and, thus, bias our samples and results. By using fewer speakers, but speakers who were trained in the natural production of early, medial and late peaks, these problems can be circumvented. That is, every pitch-accent pattern that these speakers produced and approved (after possible self-correction) was simply accepted as a successful rendering of the intended early, medial, or late peak.

The 4 speakers were between 25 and 41 years old. Two speakers were female and two were male. All 4 grew up in Northern Germany and were native speakers of Standard Northern German. All had normal hearing and speech skills and worked as PhD students or belonged to the scientific staff of the Dept. of Linguistics at Kiel University.

## **2.2 Reading Material**

The reading material consisted of 20 target sentences, which were realized in isolation without any pitch-accent supporting context. All target sentences had 6-7 syllables. The syllables were embedded in a syntactic structure of personal pronoun (she/they), verb, and noun (direct object), like, for example, “Sie war mal Malerin” (She was once a painter), “Sie leben in Sambia” (They live in Zambia), or “Sie haben Sonnenbrand” (They have a sunburn). The noun elicited the relevant nuclear pitch accent on its initial CV(C) syllable. Thus, the nuclear accent was always in the second half of each target sentence and occurred 2-3 syllables before the end of the sentence. The syllable preceding the noun was always unaccented and represented a so-called “weak form”. That is, it was either a particle, a preposition, or a grammatical suffix morpheme.

All target sentences were phonetically largely voiced, especially in the triplet of pre-accented, accented and post-accented syllable.

The set of 20 target sentences was completed by 6 syntactically and phonetically similar sentences, half of which were placed as dummies before and after the 20 target sentences. The frame of three initial and final dummy sentences was to avoid the prosodic list effects that occur “almost invariably” when speakers read sequences of isolated target sentences (Ladefoged, 2003, p. 7).

Overall, the reading material comprised 26 individual sentences.

### **2.3 Procedure**

The recording of the reading material took place for each of the 4 speakers in individual sessions. At the beginning of the session, the speaker was given the list of 26 sentences and asked to familiarize him/herself with the sentences for about 3-5 minutes.

Subsequently, the speaker was instructed to realize the sentences as separate (i.e. context-free) statements at a normal speaking rate and with a normal, clearly pronounced reading style and intonation (see Mixdorff & Pfitzinger, 2005 and Barbosa, 2015 for the prosodic characteristics of read as compared to spontaneous speech). Furthermore, the speaker received the information that there would be three rounds of recording. In the first round, each statement was to be produced with a medial peak as the nuclear pitch accent, in the second round with an early nuclear pitch accent, and in the third round with a late nuclear pitch accent. The pitch-accent elicitation order represented the subjective difficulty level with which the three accent categories can be produced (from less to more difficult). The order was constant across all 4 speakers (a complete permutation of the elicitation order would not have been possible with only 4 speakers anyway).

Then the speaker was told that, in addition to producing the sentence, s/he should in parallel tap the rhythm of each sentence with his/her index finger in syllable-by-syllable fashion. To that end, the experiment used an innovative device - a modified DJ drum pad AKAI MPD18 - that recorded the onset, offset, and power (reflected in the amplitude of the sound that it generates) of the speaker’s finger tapping in parallel to his/her speech signal, in a way similar way as did the “Sentograph” device that had been developed by Manfred Clynes in 1925 (see Kopiez & Lehmann, 2013). The drum-pad button, which was to be used for finger tapping, was on the top right corner of the device, where it was most convenient to reach for the speaker’s index finger. The button was also marked in red color.

The drum pad was placed on the table to the right of the speaker (all 4 speakers were right-handed). The speaker again received 3-5 minutes in order to familiarize him/herself with this simultaneous speaking-and-tapping task, using sentences of his/her choice from the list of 26. The speech tempo was initially slowed down significantly by this dual-task paradigm. However, at the end of this second familiarization phase, it had returned to the normal level of each speaker, i.e. about 4 syllables per second.

After the two familiarization phases, the actual speech recording began. The 26 sentences were presented to the speaker individually on a PC screen in font size 38 (Times New Roman, see Berger et al., 2016 for the advantages of the chosen typeface in speech-elicitation tasks). The speaker received the sentences in a constantly re-randomized order, i.e. an order that was always new in each round of recording and for each speaker. A separate re-randomization was also performed for the 6 dummy sentences. However, they remained consistently placed in triplets at the beginning and end of the list. The participant spoke into a gooseneck microphone (Sennheiser MD 421-U) that was positioned to the left of the screen. Figure 3 shows a sample photo of the recording setting.



Figure 3. Edited photograph showing the recording setting with the drum pad being placed to the right of the speaker and the gooseneck microphone being located to the left of the speaker, next to the screen in the center on which the 26 target/dummy sentences were displayed individually.

The recording was carried out in a self-paced fashion. That is, the speaker pressed a button and moved on to the next sentence on the screen whenever s/he was satisfied with the production of the current sentence (especially regarding its nuclear pitch-accent pattern). On average about 15 % of the sentences (4 out of 26), were realized several times by speakers, either because the speakers corrected a slip of the tongue, or because the nuclear pitch accent was not implemented well or clearly enough in the ears of the speaker. Some sentences were also re-read because of a miscoordination between the tapping of the finger and syllables in speech production.

#### **2.4 Data Analysis**

The sound files of the recording sessions were stored as stereo signals. On the left channel was the speech signal, and on the right channel was the time-aligned finger-tapping signal. The latter signal was recorded in the form of a sinusoid (with a constant frequency). The beginning and the end of the sinusoid marked those points in time at which the speaker had touched and released the button of the drum pad; and the amplitude of the sinusoid indicated the maximum power with which the button of the drum pad was pressed down by the finger. Figures 4(a)-(b) presents two examples of recorded stereo signals. The upper example shows the tapping-and-speaking signal of “Sie mögen Blumen sehr” (They like flowers very much) produced with a nuclear late-peak accent on “Blu-” [blu:]. The lower example shows the tapping-and-speaking signal of “Sie war mal Lehrerin” (She was once a teacher) produced with a nuclear medial-peak accent on “Leh-” [le:].

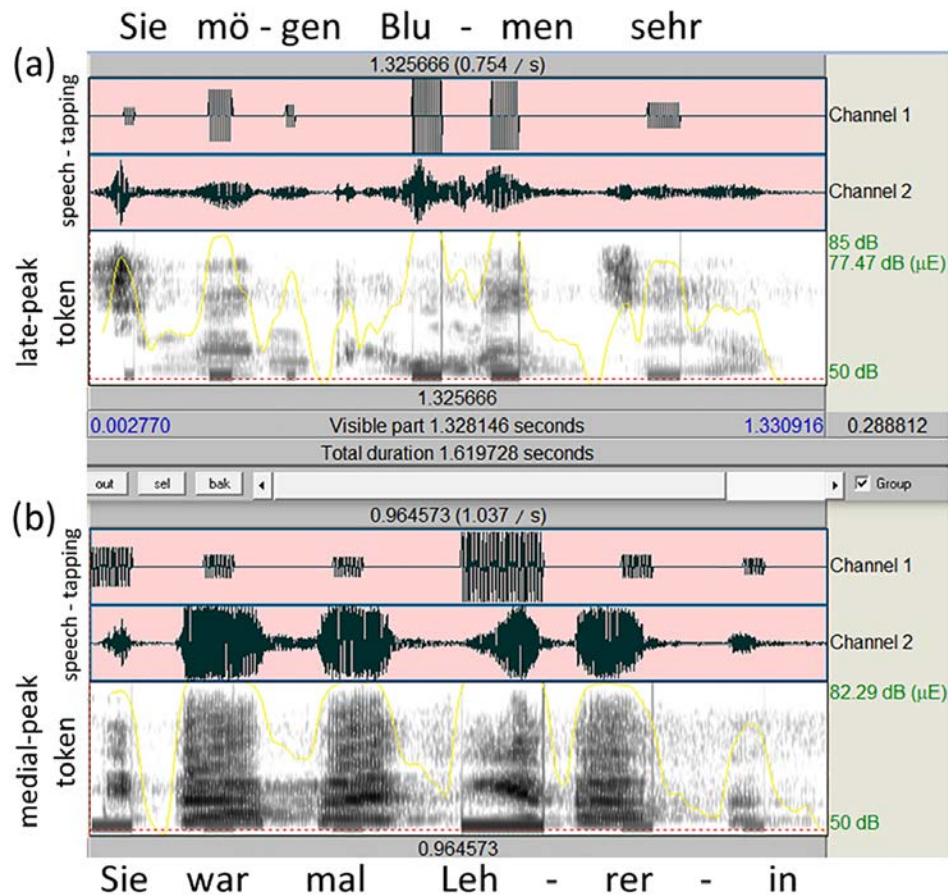


Figure 4. Recorded stereo files integrating the finger-tapping signal and the speech signal; (a) shows an example of a target sentence realized a late peak (Sie mögen Blumen sehr; they like flowers a lot), (b) shows an example of a target sentence realized a medial peak (Sie war mal Lehrerin; she was once a teacher).

The stereo signals of the 20 target sentences per participant were annotated with the Textgrid function in PRAAT (Boersma & Weenink 2018). Marked intervals were, firstly,

- the durations of the pre-accented, accented, and post-accented syllables, segmented on the basis of the acoustic speech signal (through a combined visual inspection of waveform and spectrogram representations);
- and the durations of the button presses on the drum pad, segmented on the basis of the acoustic sinusoid signal (through visual inspection of the waveform only).

The Textgrid files were used to measure (in ms) the durations of the syllables and button presses automatically by means of a PRAAT script. A similar PRAAT script was also used to determine the intensity maxima

of all annotated syllables and button presses (RMS, in dB, length of analysis window 40 ms, mean pressure subtracted). Prior to the intensity measurements, all speech files were intensity normalized (in Adobe Audition) by boosting the largest signal elongation to the maximum of the recording's dynamic range and then upscaling all other signal elongations proportionally. In this way, we removed differences due to speaker-individual loudness levels from the analysis. It was not possible to compensate, in a similar post-processing step, also for possible head or body movements of a speaker during the recording. However, given the constant contact of the speaker's arm and hand to the table and the drum pad, and because of the speaker's constant focus on the target sentences on the screen, each speaker maintained a fairly stable posture during the recording, and head movements were minimal. In relation to the mouth-to-microphone distance of about 70 cm, changes in this distance of a few centimeters represented a negligible and in any case randomly fluctuating variable.

Altogether 1,440 duration and intensity measurements were taken in the acoustic analysis, 720 for the speech data (240 per pitch accent), and 720 for the finger-tapping (i.e. drum pad) data.

### **3. Results**

For statistical analysis of the measurements, we conducted repeated-measures MANOVAs based on the two within-subjects factors Pitch-Accent Category (3 levels: early, medial, late) and Syllable (3 levels: pre-accented, accented, post-accented). The factor Speaker was included as a covariate. One MANOVA was conducted for each type of analyzed data, i.e. the speech data and the finger-tapping data. The two dependent variables were in both MANOVAs the measured duration and intensity values. Each MANOVA was supplemented by a pair of univariate repeated-measures ANOVAs. They were based on the same two within-subject factors as the MANOVA, but looked separately at the duration and intensity parameters. Moreover, multiple post-hoc comparisons (t-test series with Bonferroni corrections of significance levels) were carried out between the levels of the two within-subject factors in each ANOVA.

Results for the speech data are depicted in Figures 5(a)-(b). The figures show along the vertical axis the mean duration and intensity values, with which the speakers have realized the triplet of pre-accented (blue), accented (red), and post-accented syllable (green) in combination with each pitch-accent category. For example, for the early-peak category in



Figure 5(a), we can see that the post-accented syllable was on average 223.4 ms long (green). The pre-accented syllable was 246.4 ms long (blue) and thus slightly longer. The accented syllable (red) was the longest of the three with an average duration of 288.4 ms. Arrows in between the vertically displayed green, blue, and red mean values indicate significant differences between mean values ( $p < 0.05$ ), as determined in the multiple post-hoc t-test comparisons. Along the horizontal axis, it is shown how the mean values for each of the three syllables (pre-accented, accented, and post-accented syllable) changed over the pitch-accent categories of the early, medial, and late peak. For example, Figure 5(b) shows that the mean intensity of the pre-accented syllable (blue) decreases from the early peak (80.6 dB) through the medial peak (74.1 dB) to the late peak (68.8 dB). Analogous to the arrows along the vertical axis, continuous lines along the horizontal axis indicate significant differences between the PA categories ( $p < 0.05$ ). Dashed lines indicate that a difference between early and medial or medial and late peak is not significant.

The MANOVA on the speech data yielded significant main effects of Pitch-Accent Category ( $F[4,630]=77.5$ ,  $p < 0.001$ ) and Syllable ( $F[4,630]=63.3$ ,  $p < 0.001$ ), as well as a significant interaction of the two within-subject factors ( $F[8,1262]=114.6$ ,  $p < 0.001$ ). According to the separate univariate ANOVAs, the two dependent variables Duration (Pitch-Accent Category:  $F[2,316]=28.9$ ,  $p < 0.001$ ; Syllable:  $F[2,316]=37.0$ ,  $p < 0.001$ ; Pitch-Accent Category \* Syllable:  $F[4,632]=59.4$ ,  $p < 0.001$ ) and Intensity (Pitch-Accent Category:  $F[2,316]=19.7$ ,  $p < 0.001$ ; Syllable:  $F[2,316]=38.4$ ,  $p < 0.001$ ; Pitch-Accent Category \* Syllable:  $F[4,632]=45.4$ ,  $p < 0.001$ ) made comparably strong contributions to the MANOVA's overall main effects and their interaction. The additionally conducted multiple post-hoc comparisons yielded significant differences between all factor levels, except for the duration and intensity levels of the medial peak. For this pitch-accent category, there were no differences between the pre- and post-accented syllables on both sides of the accented one.

It can clearly be seen in the Figures 5(a)-(b) that the measured duration and intensity levels were consistently higher in the accented syllable, irrespective of pitch-accent category. More specifically, accented syllables were on average about 40-50 ms longer (284-292 ms) and had 6-8 dB higher intensity maxima (86-87 dB) than the surrounding pre- and post-accented syllables. In contrast, pre- and post-accented syllables differed strongly in their duration and intensity characteristics depending on the pitch-accent category with which they co-occurred. For early-peak

productions, pre-accented syllables were about 20 ms longer and 12 dB higher in intensity than their post-accented counterparts. An inversely asymmetrical duration and intensity pattern emerged for the late-peak productions. Here, it was the post-accented syllable that exceeded the mean duration and intensity levels of the pre-accented syllable; on average about 50 ms in duration and 8 dB in intensity.

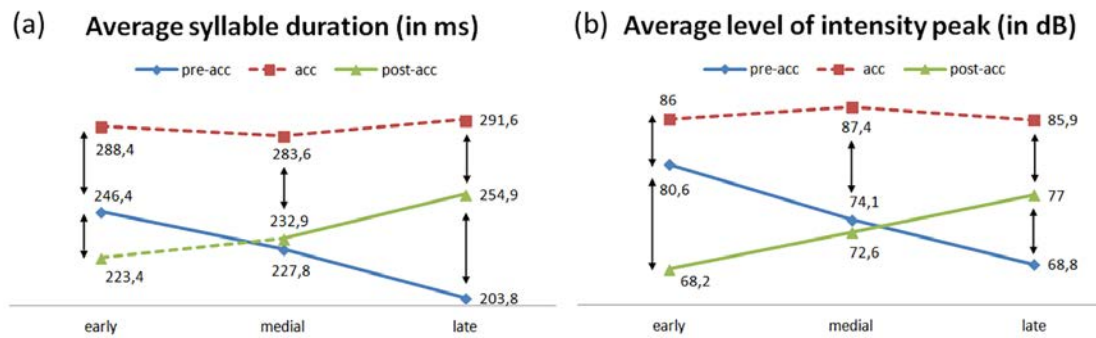


Figure 5. Average syllable durations (a) and average intensity maxima (b) across all 4 speakers, displayed separately for triplet of pre-accented, accented, and post-accented syllables produced in combination with early, medial, and late pitch accents. Continuous lines and vertical arrows show significant differences ( $p < 0.05$ ) between Pitch-Accent Category or Syllable conditions. Each data point represents 80 tokens.

The results summary of the finger-tapping data is provided in Figures 6(a)-(b). Like in Figure 5(a)-(b) above, the vertical axes in Figures 6(a)-(b) show mean value differences across the triplet of pre-accented, accented, and post-accented syllable (significant ones being marked by vertical arrows). The horizontal axes show how mean values vary across the triplet of early, medial, and late peak (with continuous lines indicating significant and dashed lines indicating not-significant differences between pitch-accent categories).

The overall results pattern closely resembles that of the speech data. The MANOVA yielded significant main effects of Pitch-Accent Category ( $F[4,630]=52.9$ ,  $p < 0.001$ ) and Syllable ( $F[4,630]=77.2$ ,  $p < 0.001$ ). Moreover, there was a significant interaction of Pitch-Accent Category and Syllable ( $F[8,1262]=95.1$ ,  $p < 0.001$ ). The supplementary ANOVAs showed that the two main effects and their interaction rely to a similar degree on both dependent variables, i.e. duration (Pitch-Accent Category:  $F[2,316]=33.5$ ,  $p < 0.001$ ; Syllable:  $F[2,316]=26.4$ ,  $p < 0.001$ ; Pitch-Accent Category \* Syllable:  $F[4,632]=82.7$ ,  $p < 0.001$ ) and intensity (Pitch-Accent Category:  $F[2,316]=40.4$ ,  $p < 0.001$ ;

Syllable:  $F[2,316]=38.6$ ,  $p<0.001$ ; Pitch-Accent Category \* Syllable:  $F[4,632]=66.2$ ,  $p<0.001$ ). Like for the speech data, the multiple post-hoc comparisons for the finger-tapping data yielded significant differences between all factor levels, except for the comparison of the medial peak's pre- and post-accented syllables. Their duration and intensity levels did not differ from each other.

Figure 6(a) shows that the finger-tapping durations are overall 30-60 ms shorter than the actual syllable durations (we assume that this is because our 4 speakers coordinated the entire downward and upward movement of their index finger with the speech syllables, not just the time during which the index finger pressed the button of the drum pad). Nevertheless, significant relative duration differences among the produced syllables emerged also in the finger-tapping data. That is, speakers pressed the button on the drum pad about 30 ms longer for the pre-accented or post-accented syllable, depending on whether they realized an early or a late pitch-accent peak, respectively. The longest button presses were consistently measured on the accented syllable, irrespective of pitch-accent type.

Likewise, Figure 6(b) shows that the power with which speaker pressed the button on the drum pad (i.e. the intensity of the sinusoid generated by the drum pad) was for all three pitch-accent categories significantly stronger on the accented syllable than on the two surrounding syllables. Besides this comparability of the three accents, the drum-pad button was pressed with greater power by the speakers (i.e. the drum pad generated a higher intensity level) on the pre-accented syllables in early-peak productions and on the post-accented syllables in late-peak productions.

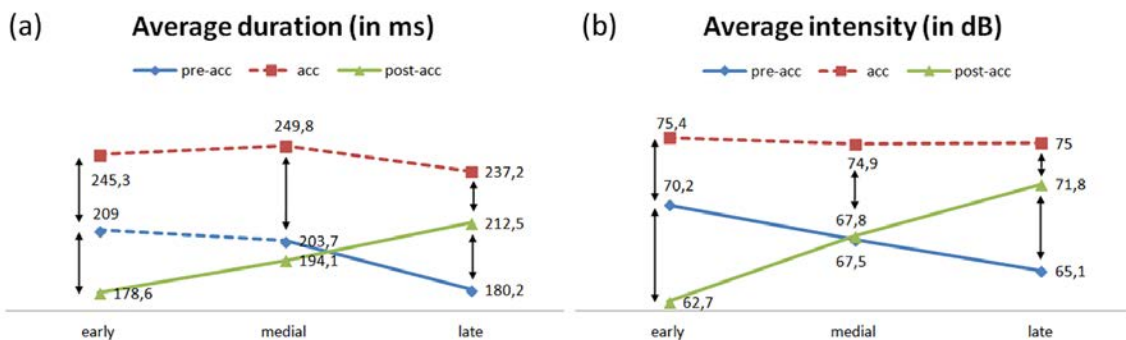


Figure 6. Average duration (a) and power (intensity) of button presses across all 4 speakers, displayed separately for triplet of pre-accented, accented, and post-accented syllables produced in combination with early, medial, and late pitch accents. Continuous lines and vertical arrows show significant differences ( $p<0.05$ ) between Pitch-Accent Category or Syllable conditions. Each data point represents 80 tokens.

Finally, note that the effect of the covariate Speaker came out highly significant in both MANOVAs (Acoustics:  $F[1,158]=120.8$ ,  $p<0.001$ ; Finger tapping:  $F[1,158]=98.6$ ,  $p<0.001$ ). That is, there were strong speaker-specific differences in how the target syllables were tapped and acoustically realized. Many of these differences were gender-related. For example, both syllable and finger-tapping durations were longer for the female than for the male speakers. Intensity levels were on average also higher for female speakers. In contrast, power levels of button presses were higher for the male than for the female speakers. The longer duration values measured for female speakers match with the longer word and sentence durations that were found for female speakers in other studies (across languages) and that are associated with females having a slower speaking rate than males (everything else being equal), see Van Borsel & De Maesschalck (2008) as well as Weirich & Simpson (2014) for a critical discussion of gender-specific speaking rates. That female speakers produced higher intensity levels is consistent with previous studies on different languages as well, see Klatt & Klatt (1990) and Hwa Chen (2007). Also the higher finger-tapping power of male speakers replicates findings in previous studies (Aoki et al., 2005).

In addition, there seemed to be some speaker-specific trade-offs in the extent to which duration and intensity/power differences are realized between pitch accents. One female speaker seemed to focus more on duration than on intensity when creating pitch-accent-specific differences in the triplet of pre-accented, accented, and post-accented syllable, whereas one male speaker seemed to prefer intensity over duration. However, based on only 4 speakers, we refrain from making any general statements about possible trade-offs between non-F0 parameters in pitch-accent production. It is interesting to keep in mind the possibility of such trade-offs for future studies, though.

#### **4. Discussion**

Niebuhr & Pfitzinger (2010) found in an acoustic analysis of nuclear German pitch accents that the three accent categories of early, medial, and late peak (nowadays established phonological categories across models of German intonation, Grice et al., 2005; Mayer, 1995; Kohler, 1991a) involve systematic changes in the duration and intensity levels of their coinciding pre-accented, accented, and post-accented syllables. With reference to the relevance of duration and intensity for perceived syllable prominence in German (see, for example, Kohler, 2008), Niebuhr & Pfitzinger called

these effects pitch-accent-specific micro-rhythms. The attribute “micro-” takes into account the fact that the actual macro-rhythm (i.e. what is typically meant by the term speech rhythm, cf. Kohler, 2009; Cumming, 2010) is, firstly, a matter of larger prosodic domains like the intonation phrase and, secondly, a matter that relies on the relatively strong perceptual prominences of accented and/or stressed syllables, not on relatively weak perceptual prominence differences between unstressed and/or unaccented syllables.

So far, Niebuhr & Pfitzinger’s idea of a pitch-accent-specific micro-rhythm was only supported by the fact that duration and intensity are prominence-related factors. There was no direct empirical evidence that the triplet of pre-accented, accented, and post-accented syllable actually forms a rhythmic pattern. Such evidence could, for example, have come from a perception experiment in which listeners judge the prominence levels of individual syllables. Previous studies showed that such judgments are possible to make for listeners with the necessary fine grading and for sequences of consecutive syllables (Jensen & Tøndering, 2005; Arnold et al., 2011). Nevertheless, the present study took an alternative approach, which was assumed to be still easier to implement and still more direct in reflecting speech rhythm: syllable-synchronous finger tapping. While speaking, participants pressed a button in a drum pad, once for each syllable they produced. These motor reflexes of speech rhythm were then analyzed in terms of the duration and power of the individual button presses (on the relevant syllable triplet) and additionally set in relation to the acoustic duration and intensity values of the coinciding syllables.

The acoustic analysis of 240 target sentences (80 tokens per pitch-accent category) replicated the findings of Niebuhr & Pfitzinger (2010) and, thus, was in accord with the hypotheses that were put forward on this basis in the present study.

- Duration and intensity levels are higher for the accented syllable than for the two framing unaccented syllables.
- For the early peak, the duration and intensity levels of the pre-accented syllable are higher than those of the post-accented syllable.
- For the late peak, the duration and intensity levels of the pre-accented syllable are lower than those of the post-accented syllable.
- For the medial peak, the duration and intensity levels of the pre- and post-accented syllables are equally low relative to those of the accented syllable.

Furthermore, and crucial for the present study, the pitch-accent-specific micro-rhythms derived from the acoustic prominence factors clearly also manifested themselves in the speaker's finger-tapping behavior. Thus, the corresponding hypotheses are supported.

- Irrespective of the pitch accent, the finger tapping for the accented syllable is always stronger and longer than for the two surrounding syllables.
- An early peak leads to an asymmetrical finger-tapping pattern with an overall declining strength and duration. That is, the finger tapping is stronger and longer for the pre-accented syllable than for the post-accented syllable.
- A late peak results in an asymmetrical finger-tapping pattern with an overall increasing strength and duration. That is, the finger tapping is weaker and shorter for the pre-accented syllable than for the post-accented syllable.
- A medial peak leads to a symmetrical finger tapping pattern. That is, the finger tapping is similarly weak and short for both the pre-accented and the post-accented syllable, and only strongly pronounced in terms of power and duration for the intervening accented syllable.

Expressed in prominence patterns, the early peak seems to be characterized by a slight increase in prominence towards the accented syllable, followed by a strong drop in prominence after the accented syllable. In contrast, the late peak is associated with a strong increase in prominence towards the accented syllable and only a slight prominence decrease after the accented syllable. In other words, for early peaks, two approximately equally strong prominent syllables follow a weakly prominent syllable, and for late peaks, a weakly prominent syllable is followed to two approximately equally prominent syllables. The medial peak is characterized by a strong prominence contrast between the pre- and post-accented syllables and the accented syllable in the center of the triplet that clearly stands out against its two neighbors.

Initial experimental data from Niebuhr & Pfitzinger (2010) and Niebuhr (2011) suggest that these pitch-accent-specific micro-rhythms are perceptually relevant. This applies both to the identification of the pitch accents and to the perception of their communicative meanings. This perceptual relevance is not sufficiently represented in current intonation models and phonologies as they are all focused on F0 alone.

Note, however, that there is an interesting parallel between the pitch-accent-specific micro-rhythms determined here and the representations of the early, middle, and late peaks in the major autosegmental-metrical (AM) model of German intonation, GToBI (Grice et al., 2005). The early peak is conceptualized in GToBI as H+L\* (or H+!H\*), the medial peak as H\*, and the late peak as L\*+H. That is, the position of the leading or trailing tone relative to the starred tone is the same as the position of the more prominent pre- or post-accented syllable relative to the accented syllable in the pitch-accent-specific micro-rhythms. H\* does not have a training or leading tone in GToBI and neither did we find a prominent pre- or post-accented syllable for this pitch accent category. However, in GToBI (as in the original AM framework of Pierrehumbert, 1980), the leading and trailing tones are not separately associated with (pre- or post-accented) syllables, and they also need not coincide with particular syllables or syllable boundaries. Thus, in order to explain and represent pitch-accent-specific micro-rhythms by means of trailing or leading tones in the AM framework, auxiliary phonological concepts such as the secondary-association concept would be required (Prieto et al., 2005); and even on this basis the complex interaction of F<sub>0</sub>, duration, and intensity in the signaling of pitch accents can probably not be adequately and fully covered. For example, the F<sub>0</sub> peaks themselves can show also considerable variation in peak shape and alignment (Niebuhr, 2007a,c), and trailing or leading tones cannot represent both F<sub>0</sub> shape characteristics and pitch-accent-specific rhythm characteristics at the same time. In addition, there are the indications in the present data for speaker-specific trade-offs in the extent to which duration and intensity/power differences are realized between pitch accents. Except for the fact that tonal targets like leading and trailing tones are only two-dimensional descriptor units, which are unable to represent continuous prosodic variation beyond the F<sub>0</sub> alignment and scaling dimensions (syllable association is a third but categorical or binary variable), modeling duration and intensity interactions by means of F<sub>0</sub>-related units seems in general to be at best a preliminary solution; provided that these interactions (trade-offs) are supported and further substantiated by follow-up studies with a larger speaker sample.

Overall, empirical evidence suggests that F<sub>0</sub> on the one hand and syllable duration and intensity (i.e. patterns of prominence or rhythm) on the other are connected but conceptually independent signaling systems of pitch accents. In combination, these signaling systems form what

Ward & Gallardo (2017) call a “prosodic construction”, i.e. a coherent multiparametric configuration of prosodic features (see the corresponding special session at the International Congress of Phonetic Sciences, ICPhS, Melbourne, 2019: <http://www.cs.utep.edu/nigel/pconstructions/icphs-configs.html>). The system of syllable duration and intensity does not seem to be an epiphenomenon of a F0 system controlled by tonal targets and their primary or secondary association.

An alternative framework may be more suitable for explaining and modeling the present findings: the perception-based Contrast Theory of Niebuhr (2007b, see also Niebuhr, 2013). The Contrast Theory is based on similar ideas and concepts as the Tonal Center of Gravity (TCoG) theory of Barnes et al. (2012). It too showcases the complex interplay of seemingly disparate aspects of the acoustic signal in the domain of perception, but its focus is more strongly on perceived prominence. The Contrast Theory’s basic assumption is that all different realization strategies that are observed for pitch-accent categories at the level of acoustics boil down to making some sections of F0 peaks or movements stand out more prominently in perception than others. For differentiating between early and medial peaks, for example, the final low F0 section and the middle high F0 section of the rising-falling F0 peaks must achieve maximum prominence respectively. The typical alignment differences between early and medial peaks (see Figures 1-2), according to the Contrast Theory, are so widespread across speakers and languages because they represent the simplest way to achieve this prominence goal, namely by moving the corresponding section of the F0 peak into an area in which its prominence is inherently enhanced by a high acoustic energy level: the accented vowel.

In the Contrast Theory, the duration and intensity differences between the pre- and post-accented syllables would only be an additional strategy to make especially the early and late peak categories phonetically and phonologically more dissimilar. Unlike the medial peak, both the early and the late peak are prosodically constructed around a low-pitched prominence. Therefore, both pitch accents additionally have a secondary high-pitched prominence. While in the early peak pattern this secondary high-pitched prominence precedes the major low-pitched prominence, it follows the major low-pitched prominence in the late peak pattern. The similarity of this concept to the GToBI representations H+L\* and L\*+H is obvious, but the essential difference between the GToBI representation and the conceptualization of the pitch accents in the Contrast Theory



is that the latter theory views pitch accents as multiparametric prosodic configurations (“constructions”) that are inseparably constituted of a pitch Gestalt and a prominence Gestalt (Niebuhr, 2007c, 2013). In addition, the Contrast Theory sees the phonologically distinctive elements of all three pitch accents not in the pitch Gestalt but in the prominence Gestalt.

The Contrast Theory also explains why, in speech production, early, medial, and late peaks show specific F<sub>0</sub>-peak shape and range properties that are also relevant in pitch-accent perception. For example, characteristic of the early peak is a shallower rise towards the F<sub>0</sub> peak maximum (Niebuhr, 2007a). In the case of the late peak, it is an expanded F<sub>0</sub> peak range that is characteristic of this category (Niebuhr & Ambrazaitis, 2006; Niebuhr, 2007c). Both strategies are also suitable to further enhance the secondary high-pitched prominence before or after the major low-pitched prominence on the accent syllable. For the medial peak, it is characteristic and perceptually advantageous when both the rising and the falling F<sub>0</sub> slope are steep. This can be understood as the avoidance of any secondary high-pitched prominences on the surrounding syllables.

The Contrast Theory, however, is not yet a fully developed intonation model. Nevertheless, it shows, in a similar way as the TCoG theory of Barnes et al. (2012), a possible way of reducing and understanding the complex acoustic signaling of pitch accents in terms of a manageable number of perceptual variables, also with respect to a trade-off between acoustic parameters, for which some indications were found in the present study. Additional trade-offs of a different kind are included in the Gestalt-based Functional Contour superposition model (SFC) of Bailly & Holm (2005) and its further developed variant, the Variational Prosody Model (VPM) of Gerazov et al. (2018). These AI-driven models take into account the possibility that each prosodic configuration at each point in time reflects not a single communicative function, but a number of simultaneously coded functions. The SFC and VPM models use a separate set of (hyper)parameters that represent how prominently each communicative function is present in the speech coded at each point in time, i.e. how strongly the corresponding signal features are pronounced by the speaker. In defining these signal features and the meaningful Gestalt-like signal configurations that they form, the SFC and VPM models go beyond the Contrast Theory and the TCoG theory in that they include also visual features, i.e. a speaker’s mimic, head, and body movements, in the overall signal configuration (which is therefore not a

mere prosodic configuration but a multi-modal signal configuration). The rich and sophisticated models like SFC and VPM already are, the less insightful they are when it comes to understanding and explaining the actual links between speech production and speech perception, i.e. why certain prosodic and visual signals are used and how the listener's ear determines their configurational combination and inter-individual as well as cross-linguistic interactions. To that end, computer-based models like SFC and VPM will have to be combined with psycho-phonetic concepts as they are represented in the Contrast Theory and the TCoG theory.

While this is still a future task, the success of SFC and VPM in modeling empirical data beyond the auditory modality – and beyond individual syllables and even rhythmic feet – further stresses the fact that pitch accents are no simple F0 patterns, and certainly no individual local target points. The present study was only a small contribution to emphasizing the actual nature of pitch accents as coherent configurations of multiple prosodic features. Many more studies have to follow, especially those with a comparative cross-linguistic perspective, as the early, medial, and late peaks are melodic elements that also occur in many other languages (but often with different communicative functions). This includes internationally used and taught languages such as English (Kleber, 2006) and Scandinavian languages such as Swedish (Ambrazaitis et al., 2012) and Icelandic (Dehé, 2010).

In addition, follow-up studies should investigate, in a cross-linguistic perspective, to what extent the micro-rhythms outlined here are actually really more “micro” than “macro” in terms of perceptual prominences, given that the pitch-accent-specific duration and intensity differences measured between pre- and post-accented syllables are not much smaller as those measured between each of these syllables and the accented one. As was mentioned above, Jensen & Tøndering (2005) and Arnold et al. (2011) have tested and shown that listeners can apply 31-point scales to represent in a sensible way perceived prominence differences between syllables in stimulus sentences. Such scales seem sensitive enough to quantify (i) how big or small the prominence gap is between the accented syllable and its pre- and post-accented syllables (especially pre-accented syllables in early-peak and post-accented syllables in late-peak contexts), (ii) how much the pre- and post-accented syllables vary in their perceived prominence levels depending on the pitch-accent category, and (iii) how much the prominence levels of pre- and post-accented syllables increase for specific-pitch accents relative

to other adjacent unaccented syllables. These experiments are currently ongoing and will be followed by speech-production experiments with a larger speaker sample before we turn to the questions of prosodic modeling outlined above.

## 5. Acknowledgments

I would like to thank Allard Jongman and the other reviewers of my paper for their constructive and insightful comments. I am also greatly indebted to Lene Boysen for her excellent work and help in collecting and analyzing the data of the present study (as part of her BA thesis at Kiel University, 2015). Furthermore, my thanks are due to Nigel Ward, Gérard Bailly, Yi Xu, Benno Peters, and Ernst Dombrowski as well as to all participants of the Symposium on Speech and Language on the occasion of Ocke-Schwen Bohn's 65th birthday at Aarhus University (May 2018) for the many inspiring discussions with me on the nature and investigation of multiparametric prosodic configurations. Last but not least, I am indebted to the editors of the Festschrift for the honor and the opportunity to make a small contribution to their collection of papers and for their motivating and patient correspondence with me.

## References

- Ambrazaitis, G., J. Frid, & G. Bruce. (2012). Revisiting Southern and Central Swedish intonation from a comparative and functional perspective. In: O. Niebuhr (ed.), *Understanding Prosody – The role of context, function, and communication* (pp. 138-158). Berlin/New York: de Gruyter.
- Aoki, T., S. Furuya, & H. Kinoshita. (2005). Finger-Tapping Ability in Male and Female Pianists and Nonmusician Controls. *Motor Control* 9, 23-39.
- Arnold, D., P. Wagner, & B. Möbius. (2011). Evaluating different rating scales for obtaining judgments of syllable prominence from naive listeners. *Proc. 17th International Congress of Phonetic Sciences, Hong Kong, China*, 252-255.
- Bailly, G. & B. Holm. (2005). SFC: a trainable prosodic model. *Speech Communication* 46, 348-364.
- Barbosa, P. A. (2015). Temporal parameters discriminate better between read from narrated speech in Brazilian Portuguese. *Proc. 18th International Congress of Phonetic Sciences, Glasgow, UK*, 1053-1057.
- Barnes, J., A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux. (2012). On the nature of perceptual differences between accentual peaks and plateaux. In O. Niebuhr

- (Ed.), *Understandig prosody – The role of context, function and communication* (pp. 93-118). Berlin/New York: de Gruyter.
- Berger, S., C. Marquard, & O. Niebuhr. (2016). INSPECTing read speech : How different typefaces affect speech prosody. *Proc. 8th International Conference of Speech Prosody, Boston, USA*, 513-517
- Boersma, P. & D. Weenink. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.39, retrieved 3 April 2018 from <http://www.praat.org/>
- Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund: Gleerup.
- Brücke, E. (1871). *Physiologische Grundlagen der neuhochdeutschen Verskunst*. Vienna: Gerold.
- Cumming, R. E. (2010). *Speech rhythm: the language-specific integration of pitch and duration*. PhD thesis, Downing College, UK. <https://doi.org/10.17863/CAM.16499>.
- Dehé, N. (2010). The nature and use of Icelandic prenuclear and nuclear pitch accents: Evidence from F0 alignment and syllable/segment duration. *Nordic Journal of Linguistics* 33, 31-65.
- Gartenberg, R. & C. Panzlaff-Reuter. (1991). Production and perception of f0 peak patterns in German. *AIPUK* 25, 29-115.
- Gerazov, B., G. Bailly, & Y. Xu. (2018). A Weighted Superposition of Functional Contours model for modelling contextual prominence of elementary prosodic contours. *Proc. 19th International Interspeech Conference, Hyderabad, India*, 1-5.
- Grice, M., S. Baumann & R. Benz Müller. (2005). German Intonation in Autosegmental-Metrical Phonology. In: Jun, Sun-Ah (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.
- Gussenhoven, C. (1999). Discreteness and Gradience in Intonational Contrasts. *Language and Speech* 42, 283-305.
- Hwa Chen, S. (2007). Sex Differences in Frequency and Intensity in Reading and Voice Range Profiles for Taiwanese Adult Speakers. *Folia Phoniatrica et Logopaedica* 59, 1-9.
- Jensen, C. & J. Tøndering. (2005). Choosing a Scale for Measuring Perceived Prominence. *Proc. 5th International Interspeech Conference, Lisbon, Portugal*, 2385-2388.
- Kleber, F. (2006). Form and function of falling pitch contours in English. *Proc. 3rd International Conference of Speech Prosody, Dresden, Germany*, 61-64.
- Kohler, K. J. (1991a). A model of German intonation. *AIPUK* 25, 295-360.
- Kohler, K. J. (1991b). Terminal intonation patterns in single-accent utterances in German: phonetics, phonology and semantics. *AIPUK* 25, 117-185.
- Kohler, K. J. (1991c). The interaction of fundamental frequency and intensity in the perception of intonation. *Proc. 12th International Congress of Phonetic Sciences, Aix-en-Provence, France*, 186-189.

- Kohler, K. J. (2005). Timing and functions of pitch contours. *Phonetica* 62, 88-105.
- Kohler, K. J. (2008). The perception of prominence patterns. *Phonetica* 65, 257-269.
- Kohler, K. J. (2009). Rhythm in Speech and Language – A New Research Paradigm. *Phonetica*, 66, 29-45.
- Kohler, K. J. (2017). *Communicative Functions and Linguistic Forms in Speech Interaction* (Cambridge Studies in Linguistics 156). Cambridge: Cambridge University Press.
- Klatt, D. H. & L. C. Klatt. (1990). Analysis, synthesis and perception of voice quality variations among male and female talkers. *Journal of the Acoustical Society of America* 87, 820-856.
- Kopiez, R. & A. C. Lehmann (2013). Der Sentograph und seine Anwendung in der musikalischen Ausdrucksforschung – Erkenntnisse aus einer Einzelfallstudie. In: V. Busch, K. Schlemmer, C. Wöllner (eds), *Wahrnehmung – Erkenntnis – Vermittlung. Musikwissenschaftliche Brückenschläge. Festschrift für Wolfgang Auhagen zum sechzigsten Geburtstag* (pp. 121-130). Hildesheim: Olms.
- Ladefoged, P. (2003). *Phonetic Data Analysis. An Introduction to Fieldwork and Instrumental Techniques*. Oxford: Blackwell.
- Mayer, J. (1995). *Transcription of German Intonation: The Stuttgart System*. Technischer Bericht, Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung.
- Meyer, E. (1898). Die neueren Sprachen (Vol. 6).
- Mixdorff, H. & H. R. Pfitzinger. (2005). Analysing fundamental frequency contours and local speech rate in map task dialogs. *Speech Communication* 46, 310-325.
- Nash, R. & A. Mulac. (1980). The intonation of verifiability. In L. R. Waugh & C. H. van Schooneveld (eds), *The melody of language: Intonation and prosody* (pp. 219-241). Baltimore: University Park Press.
- Niebuhr, O. (2006). The role of the accented-vowel onset in the perception of German early and medial peaks. *Proc. 3rd International Conference Speech Prosody, Dresden, Germany*, 109-112.
- Niebuhr, O. & G. Ambrazaitis. (2006). Alignment of medial and late peaks in German spontaneous speech. *Proc. 3rd International Conference Speech Prosody, Dresden, Germany*, 161-164.
- Niebuhr, O. (2007a). The signalling of German rising-falling intonation categories – The interplay of synchronization, shape, and height. *Phonetica*, 64, 174-193.
- Niebuhr, O. (2007b). Categorical perception in intonation: a matter of signal dynamics? *Proc. 7th International Interspeech Conference, Antwerp, Belgium*, 109-112.

- Niebuhr, O. (2013). The acoustic complexity of intonation. In E-L. Asu, & P. Lippus (eds), *Nordic Prosody XI* (pp. 25-38). Frankfurt: Peter Lang.
- Niebuhr, O. & A. Michaud. (2015). Speech data acquisition: the underestimated challenge. *KALIPHO (Kieler Arbeiten in Linguistic und Phonetik)* 3, 1-42.
- Parrell, B., L. Goldstein, S. Lee, & D. Byrd. (2014). Spatiotemporal coupling between speech and manual motor actions. *Journal of Phonetics* 42, 1-11.
- Peters, B. (1999). Prototypische Intonationsmuster in deutscher Lese- und Spontansprache. *AIPUK* 34, 1-173.
- Peters, B., K. J. Kohler, & T. Wesener. (2005). Melodische Satzakkentmuster in prosodischen Phrasen deutscher Spontansprache. *AIPUK* 35a, 7-54.
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. PhD Diss, MIT, USA.
- Prieto, P., M. D'Imperio, & B. Gili-Fivela. (2005). Pitch accent alignment in Romance: primary and secondary associations with metrical structure. *Language and Speech* 48, 359-396.
- Samlowski, B. & P. Wagner. (2015). Promdrum – exploiting the prosody-gesture link for intuitive, fast and finegrained prominence annotation. *Proc. 8th International Conference of Speech Prosody*, Boston, USA, 1-5.
- Scripture, E. W. (1902). *The Elements of Experimental Phonetics*. New York: Charles Scribner's Sons.
- Wagner, P. (2008). *The rhythm of language and speech: Constraining factors, models, metrics and applications*. Habil. thesis (Habilitationsschrift), University of Bonn, Germany.
- Van Borsel, J. & D. De Maesschalck. (2008). Speech rate in males, females, and male-to-female transsexuals. *Clinical Linguistics & Phonetics* 22, 679-685.
- Ward, N. G. & P. Gallardo. (2017). Non-Native Differences in Prosodic-Construction Use. *Dialogue and Discourse* 8, 1-30.
- Weirich, M. & A. P. Simpson. (2014). Differences in acoustic vowel space and the perception of speech tempo. *Journal of Phonetics* 43, 1-10.



# **MORPHOLOGY & SYNTAX**

Handling editor: Anne Mette Nyvad





## **A Sound Approach to Text Processing: Between Experiments and Experience**

Laura Winther Balling  
Copenhagen Business School

### **Abstract**

The area of text processing is an interesting one both for psycholinguists attempting to understand how language works and for those who focus on making texts accessible. However, understanding a text is a complex process that involves several different aspects, of which I discuss three main ones: comprehension, processing speed and ease, and reader reception, along with ways to study these aspects based on the reader and the text. A main argument in this chapter is that experiments should attempt to take these multiple aspects into account, and I describe two approaches that I have used to do so, and discuss their pros and cons. Based on this, some avenues of further research are outlined.

### **1. Introduction**

Since working with Ocke-Schwen Bohn as his PhD-student, I have shared with him a devotion to understanding language through experiments. Experiments help us understand language processing on many levels, including speech perception, word recognition, and sentence processing. However, when we approach the level of text processing, we may be reaching the limits of what experiments can do, or at least what experiments can do alone: strict and sometimes artificial experimental manipulations and standard behavioural measures like response times provide a mechanistic and therefore too limited view of text processing. My argument in this chapter is that a sound approach to text processing should attempt to take

into account both what experiments tell us and what readers experience, attempting to bridge the gap between the language processing that happens in the psycholinguistic lab and the language processing that happens “in the wild”.

The study of language processing in the psycholinguistic lab has traditionally been based on closely matched items in factorial designs, for instance comparing two different types of dative constructions using the same lexical material (e.g. Balling & Kizach, 2017). Stimuli appear without context and are generally constructed by the researchers rather than sampled from actual language use. In addition, the tasks that participants perform in the lab are often substantially different from language processing in the wild, including tasks such as acceptability ratings, which require the inclusion of ungrammatical sentences that we generally would not encounter in real written discourse; self-paced reading, which only gives access to one word at a time; and lexical decision, where we read or listen to both real words and constructed nonsense words and decide which are which. There are many advantages to the experimental approach, particularly in the ability to isolate and understand a particular aspect of processing, but in the case of text processing, this isolation arguably comes with too severe drawbacks.

In the following, I will discuss the problem of the gap in the study of text processing between language processing in the lab and language processing in the wild and ways in which it may be, if not bridged, then at least taken into account. The point of departure is an account in section 2 of what I see as the primary reasons to study text processing, followed by an outline in section 3 of three main aspects of text processing and how these may be studied. Section 4 focuses on attempts to bridge the gap between the lab and the wild, including both naturalistic experiments and other possible avenues of research. I focus here on the processing of written text, but some of the issues also apply to the study of oral discourse processing.

## **2. Why study text processing?**

There are at least two main reasons for studying text processing: One is the general linguistic one of wanting to know how language works, with the specific psycholinguistic focus on understanding the relevant cognitive mechanisms. The level of text processing is particularly interesting – and complicated – in drawing on (the combination of) many other levels of processing. In this sense, the study of text processing is an attempt to understand the puzzle of how humans manage to read several hundred words

a minute while having to perform a range of different and in themselves rather complicated processing tasks, including recognising letters and combining them to form words, accessing the lexical representations of each of those words in a vocabulary of up to 150,000 words (Harley, 2008: 7), integrating them in phrasal and syntactic structures, and processing the discourse relations between them. Here, the influential Lexical Quality Hypothesis (Perfetti, 2007; Perfetti & Hart, 2002) argues that the quality and accessibility of the lexical representations is particularly important, but the other levels of processing and the coordination between them should of course not be overlooked.

The second reason is more practically oriented, namely the aim to improve text comprehension. This is for instance expressed in the Plain Language movement (see for instance *Federal Plain Language Guidelines*, 2011 for the US; Kjærgaard, 2016 for a discussion of the movement in the Nordic countries), as well as writing guides (e.g. Jacobsen & Jørgensen, 1992; Rozakis, 2000; Sorenson, 2010; Williams, 2005) and language policies, of which for instance those of the Danish courts (Kjærgaard, 2011a, 2011b, 2012) and Denmark's taxation authority have been carefully studied (Kjærgaard, 2015). These publications are not concerned with text processing *per se*, but the aim to improve comprehension and make texts more accessible should involve (psycholinguistic) considerations of how texts are actually processed and comprehended when read by different users.

### **3. Three main aspects of text processing**

A main challenge when studying text processing is that it covers many different facets, not all of which are directly measurable. An obvious main issue – and in some respects in fact *the* main issue – is comprehension in the sense of understanding the contents of the text. This is an everyday activity for most literate people, but it remains hard both to define and to measure.

With respect to *defining* comprehension, it is important to consider both the depth and breadth of comprehension. The balance between these depends crucially on the purpose of reading – is it to look for a detail, get an overall sense of the topic, experience a narrative, or something else entirely – and on the motivation for reading, which could be interest or obligation or somewhere in between. Alternatively, we may think about purpose and motivation in terms of the three different types of goals for the reading of a given text outlined by Graesser, Singer, and Trabasso (1994):

default which is the goal of constructing an adequate situation model (see more on this concept below) for the text and is default in the sense of being generally applicable to most, if not all types of text processing; genre-based goals which are constrained by the type or genre of the text in question; and idiosyncratic goals which come close to what I refer to as the motivation for reading. In connection with the purposes and motivations for text comprehension, learning is an obvious issue in many contexts, but again, this is a phenomenon that comes in many varieties.

The variety of purposes and motivations for reading, and their influence on the process, mean that we may have to accept that text comprehension cannot be defined in isolation from the purpose of and motivation for the reading activity. This in turn becomes a methodological challenge in the empirical study of text processing, in that we must somehow define the level of reading expected of our participants, either by explicitly describing this or through specific task demands.

When it comes to *measuring* comprehension, there are multiple options which again depend on or may define the purpose of reading. An obvious choice in both text and sentence processing experiments (Bråten & Anmarkrud, 2013; Cop, Drieghe, & Duyck, 2015; Pham & Sanchez, 2018; Veldre & Andrews, 2018, to name but a few, from different domains) is multiple-choice or other forced-choice questions; this is a relatively quick and straightforward approach but potentially measures only rather superficial comprehension and relatively passive knowledge. More in-depth processing may be indexed by asking open questions (Balling, 2018) or requiring readers to recall the contents of texts they have read, and then scoring that recall for how many and how important ideas are recalled (e.g. Spyridakis & Isakson, 1998). Apart from the obvious drawback of being a more time-consuming research process, scoring of the replies relies to some extent on interpretation, particularly when it comes to the distinction between important and less important ideas.

In addition to the more generic objections to multiple-choice comprehension assessment, there is also evidence that different measures of comprehension measure different aspects of comprehension or different aspects of the text structure: Kintsch & Yarborough (1982) found that readers who had encountered texts with “good”, conventional rhetorical structure performed better on questions about topic and main ideas of that text than those who read texts with “bad”, unconventional rhetorical structure, while cloze test performance remained the same irrespective of rhetorical structure. It seems that overall rhetorical structure supports the

macro-processing indexed by recall rather than micro-processing indexed by cloze test performance. A further, extreme example of the difference between micro- and macro-level processing is that of quoting the Quran in Arabic without (otherwise) knowing Arabic (Kintsch, 1998: chapter 9), where a certain micro-level of learning is “measured” by repetition, but certainly not the kind of macro-level comprehension and learning that we are usually interested in when studying text processing.

A second important aspect of text processing is the ease and speed of the processing, which is partially an index of ease of comprehension but also depends on the efficiency of decoding as well as the more general language skills of the participants<sup>1</sup>, which in turn vary with their reading proficiency. Experimental methods, including both standard behavioural measures like word reading time in self-paced reading and more advanced measurements like eye movements, are ideal for measuring speed, but cannot in themselves help us distinguish between text comprehension and decoding processes. To draw that distinction, there are broadly speaking two approaches: One is to measure decoding skills through one or typically more auxiliary tasks (see for instance the broad range of tasks used by Kuperman & Van Dyke, 2011). The other option is to conduct experiments with groups of participants with presumably similar decoding skills which is the typical approach when we run experiments with college students (e.g. most of the studies referenced in this text). However, even in such relatively homogeneous samples, we may well see substantial variation in decoding and general language processing skills, and the generalisability to “reading” in the abstract – to the extent that such a thing even exists, as discussed above – becomes questionable. We should also note that, although there is a correlation between text processing skills and general language skills, there is a substantial group of readers that read texts more poorly than we would expect based on their general language skills, as indicated by word reading ability (Perfetti & Adlof, 2012).

A third aspect of text processing, which is usually overlooked in the literature that focuses on text and discourse processes, but emphasised when the focus is on Plain Language and related approaches, is the reader’s reception of the text and their resulting image of the sender. The methods for studying this are generally decidedly not experimental, but include comparative text analysis (e.g. Kjærsgaard, 2011b), qualitative

<sup>1</sup> Because my focus is on text processing, I take the liberty of conflating these two, potentially quite different issues of decoding and general language skills, though in other contexts, it may be highly relevant to distinguish between them.

interviews (e.g. Garwood, 2014), and questionnaires (e.g. Kjærgaard, 2015). A particularly interesting approach in this field is the use of think-aloud protocols, a method that has also been used as a quasi-experimental paradigm in cognitive psychology (see e.g. Ericsson & Simon, 1980), to study the reading and reception of texts qualitatively (Kjærgaard, Gravengaard, Dindler, & Hjuler, 2018; Schriver, 1991).

One way of conceptualising these three different aspects of text processing – comprehension, processing speed and reception – is in relation to the general model of discourse processing that originates with van Dijk & Kintsch (1983) with later developments by Kintsch (1998) and others. This model includes five levels: surface code, text base, situation model, genre and rhetorical structure, and pragmatic communication. The surface code is the explicit lexical and syntactic contents of the text, which feed into the text base which is the reader's representation of the core semantic units of the text. The third level is the situation model which is the reader's representation of both the explicit contents of the text and the inferences drawn based on the text and existing knowledge. The fourth and fifth levels, like the first level, are oriented more towards the text/discourse than towards the receiver, with the fourth level being genre and rhetorical structure of the text or discourse, at various levels of granularity, and the fifth level the pragmatic communication, i.e. the message that the sender is trying to convey with the text or discourse.

This is not the only possible model of discourse processing, but it is one that is relatively broadly accepted and which provides a useful framework for understanding the elements involved in discourse and text processing. It also offers meaningful explanations for the different ways text processing may be impeded or even break down (Graesser & Millis, 2011). In relation to the aspects described here, speed and efficiency relate mostly to the first two levels of the model, while reader reception concerns levels 4 and 5, with comprehension understood as making sense of the text drawing on all levels but with a focus on level 3.

While the preceding parts of this section have focused on the readers and the reading process, an obvious further issue to consider is properties of texts. Here, an extensive literature has attempted to formulate readability indices that can measure the difficulty of a text, i.e. measure how difficult a reader will find a text to comprehend based on properties of that text. In a Danish context, the most common index is LIX (Björnson, 1968, cited by Klare, 1984), while in the US the most commonly used indices seem to be Flesch Reading Ease scale and the Flesch-Kincaid grade level scale that

was derived from that (Flesch 1943, Kincaid et al. 1975, both cited by Bailin & Grafstein, 2016). Most readability indices rely on some combination of word length or frequency with sentence length; for an overview see Bailin & Grafstein (2016), who also discuss many potential criticisms of standard readability measures. The major issue is the reliance on word and sentence lengths, which are correlated with, respectively, vocabulary difficulty and syntactic complexity but not perfectly so. For instance, relatively long and low-frequent words that consist of multiple well-known morphemes are not necessarily difficult to read because of their length (Bailin & Grafstein, 2016), and may in fact be easier to understand than their length would predict, due to their morphological structure supporting recognition (Balling, 2008). Similarly, longer sentences with simpler structure tend to be easier to read than shorter sentences with more complex syntactic structures, but this is not reflected in simple readability measures. In addition, the formulaic nature of the readability formulas means that the word and sentence length measures as well as frequency are assumed to have straightforwardly linear incremental effects, which is not necessarily the case (Bailin & Grafstein, 2016).

Another major issue, for readability formulas and for text processing in general, is text coherence, i.e. the logical structure of the text, and the explicit cohesive devices used to mark coherence. These are not captured by traditional readability measures, but are likely to play a central role to making sense of texts. A more recent attempt at automated capture of text readability, Coh-Metrix (for a comprehensive overview, see McNamara, Graesser, McCarthy, & Cai, 2014), does, as the name suggests, focus to a large extent on coherence, including multiple measures of markers of coherence, such as causal and referential cohesion. This approach is more refined than classical readability formulas, and generally also predicts text difficulty better, to the extent that we can measure that. However, the Coh-Metrix approach also suffers from one of the same fundamental problems as the more traditional readability formulas, namely the assumption that readability can be measured through some mechanistic combination of formal properties of the text (Bailin & Grafstein, 2016). Coh-Metrix uses more and more fine-grained variables, but it remains an issue for discussion whether this class of approaches really capture what we want to capture, and whether it is meaningful to attempt to measure readability based on texts alone, to the exclusion of the text user.



## **4. Bridging the gap**

### **4.1 Naturalistic experiments**

One way to attempt to bridge the gap between the lab and the wild in text processing research is to use experiments that are more naturalistic than the classical experiments described in the introduction. One way to do so is working with eye-tracking rather than experiments whose key measurements are based on explicit responses, like grammaticality judgment and self-paced reading. This method has been used more for studies of word and sentence processing than for studies of text, but at least since Rayner, Chace, Slattery, & Ashby (2006) also to investigate text and discourse processing. Rayner and colleagues found more and slightly but significantly longer fixations for complex texts, indicating that eye movements can be used as measurements of global text difficulty and text comprehension.

However, the use of eye-tracking methodology does not in itself make an experiment naturalistic. It does probably makes the reading process more similar to real-life reading processes than classical experimental tasks, but further steps are needed. One of them is investigating texts that are sampled from actual language use rather than constructed by the experimenter. For instance, I used authentic, only slightly edited descriptive and expository texts to investigate the effect of writing advice – such as ‘avoid passives’ and ‘avoid nominalisations’ that tend to show up in writing guides and language policies – on reading comprehension, in L1 Danish (Balling, 2013a) and L2 English (Balling, 2018). These studies are in some ways experimental in the sense outlined in the introduction, primarily because the investigation is based on two groups of participants each reading a different version of the same (sentential or phrasal) constructions. These experiments are nonetheless more naturalistic, and hence presumably more ecologically valid, than traditional experiments because they are based on authentic texts with minor experimental manipulations.

This use of authentic texts relies on three key design and analysis decisions: firstly, the experiments used eye tracking of reading. Secondly, the design and analysis relied on a regression approach where a range of relevant variables could be statistically controlled in the statistical analysis; since many predictor variables – including the frequency, predictability and length of words and constructions – by definition cannot be controlled beforehand in authentic texts, the statistical control becomes an absolute necessity. While length and frequency are relatively standard measures, predictability is harder to work with, leading to the third key

design decision of controlling predictability through conditional trigram frequency (originally inspired by MacDonald & Shillcock, 2003). The basic logic of this approach is that we index the predictability of a target word by taking the joint frequency of the target word and the two words preceding it and dividing it by the joint frequency of the two preceding words. For instance, for the highly predictable target word ‘fløde’ (cream) in the phrase ‘rødgrød med fløde’ (roughly translated as jelly with cream, a Danish dessert whose name is famously hard for non-Danes to pronounce):

$$p(\text{fløde}) = \frac{\text{freq}(\text{'rødgrød med fløde'})}{\text{freq}(\text{'rødgrød med'})}, \text{ or } p(\text{cream}) = \frac{\text{freq}(\text{'jelly with cream'})}{\text{freq}(\text{'jelly with'})},$$

In other words, how often out of the times we find jelly with X is that X actually cream. In this case quite frequently, but of course the measure may also be used for very low predictabilities, and crucially also to gauge the differences between different low predictabilities by using tools from natural language processing (particularly the modified Kneser-Ney smoothing of Chen & Goodman, 1998) to deal with non-attested word bigrams and trigrams. This is in contrast to the standard method of measuring predictability, namely asking a group of participants to fill in cloze tests for the target words. The cloze method tends to assign the same zero probability to many words which are associated with probabilities which are different but not high enough for the word to show up in a cloze test (Yan, Kuperberg, & Jaeger, 2017). This lack of sensitivity at the low end of the scale is particularly problematic in view of the evidence that predictability effects are logarithmic in nature (Smith & Levy, 2013). The word trigram-based method described here has the additional advantage over cloze testing with human participants that, once the language model is trained, the extraction of the predictability measure for the relevant words is extremely fast. In the text processing experiments of (Balling, 2013a, 2018), the trigram-based predictability measure was averaged across the target constructions to index the average predictability of the words in the constructions.

These three design features – the use of eye-tracking, statistical control in regression analyses, and trigram models to index predictability – were used to allow the comparison of different types of target constructions in authentic descriptive and expository texts. The texts were only slightly edited to vary the versions of the target constructions between those forms that are recommended by writing guides and those that are labelled as

problem constructions, for instance actives vs. passives and sentential vs. nominal constructions (see an overview of the most prominent construction types in table 1). The original study by Balling (2013) showed no difference in fixation time between the recommended and problem constructions for highly skilled L1 readers of Danish. Balling (2018) tested a similar group of readers in their L2 English, investigating parallel differences for a lower-proficiency language but for readers with presumably similar decoding and general language skills. Again, this study did not show an effect on the fixation time on the different types of constructions. As a further attempt to encourage naturalistic but still somewhat controlled reading, the 2018 study used a set hypothetical but realistic comprehension frame for the texts and open questions to measure comprehension.

<b>Problem</b>	<b>Recommendation</b>	<b>Example</b>
Nominalisation	Verbal construction	- <i>is in relation to</i> + <i>relates to</i>
Reduced relative clause	Full relative clause	- <i>information contained</i> + <i>information that is contained</i>
Passive verb	Active verb	- <i>amounts covered</i> + <i>amounts we cover</i>
Long or complex words or sentences	Shorter sentences or words	- <i>be different</i> + <i>differ</i>

Table 1. Examples of the construction types investigated in Balling (2013a) and (2018), adapted from Balling (2018, table 1)

There are various possible reasons for this failure to detect an effect, aside from the substantive interpretation that the differences investigated do not in themselves matter, on which more below. These possible reasons fall into two groups: specific problems with these specific studies, and more general issues with this type of naturalistic experiment. Among the specific reasons is the obvious one that the power of the experiments may not have been sufficient to detect effects of this manipulation; the fact that the experiments did show effects of other predictors like construction length and the position of the construction in the sentence makes this explanation less likely, although it does remain a possibility. Turning to the design characteristics of the experiments, another possibility is that the texts were of too high quality (the manipulations on purpose did not disrupt the coherence of the texts), that the readers were too proficient to be affected

by these relatively minor manipulations, and, related to both these points, the possibility that the relevant manipulations – such as active vs. passive – pertain so exclusively to the surface code of the text that they do not affect processing in any measurable way, not even the relatively mechanistic reading time measures employed in these studies.

There are also more general potential problems with the two studies that arise because of the attempt to make the experiments as naturalistic as possible. One issue is the averaged conditional trigram probability used to index predictability: while this index works well as a predictor of the predictability of single words (Balling, 2013b), the averaged measure used in these two experiments and elsewhere (Balling & Kizach, 2017) may not be sensitive enough and is often only borderline significant. Another issue is that because of the use of authentic texts and a naturalistic set-up, the data are potentially quite noisy, particularly those from the 2018 study where participants were presented with full pages of text, while the 2013 study used sentence by sentence presentation which gives cleaner, or at least more “cleanable” eye-tracking data, but again also less natural reading.

#### **4.2 Going more experimental: manipulating voice and givenness**

Although we must always be careful with interpreting null results like the ones discussed above, it is conceivable that the construction type differences in themselves do not actually make a difference to text processing and comprehension. Nevertheless, it also remains a possibility that differences such as the one between active and passive constructions do in fact matter, but only when considered in conjunction with the key factor of text coherence. This possibility was investigated in a more strictly controlled experiment where the use of active vs. passive voice was manipulated in conjunction with the givenness of the agent and theme roles (Balling, in preparation).

The experiment used sentence-by-sentence self-paced reading of short constructed texts that were partly based on authentic texts from news outlets. Each text consisted of six pairs of sentences: target sentences with transitive main verbs in either the active or the passive voice and, immediately preceding each target sentence, a context sentence which set up either the agent or the theme of the target sentence as explicitly given, see table 2 for an example quadruple of related sentences. The dependent variable was reading time on the target sentences. In addition to the main 2\*2 manipulation of voice and givenness, the analysis also included control

variables such as sentence length in characters, trial number (arguably indexing structural priming because the structures of target sentences were quite similar), and reading time on the immediately preceding context sentence.

CONTEXT SENTENCES		TARGET SENTENCES	
Agent of target sentence given	Another focus area is [DNA-investigations] <sub>agent</sub>	Active	In the individual herd, [determine] <sub>verb active</sub> [DNA-investigations] <sub>agent</sub> [family relations between the giraffes] <sub>theme</sub>
Theme of target sentence given	Another focus area is [family relations between the giraffes] <sub>theme</sub>	Passive	In the individual herd, [determine] <sub>verb passive</sub> [family relations between the giraffes] <sub>theme</sub> by [DNA-investigations] <sub>agent</sub>

Table 2. An example of a target sentence in active and passive voice versions, with context sentences setting the agent or the theme up as given. Each of the target sentences occurred with each of the context sentences, in different versions of the texts. Translated from the original Danish (preserving Danish V2 word order).

The underlying assumption of the manipulation is that a target sentence is easier to read if it is more coherent with the immediately preceding context sentence, and that such coherence may be at least partially achieved if the subject of the target sentence, which occurs as the first NP associated with the target verb, is explicitly given by the context sentence. This leads to the hypothesis that active sentences will be easier to understand if the agent – which takes the subject position in active sentences – is given, while passive sentences are easier to read if the theme – which is the subject of passive sentences – is given by the context sentence. However, this was not the case: a mixed-effects regression model (Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2016) in the R statistical environment (R Core Team, 2016) showed no significant interaction between voice and which thematic role was given by the context, no difference between active and passive sentences, and a main effect advantage for sentences in which the theme rather than the agent was given, an effect which is not in itself really interpretable.

This experiment was an attempt to further investigate the absence of an effect in the eye-tracking experiments described in section 4.1. At the same time, it was also an additional exploration of the continuum between experiments and experience, attempting to address two of the problems

with the previous studies – the predictability measurement problem and the noisiness of the data – that arose because of their naturalistic approach. The predictability issue was at least partly addressed by the systematic manipulation of voice and givenness on the same lexical material, and indeed an aggregated conditional trigram probability measure was not significant in these analyses. The same systematic manipulation could also contribute to less noisy data, compared to the many different types of constructions and the variations of them used in the experiments reported in section 4.1. However, this systematicity came at the cost of naturalness, with the constructed texts arguably coming across as too artificial to the readers. Finally, the reading time in the sentence-by-sentence self-paced task, which was partly chosen with the practical objective of being able to run multiple experiments simultaneously and thus get more participants, may be too insensitive to the after all relatively minor manipulation. Although the explicit givenness of first NP associated with the target verb does probably improve coherence, the difference between the two NPs was in practice one of relative givenness: in order to get the texts to work as texts, the not explicitly given NP was in many cases implicitly given.

### **4.3 Other avenues of research**

While the approaches described in sections 4.1 and 4.2 should not be entirely discounted, the problems with them are also such that other avenues of research should be explored. There are several interesting possible perspectives, but common to them is the need to take into account multiple aspects of text processing.

One interesting way of doing this, which stays squarely in the experimental camp, is the approach of Kuperman, Matsuki, & Van Dyke (2018) who investigate the effects of the readers' cognitive and linguistic ability, the linguistic properties of the text, and the temporal dynamic of the reading process, and crucially also the interaction and relative weights of these. This goes some way towards understanding the many levels of text processing and the emphasis on interactions is both novel and crucial. More generally, reader proficiency in a broad sense is an important aspect to take into account, and on trend in relation to the recent emphasis on individual differences in language processing (for an overview, see Kidd, Donnelly, & Christiansen, 2018). However, the clear experimental focus of Kuperman et al. (2018) still means that the more experience-related issues of in-depth comprehension and reader reception are not clearly addressed.

Given the interdependence of the different aspects of text processing outlined in this chapter, an ideal would be joint consideration of the multiple aspects – including comprehension, speed, and reception – while looking at both the narrowly processing-oriented aspects measured in experiments and the experience of language users in the wild. This ideal may be partially implemented by mixed-methods approaches, though it remains to be worked out exactly how to interpret potentially diverse results together. As a compromise, the reception aspect and variations in comprehension beyond what may be measured in multiple-choice questions should as a minimum be considered as valid concerns in relation to text processing; conversely, the more text processing oriented aspects should in turn be seriously evaluated rather than just assumed in writing guides and language policies.

## 6. References

- Bailin, A., & Grafstein, A. (2016). *Readability: Text and Context*. Houndmills: Palgrave Macmillan.
- Balling, L. W. (2008). *Morphological Effects in Danish Auditory Word Recognition*. PhD-thesis, University of Aarhus.
- Balling, L. W. (2013a). Does Good Writing Mean Good Reading? An Eye-tracking Investigation of the Effect of Writing. *Fachsprache*, 35(1-2), 2-23.
- Balling, L. W. (2013b). Reading authentic texts: What counts as cognate? *Bilingualism: Language and Cognition*, 16(3), 637-653. <https://doi.org/doi:10.1017/S1366728911000733>
- Balling, L. W. (2018). No Effect of Writing Advice on Reading Comprehension. *Journal of Technical Writing and Communication*, 48(1), 104-122. <https://doi.org/10.1177/0047281617696983>
- Balling, L. W., & Kizach, J. (2017). Effects of Surprisal and Locality on Danish Sentence Processing: An Eye-Tracking Investigation. *Journal of Psycholinguistic Research*, 46(5), 1119-1136. <https://doi.org/10.1007/s10936-017-9482-2>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67(1), 1-48. <https://doi.org/10.18637/jss.v067.i01>
- Björnson, C. H. (1968). *Läsbarhet*. Stockholm: Liber.

- Bråten, I., & Anmarkrud, Ø. (2013). Does naturally occurring comprehension strategies instruction make a difference when students read expository text? *Journal of Research in Reading*, 36(1), 42-57. <https://doi.org/10.1111/j.1467-9817.2011.01489.x>
- Chen, S. F., & Goodman, J. (1998). An empirical study of smoothing techniques for language modeling. *Harvard University Technical Report, TR-10-98*.
- Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS ONE*, 10(8), 1-38. <https://doi.org/10.1371/journal.pone.0134008>
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215-251. <https://doi.org/http://dx.doi.org/10.1037/0033-295X.87.3.215>
- Federal Plain Language Guidelines*. (2011). Retrieved from <http://www.plainlanguage.gov/howto/guidelines/FederalPLGuidelines/FederalPLGuidelines.pdf>. April 30, 2012.
- Flesch, R. (1943). *Marks of Readable Style: A Study in Adult Education*. Teachers College, Columbia University.
- Garwood, K. (2014). *Plain, but not Simple: Plain Language Research with Readers, Writers and Texts*.
- Graesser, A. C., & Millis, K. (2011). Discourse and Cognition. In T. A. van Dijk (Ed.), *Discourse Studies: A Multidisciplinary Introduction* (pp. 126-142). London: SAGE Publications. <https://doi.org/10.4135/9781446289068.n16>
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371-395. <https://doi.org/10.1037/0033-295X.101.3.371>
- Harley, T.A. (2008). *The Psychology of Language. From Data to Theory* (3rd ed.). Hove & New York: Psychology Press.
- Jacobsen, H. G., & Jørgensen, P. S. (1992). *Håndbog i Nudansk* (2nd ed.). København: Politikens Forlag.
- Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Sciences*, 22(2), 154-169. <https://doi.org/10.1016/j.tics.2017.11.006>
- Kincaid, J.P., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). *Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel*. Research Branch Report 8-75. Millington, TN, Naval Technical Training, U. S. Naval Air Station, Memphis, TN.
- Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- Kintsch, W., & Yarborough, J. C. (1982). Role of rhetorical structure in text comprehension. *Journal of Educational Psychology*, 74, 828-834.



- Kjærgaard, A. (2011a). Det laaange seje træk, del 2. Mere om Sprogpolitik for Danmarks Domstole. *Nyt Fra Sprognævnet*, 2011/2, 7-12.
- Kjærgaard, A. (2011b). Nytter det? – Om de tekstlige effekter af sprogpoltiske projekter i offentlige institutioner. *Nydanske Sprogstudier*, 40, 90-116.
- Kjærgaard, A. (2012). Fra lidenskab til lige gyldighed – En caseanalyse fra Danmarks Domstole af et sprogpoltisk projekts (manglende) gennemslagskraft. *Sakprosa*, 4(1), 1-28.
- Kjærgaard, A. (2015). Påvirker omskrivninger af tekster fra det offentlige borgernes forståelse – og hvordan? *Sakprosa*, 7(2), 1-25.
- Kjærgaard, A. (2016). The organisation of the plain language movement in Denmark. In P. Nuolijärvi & G. Stickel (Eds.), *Language use in public administration. Theory and practice in the European states. Contributions to the EFNIL Conference 2015 in Helsinki*. (pp. 123-134). Helsinki: EFNIL.
- Kjærgaard, A., Gravengaard, G., Dindler, C., & Hjuler, S. (2018). Tænke højt-protokoller. En metode til at undersøge modtageres tekstforståelse og -oplevelse. *Nydanske Sprogstudier*, 54.
- Klare, G. R. (1984). Readability. In P. D. Pearson (Ed.), *Handbook of Reading Research*, pp. 681-741. Mahwah, NJ: Lawrence Erlbaum.
- Kuperman, V., Matsuki, K., & Van Dyke, J. A. (2018). Contributions of reader- and text-level characteristics to eye-movement patterns during passage reading. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 44(11), 1687-1713. <https://doi.org/10.1037/xlm0000547>
- Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, 65(1), 42-73. <https://doi.org/10.1016/j.jml.2011.03.002>
- Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016). lmerTest: Tests in Linear Mixed Effects Models. Retrieved from <https://cran.r-project.org/package=lmerTest>
- MacDonald, S. A., & Shillcock, R. C. (2003). Low-level predictive inference in reading: the influence of transitional probabilities on eye movements. *Vision Research*, 43, 1735-1751.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. New York: Cambridge University Press.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357-383. <https://doi.org/10.1080/10888430701530730>
- Perfetti, C. A., & Hart, L. (2002). The Lexical Quality Hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213).
- Perfetti, C., & Adlof, S. M. (2012). Reading Comprehension: A Conceptual Framework from Word Meaning to Text Meaning. In J. Sabatini & E. Albro (Eds.), *Assessing reading in the 21st century: Aligning and applying advances in the reading and measurement sciences* (pp. 3-20).

- Pham, H., & Sanchez, C. A. (2018). Text Segment Length Can Impact Emotional Reactions to Narrative Storytelling, to appear in *Discourse Processes*, 0(0), 1-19. <https://doi.org/10.1080/0163853X.2018.1426351>
- R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from <http://www.r-project.org/>
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye Movements as Reflections of Comprehension Processes in Reading. *Scientific Studies of Reading*, 10, 241-255.
- Rozakis, L. (2000). *Complete Idiot's Guide to Writing Well*. East Rutherford, NJ, USA: Penguin Putnam.
- Schrivver, K. A. (1991). Plain Language for Expert or Lay Audiences: Designing Text Using Protocol-Aided Revision, Technical Report No. 46, Center for the Study of Writing. Available from <https://files.eric.ed.gov/fulltext/ED334583.pdf>
- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302-319. <https://doi.org/10.1016/j.cognition.2013.02.013>
- Sorenson, S. (2010). *Webster's New World Student Writing Handbook* (5th ed.). Hoboken, New Jersey: Webster's New World.
- Spyridakis, J. H., & Isakson, C. S. (1998). Nominalizations vs. denominalizations: do they influence what readers recall? *Journal of Technical Writing and Communication*, 28, 163-188.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.
- Veldre, A., & Andrews, S. (2018). Beyond cloze probability: Parafoveal processing of semantic and syntactic information during reading. *Journal of Memory and Language*, 100, 1-17. <https://doi.org/10.1016/j.jml.2017.12.002>
- Williams, J. M. (2005). *Style. Ten lessons in clarity and grace* (8th ed.). New York: Pearson Longman.
- Yan, S., Kuperberg, G. R., & Jaeger, T. F. (2017). Prediction (Or Not) During Language Processing. A Commentary On Nieuwland et al. (2017) And Delong et al. (2005). *bioRxiv*. Retrieved from <http://biorxiv.org/content/early/2017/05/30/143750.abstract>.



## On the Need for Experimental Syntax

Ken Ramshøj Christensen  
Aarhus University

### Abstract

The use of expert intuition as a source of evidence in theoretical syntax has long been criticized. Here I review some of the main points of the debate. Using examples from research done by me and collaborators, I argue that an experimental approach is essential when studying subtle structural contrasts, in particular when doing comparative studies. The same applies to linguistic illusions where people are misled and interpret meaningless nonsense as meaningful. However, without expert intuition, experimental syntax would not get off the ground; it is based on expert intuition and syntactic analysis.

### 1. Introduction

Over the years, it has been debated whether the use of introspection is a reliable and valid source of data in theoretical syntax (Schütze, 1996). According to Gibson & Fedorenko (2010, 2013), the “standard” methodology in syntax, i.e. introspection in the form of expert intuitions about acceptability or grammaticality, is “weak”. However, as Sprouse & Almeida (2017) note in their response to Branigan & Pickering (2017), the claim that this is the “standard” approach “is a caricature of linguistic methodology that, to our knowledge, has never been supported by evidence. Nonetheless, a charitable interpretation of this claim reveals two separate concerns”, namely, “the routine use of small sample sizes” and “the susceptibility of [acceptability judgments] to investigator bias”. First of all, the contrast in grammaticality or acceptability between two otherwise minimally different sentences may be due to semantic properties

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 373-388). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.

of the individual lexical items (selection bias), rather than to the syntactic phenomenon in question. Multiple instances of the construction in question should be evaluated in order to make generalizations. Secondly, there is high a risk of confirmation bias on the part of the researcher seeking to support (or refute) some hypothesis, and indeed expert intuitions are normally not considered data in other branches of science. Asking a few colleagues or students may also bias the data, because they might be inclined to agree merely because they (more or less) subconsciously want to please the researcher. According to Gibson & Fedorenko (2010, p. 233), “the lack of validity of the standard linguistic methodology has led to many cases in the literature where questionable judgments have led to incorrect generalizations and unsound theorizing, especially in examples involving multiple clauses, where the judgments can be more subtle and possibly more susceptible to cognitive biases”. The remedy, they argue, is to adopt a quantitative approach, e.g. by using corpus studies and experiments with multiple items and participants.

While Culicover & Jackendoff (2010) agree that grammaticality judgments should always be made on properly controlled data, they also argue that sometimes, subjective judgments are sufficient and just as good as experimental data. For one, corpus data may not always be very helpful. Certain sentence types, phrases, and words, which people nonetheless have clear intuitions about (a classic example is the parasitic gap), are very rare and may indeed not be found in a corpus, but very little (if anything) can be deduced about the grammatical status of such items from their non-occurrence in a corpus (Newmeyer, 2003).

Furthermore, as I shall argue in detail below, some intuitions are very robust and stable across subjects, including intuitions about grammatical illusions. This is true for language as well as for other cognitive domains, such as vision. Consider the diagrams in Figure 1. There is no need for a large sample of intuitions to ascertain that people consistently see a white triangle (which is not actually there) in the Kanizsa triangle, that the Necker cube is ambiguous (the lower left square is either the front or the back of the transparent box), or that the Devil’s tuning fork is an impossible object (once you look closer):

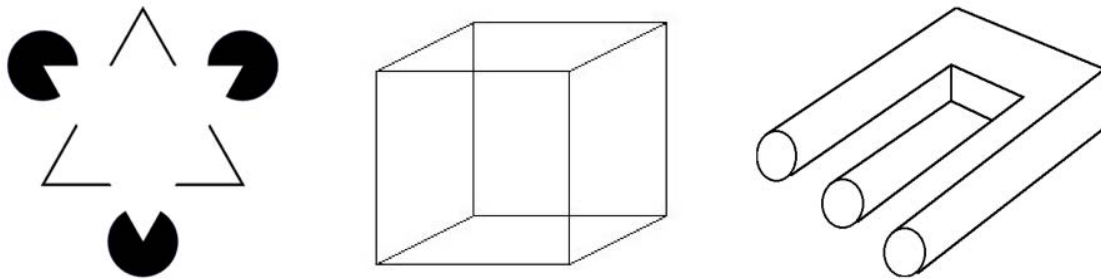


Figure 1. From left to right: the Kanizsa triangle, the Necker cube, and the Devil's tuning fork (also known as an impossible trident or a blivet)

In the same way, subjective judgments can form the basis for theory development, which may inspire experiments; just like optical illusions provide can be used to test visual theories, intuitions about sentences can be used to test grammatical theories (Townsend & Bever, 2001, p. 184). Indeed, “grammaticality judgments are the raw material for hypotheses about the structure of the language faculty. Without such judgments, the experimental enterprise cannot get off the ground” (Culicover & Jackendoff, 2010, p. 234). Along the same lines, Phillips (2009) argues that there is no crisis in theoretical linguistics. Before empirical claims become widely accepted generalizations, they are “scrutinized” by the linguistic community, and the standard methodology in theoretical syntax has not led to “unsound theorizing”. In fact, “carefully constructed tests of well-known grammatical generalizations overwhelmingly corroborate the results of ‘armchair linguistics’” (Phillips, 2009, p. 53) – in at least 95% of cases, according to Sprouse & Almeida’s (2013) analysis of 1743 judgment pairs, but see Gibson et al. (2013). A replication rate of 95% is, to put it mildly, very impressive – far better than that of other sciences, including psychology (39%) and cancer biology (10%), as well as chemistry, physics, and medicine (Baker, 2016; Open Science Collaboration, 2015). Similarly, Featherston (2009, p. 131) argues that quantitative data and statistical analyses are indeed powerful tools, “but still just tools”, which “produce a quantitative measure of how well some data supports our hypotheses”. He suggests that “linguists use data and apply statistical tests, but do not forget that both the starting point and the end point of a study must be a grammatical analysis”.

We should be methodologically tolerant because subjective introspection and experimental methods corroborate each other with an impressive level of convergence. A recent similar debate about whether syntactic priming is superior to and should replace acceptability judgments, or whether the two (and other) methodologies in fact supplement each

other can be found in the target paper by Branigan & Pickering (2017) and the many open peer commentaries, e.g. Adger (2017), Ambridge (2017), Hagoort (2017), Sprouse & Almeida (2017).

So, should we just “relax, lean back, and be a linguist” (Featherston, 2009)? Well, that depends. Although there may not be any real crisis (or at least, no more than in the sciences in general), there is still a serious issue. Gibson et al. (2013) argue that the 95% replication rate reported by Sprouse & Almeida (2013) is inflated due to the inclusion of theoretically irrelevant examples such as those in (1) below (where \* means ungrammatical): everyone agrees about their acceptability and as such, they have no bearing on the falsification of hypotheses or on the choice between theories.

- (1) a.       \*Was kissed John  
       b.       John was kissed.

Like with the Kanizsa triangle in Figure 1, there is actually no need for an experiment or a survey to argue that (1)a is ungrammatical in English, whereas (1)b is completely well-formed; intuitions from the expert in the “armchair” will do. Furthermore, such examples “are not representative of the forefront of syntactic research because all current linguistic theories correctly predict [such] contrasts” (Gibson et al., 2013, p. 3).

However, even with an acceptable error rate of 5%, as is the norm in psychology and social science in general (reflected in the standard threshold of statistical significance,  $p < 0.05$ ), non-quantitative methods have no means of discovering what the errors are and correcting them. Behavioral, quantitative studies are required to test whether the subjective intuitions match reality. Furthermore, the more judgment pairs (intuitions) from a single speaker in a paper, the higher the risk of errors and an increasing uncertainty about what the data is. Assuming 5% error in a set of 1743 judgment pairs (Sprouse & Almeida, 2013; Sprouse, Schütze, & Almeida, 2013), 87 will be incorrect. That may not sound as a lot, but according to Gibson et al. (2013), in such large data set, there are  $5.26 \cdot 10^{148}$  possible ways of 5% being wrong (choosing 87 from 1743). A truly “unfathomable” number (Gibson et al., 2013, p. 233) – even when compared to the number of fundamental particles in the observable universe:  $10^{80}$  (Mastin, 2018), or to the much smaller number of stars:  $10^{22}$  (ESA, 2016). Even in a (short) book with a mere 100 example pairs, the number of ways of having 5% errors (5% ‘wrong’ subject/expert intuitions) is larger than 75 million. However, the findings reported by Sprouse & Almeida (2013) have been replicated by Mahowald, Graff, Hartman, & Gibson (2016) who also suggest an experimental method which makes it possible to make statistically valid

generalizations about acceptability from a very small sample. However, Mahowald et al. (2016, pp. 630-631) emphasize that their method requires clear (contrasts in) acceptability judgements from the researcher based on “informal investigation” and that “statistics should supplement, not replace, careful thought about syntax and semantics”.

The message here is that with complex theories such as generative grammar, there is a need for a very high degree of reliability, and fine-grained syntactic contrasts of theoretical importance call for quantitative experimental methods. In this paper, I will illustrate the need for experimental syntax using work by myself and my collaborators.

## 2. Escapable islands

A syntactic island is a configuration that blocks extraction (Chomsky, 1986, 1995; Hofmeister & Sag, 2010; Rizzi, 1990; Ross, 1967; Sprouse & Hornstein, 2013). They are ‘inescapable’ (or at least difficult to escape from) in the sense that phrases cannot be moved out of them; they are ‘marooned’ (in somewhat the same sense that a pirate marooned on a deserted island cannot escape).

One famous example is the wh-island blocking extraction from a complement clause (Christensen, Kizach, & Nyvad, 2013a). As shown in (2), it is fully acceptable to have an embedded object question or an embedded adjunct question; the wh-element undergoes (short) movement to the left edge of the embedded clause. It is also possible to extract the question element from the embedded clause into the matrix clause, as shown in (3) where the wh-element undergoes long movement: from the base position to the edge of the embedded clause and then to the edge of the matrix clause. Crucially, long movement proceeds in short local incremental steps. When the two types of extraction (short and long movement) are combined, as in (4), problems arise because the long movement cannot take place in short local steps, as illustrated in Figure 2.

- (2) I know [she can solve the problem in this way].  
 a. I know [which problem<sub>1</sub> she can solve \_\_<sub>1</sub> in this way].  
 b. I know [how<sub>1</sub> she can solve this problem \_\_<sub>1</sub>].
- (3) a. Which problem<sub>1</sub> do you think [\_\_<sub>1</sub> she can solve \_\_<sub>1</sub> in this way]?  
 b. How<sub>1</sub> do you think [\_\_<sub>1</sub> she can solve this problem \_\_<sub>1</sub>]?
- (4) a. ?Which problem<sub>1</sub> do you wonder [how<sub>2</sub> she can solve \_\_<sub>1</sub> \_\_<sub>2</sub>]?  
 b. \*How<sub>2</sub> do you wonder [which problem<sub>1</sub> she can solve \_\_<sub>1</sub> \_\_<sub>2</sub>]?



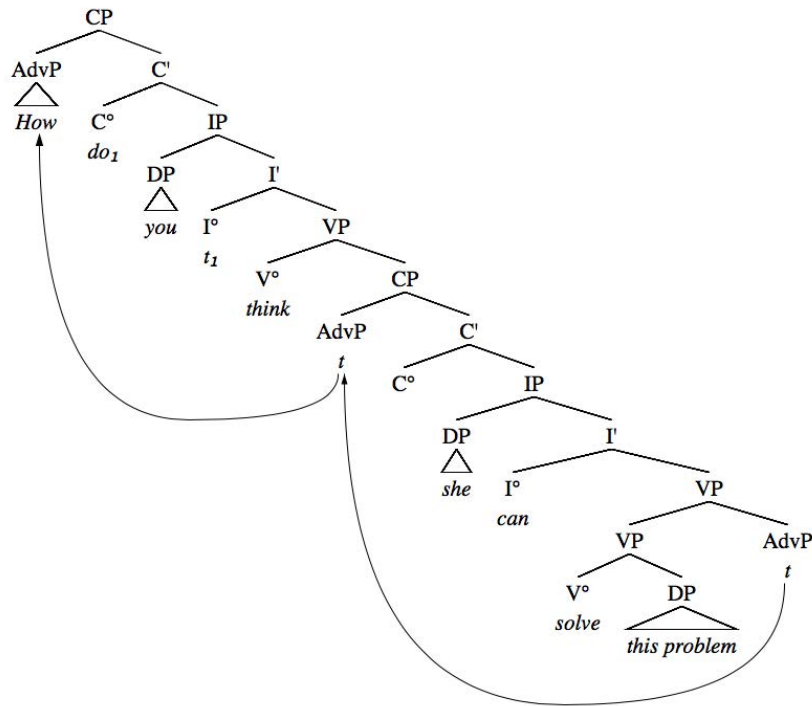


Figure 2.1. The syntactic structure of (3)b. Note that movement takes place in two successive (local) steps. Right: The syntactic structure of the ungrammatical (4)b. Here, long movement is not acceptable because it has to skip the position occupied by ‘which problem’.

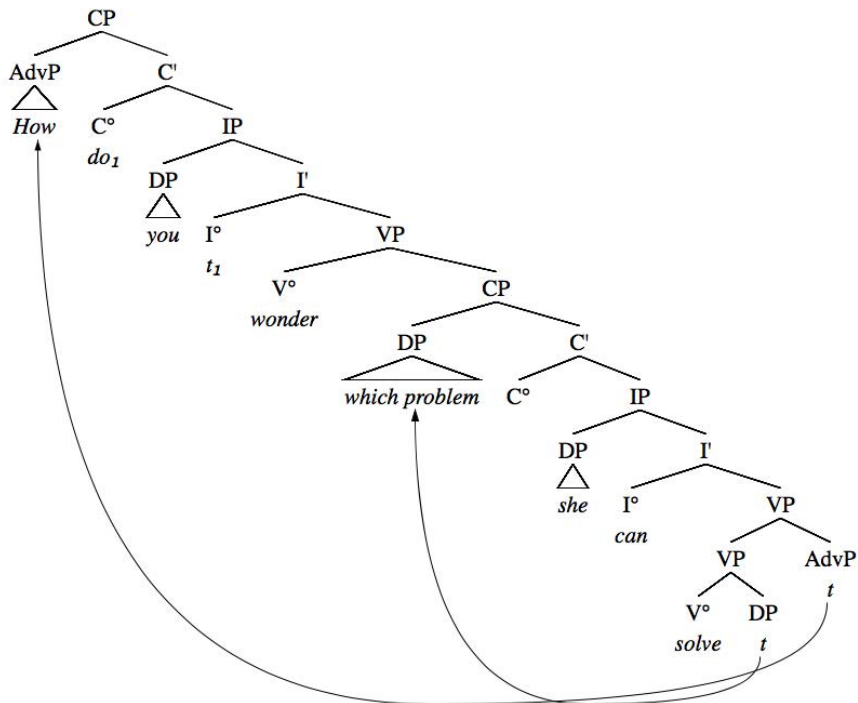


Figure 2.2. The syntactic structure of the ungrammatical (4)b. Here, long movement is not acceptable because it has to skip the position occupied by ‘which problem’.

In (4)a, moving the embedded *wh*-object (*which problem*) across the *wh*-adjunct (*how*) is unacceptable (sometimes the diacritics say ?? or ?\* indicating even lower levels of acceptability), if not fully ungrammatical. In (4)b, extracting the embedded *wh*-adjunct across the extracted *wh*-object is completely ungrammatical. The asymmetry in (4) is standardly assumed to be universal. It is indeed possible to find Danish examples that, at least to some speakers, match the asymmetry in (4), see e.g. Vikner (1995, p. 19). However, as shown by Christensen, Kizach & Nyvad (2013a; 2013b), it does not seem to hold in general for Danish.

In our studies, which involved three acceptability judgement experiments with multiple participants (60, 32, and 30), multiple different sentence tokens per condition (16, 12, and 16), and different scales of acceptability (two with a 5-point Likert-scale, one with a binary one), we found no statistically significant difference between the two island violations illustrated in (4). (Note that we replicated our initial results twice.) That is, the acceptability of the sentence pair in (5) is symmetric (they are equally acceptable), unlike the English structurally equivalent pair in (4) (asymmetric acceptability). People also found both significantly better than clearly ungrammatical control sentences.

- (5) a. ?Hvad<sub>1</sub> ved hun godt [hvor<sub>2</sub> man kan leje \_\_<sub>1</sub> \_\_<sub>2</sub>]?  
*What knows she well where one can rent?*  
 “What does she know where you can rent?”
- b. ?Hvor<sub>2</sub> ved hun godt [hvad<sub>1</sub> man kan leje \_\_<sub>1</sub> \_\_<sub>2</sub>]?  
*Where knows she well what one can rent?*  
 “Where does she know what you can rent?”

Extraction from a relative clause, as in (6)b below, is also assumed to be universally blocked due to the Complex Noun Phrase Constraint (Phillips, 2013; Ross, 1967):

- (6) a. She wanted to meet the man [who recorded the conversation]?  
 b. \*What<sub>1</sub> did she want to meet the man [who recorded \_\_<sub>1</sub>]?

Essentially, the problem is the same as illustrated in Figure 2: *who* in the relative clause blocks successive local movement of *what*. Though extractions from relative clauses have (famously) been reported to be acceptable in the Scandinavian languages (Engdahl, 1997; Engdahl & Ejerhed, 1982; Erteshik-Shir, 1973), such counter examples have been

argued to be merely ‘apparent’ counter examples; they do not involve extraction from relative clauses but from a different structure altogether, namely, small clauses (Kush & Lindahl, 2011; Kush, Omaki, & Hornstein, 2013). However, our experiment (Christensen & Nyvad, 2014), using examples such as (7), supports the idea that extractions from relative clauses are in fact grammatical in Danish, and that they are not merely apparent counter examples involving extractions from small clauses. This has subsequently also been shown for Swedish (Müller, 2015).

- (7) a. Pia har engang set/mødt en pensionist [som/der havde sådan en hund].  
*Pia has once seen/met a pensioner COMP had such a dog*  
 “Pia once met a pensioner who had such a dog.”
- b. Sådan en hund<sub>1</sub> har Pia engang set/mødt en pensionist [som/der havde \_\_<sub>1</sub>].  
*Such a dog has Pia once seen/met a pensioner COMP had*  
 “Such a dog Pia once met a pensioner who had.”

In our experiment (acceptability judgement on a 7-point Likert scale, 112 participants, 16 sentence tokens per condition) showed that the level of acceptability of extractions such as (7)b is highly dependent on the choice of matrix verb. The higher the frequency of usage of the main verb (measured as the number of occurrences in the online Danish corpus, KorpusDk), the more acceptable it is to have extraction from the relative clause inside the object. Consequently, the contrast in acceptability depends on lexical properties of the main verb, not on the construction as such. (It is also very easy to make a simple, fully acceptable sentence much less acceptable simply by using rare or less frequent words, compare *This man bought a new hat for his son* and *The gentleman purchased a novel bonnet for his offspring*.) This experiment also shows that it is important to include not only multiple participants but also multiple different tokens for each condition to avoid lexical confounds.

In short, Danish allows extraction from embedded questions, which are normally considered to be universally ungrammatical, and there is no argument–adjunct asymmetry in the extractions, also considered to be universal. Similarly, Danish allows extraction from relative clauses, also normally considered to be universally ungrammatical. These extraction patterns have serious implications for syntactic theory in general and for the syntactic theory of Danish in particular as they suggest a parametric difference between the two languages (Nyvad, Christensen, & Vikner, 2017;

Vikner, Christensen, & Nyvad, 2017) – “some islands have bridges that allow elements to escape, and this seems to be the case in the Scandinavian languages in particular” (Christensen & Nyvad, 2014, p. 42). Basically, the embedded CP layer in the tree in Figure 2 can be recursive in Danish but not in English, which also accounts for other independently observed phenomena (including stacked complementizers in Danish, e.g. *fordi at* ‘because that’). But to see these effects and to avoid wrong generalizations, we need careful experiments and quantitative analyses. It is not clear how it could have been done without experimental syntax.

### 3. From the borderlands of understanding

In this section, I argue that that quantitative intuition data can be used to address otherwise counter-intuitive interpretations of so-called linguistic illusions. While it is intuitively true that language usually makes sense, that it is usually meaningful, it is not always true. During parsing (the incremental construction of a syntactic representation in language comprehension), we sometimes make intermediate, semantically anomalous interpretations. In (8), for example, we initially and temporarily interpret *where* as a modifier of the matrix verb *believe* (this is called ‘early attachment’), even though it is a very unlikely and strange interpretation (??*Where did she believe? In the kitchen*). The extracted element is not compatible with the matrix verb. Subsequently, after encountering the rest of the sentence, we reanalyze it as modifying the embedded verb phrase (*buried the cat where?*). Despite the fact that sentences such as (8) are unambiguously grammatical, native speakers judge them as less than fully acceptable. Matrix verb incompatibility reduces acceptability (Christensen et al., 2013a; Fanselow & Frisch, 2006). (Here, it also seems difficult, though perhaps not impossible, to establish the systematic relationship between matrix verb incompatibility and reduced acceptability without an experimental approach.)

(8) Where did she believe that he had buried the cat?

Now, compare the two sentences in (9). In our experiment (Kizach, Nyvad, & Christensen, 2013) (60 participants, 16 different sentences, self-paced reading), we found that people initially attached *the pig in the pen* as the object of *noticed* and then reanalyzed it as the subject of *needed water* in (9)a. The matrix verb *notice* is compatible with either a nominal or a clausal object. In (9)b, on the other hand, people did not initially attach *the pig in the pen* as the object of *presumed*, because *presume* requires a clausal object.

- (9) a. Alice noticed the pig in the pen needed water.  
 b. Alice presumed the pig in the pen needed water.

Crucially, though, such counter-intuitive interpretations are only made if they do not violate the syntactic structure. In other words, because the syntax of the verb dictates that the object must be a clause, we do not make strange semantic interpretations. If on the other hand, the syntax allows for it, we do make strange temporary interpretations that affect the overall acceptability.

We can even be systematically tricked by certain syntactic constructions, sometimes into believing that certain sentences that are meaningless are actually meaningful. Because people disagree on the interpretations as well as on the acceptability of such examples, these counter-intuitive findings are only accessible with an experimental quantitative approach. Compare (10) and (11). While there is no doubt that (10) is ambiguous between meaning either that she used the bag to hit him with, or that she hit the bag-carrying man, people disagree on the interpretation of (11), which in fact does not have one.

(10) She hit the man with the bag.

(11) More people have been to Paris than I have.

(11) is a so-called comparative illusion (Phillips, Wagers, & Lau, 2011) or dead end (Christensen, 2010, 2016); see also Townsend & Bever (2001, p. 184) and Saddy & Uriagereka (2004).<sup>1</sup> At first sight, (11) seems to be elliptical; something has been left out after *than I have*, like in (12) where *have been* is elided (i.e. is not repeated) between *than* and *Copenhagen*; (12) means *More people than have been to Copenhagen have been to Paris*, where the *than* phrase is reconstructed in the middle of the sentence.

(12) More people have been to Paris than to Copenhagen.

If the same procedure is applied to (11), the result is seriously anomalous or incongruous: *\*More people than I have been to Paris have been to Paris*.

---

<sup>1</sup> The earliest mentioning (but not analysis) of this illusory construction that I know of is Montalbetti (1984, p. 6): “To Hermann Schultze, my eternal gratitude for uttering the most amazing \*/?sentence I’ve ever heard: More people have been to Berlin than I have. (Some have taken this sentence to be a proof of the autonomy of syntax!)”.

The sentence types in (10) and (11) are linguistic versions of the Necker cube and the Devil's tuning fork in Figure 1, respectively: The former is structurally ambiguous, the latter is globally incongruous or impossible.

I have investigated how people interpret sentences such as (11) in a series of studies, including an fMRI study (speeded acceptability, participants: n=19) (Christensen, 2010), an informal questionnaire (n=63) (Christensen, 2011), and an internet survey (multiple choice task, n=545) and two experiments (speeded acceptability, n=32 and 60) (Christensen, 2016). The results consistently showed that many people are tricked by the illusion and find sentences such as (11) meaningful. However, they do not agree on the interpretation. Interestingly, people seem to choose from a small set of mutually incompatible interpretations: 'Some people except me have been to Paris', 'More people than just me have been to Paris', or 'Some people have been to Paris more often than I have' – or they say that it is indeed meaningless. They do not find it ambiguous. This situation is very different from the one for (10), which people agree is meaningful and ambiguous.

Another type of illusion where people are systematically tricked is the so-called depth charge sentence (Kizach, Christensen, & Weed, 2016; Natsopoulos, 1985; Wason & Reich, 1979). Consider (13):

(13) No head injury is too trivial to be ignored.

Most people say that (13) means the same as (14), which is impossible. To ignore and to treat are definitely not the same, and in some contexts, they are opposites.

(14) No head injury is too trivial to be treated.

In our experiment, which included 19 participants and 150 sentences (moving window reading task), we manipulated three factors that together give rise to the depth charge effect: the number of negations ((13) has three: *no*, *trivial* [=not important], *ignore* [=not attend to]), the plausibility of the relation between the subject and the verb (*head injury* and *be ignored*: not plausible), and the logic of the relation between the adjective and the verb (the more *trivial* the less we *ignore*: illogical). When a sentence is maximally complex (i.e., when there are multiple negations, the relation between subject and verb is implausible, and the relation between adjective and verb is illogical), the majority of the participants misunderstood the sentence to mean the same as (14), but were at the same time certain of their

answers. Given that people have strong opinions about the interpretation, some have argued that their interpretation is true. And who am I to tell them otherwise? However, how can (13) and (14) be synonymous? As the experiment shows, the interpretation differs systematically, plus manipulating the three factors, leads to predictable increases in error rates. Again, these findings would not be possible without experiments and quantitative data. (Our study also confirms the two previous studies of the phenomenon, again showing a high replication rate for linguistic studies.)

#### 4. Conclusions

A sound approach that avoids “unsound theorizing” due to bias and secures a high degree of reliability and validity is experimental and quantitative. When people disagree significantly on the level of acceptability or grammaticality, or on the interpretation, an experimental quantitative approach is indeed required. Otherwise, it is difficult (if not impossible) to know what the data actually is or to detect whether or not the reported acceptability or interpretation is indeed real. This is particularly important with subtle distinctions of theoretical importance, such as the status of island violations, which are used to argue for universal properties of and constraints on human language. Likewise, when people disagree on acceptability and interpretation of linguistic illusions, we need experiments in order to determine the ‘borderlands’ of linguistic comprehension and to discover how linguistic processing interacts with general cognition. Finally, it should be kept in mind that none of the experimental findings discussed above (or elsewhere for that matter) would be possible without subjective intuition about acceptability and interpretation. Without expert intuition, the experimental enterprise would not get off the ground.

#### Acknowledgements

I would like to thank Ocke-Schwen Bohn for many interesting discussions about the validity of introspection and the nature of science. Thanks also to Anne Mette Nyvad, Sten Vikner, and Johannes Kizach, and to the anonymous reviewer for constructive criticism.

#### References

- Adger, D. (2017). The limitations of structural priming are not the limits of linguistic theory. *Behavioral and Brain Sciences*, 40, 18-19 (e283). <http://dx.doi.org/10.1017/S0140525X17000310>

- Ambridge, B. (2017). Horses for courses: When acceptability judgments are more suitable than structural priming (and vice versa). *Behavioral and Brain Sciences*, 40, 19 (e284). <https://doi.org/10.1017/S0140525X17000322>
- Baker, M. (2016). Is there a reproducibility crisis? *Nature*, 533(7604), 452-454. <https://doi.org/10.1038/533452a>
- Branigan, H. P., & Pickering, M. J. (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences*, 40. <https://doi.org/10.1017/S0140525X16002028>
- Chomsky, N. (1986). *Barriers*. Cambridge, Mass: MIT Press.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Mass.: MIT Press.
- Christensen, K. R. (2010). Syntactic reconstruction and reanalysis, semantic dead ends, and prefrontal cortex. *Brain and Cognition*, 73(1), 41-50. <https://doi.org/10.1016/j.bandc.2010.02.001>
- Christensen, K. R. (2011). Flere folk har været i Paris end jeg har. In Hansen, Inger Schoonderbeek & Widell, Peter (Eds.), *13. Møde om Udforskningen af Dansk Sprog* (pp. 113-136). Aarhus: Nordic Department, Aarhus University. Retrieved from [http://auinstallation42.cs.au.dk/fileadmin/projekter/Muds.dk/rapporter/MUDS\\_13.pdf](http://auinstallation42.cs.au.dk/fileadmin/projekter/Muds.dk/rapporter/MUDS_13.pdf)
- Christensen, K. R. (2016). The dead ends of language: The (mis)interpretation of a grammatical illusion. In S. Vikner, H. Jørgensen, & E. Van Gelderen (Eds.), *Let us have articles betwixt us – Papers in Historical and Comparative Linguistics in Honour of Johanna L. Wood* (pp. 129-160). Aarhus: Dept. of English, School of Communication & Culture, Aarhus University. Retrieved from <http://ebooks.au.dk/index.php/aul/catalog/download/119/107/466-1?inline=1>
- Christensen, K. R., Kizach, J., & Nyvad, A. M. (2013a). Escape from the Island: Grammaticality and (Reduced) Acceptability of wh-island Violations in Danish. *Journal of Psycholinguistic Research*, 42(1), 51-70. <https://doi.org/10.1007/s10936-012-9210-x>
- Christensen, K. R., Kizach, J., & Nyvad, A. M. (2013b). The processing of syntactic islands – An fMRI study. *Journal of Neurolinguistics*, 26(2), 239-251. <https://doi.org/10.1016/j.jneuroling.2012.08.002>
- Christensen, K. R., & Nyvad, A. M. (2014). On the nature of escapable relative islands. *Nordic Journal of Linguistics*, 37(01), 29-45. <https://doi.org/10.1017/S0332586514000055>
- Culicover, P. W., & Jackendoff, R. (2010). Quantitative methods alone are not enough: Response to Gibson and Fedorenko. *Trends in Cognitive Sciences*, 14(6), 234-235.
- Engdahl, E. (1997). Relative clause extraction in context. *Working Papers in Scandinavian Syntax*, 60, 59-86.
- Engdahl, E., & Ejerhed, E. (Eds.). (1982). *Readings on unbounded dependencies in Scandinavian languages*. Umeå; Stockholm, Sweden: Universitetet i Umeå; Almqvist & Wiksell International [distributor].
- Ertshik-Shir, N. (1973). *On the nature of island constraints* (PhD dissertation). Massachusetts Institute of Technology.



- ESA. (2016). How many stars are there in the Universe? Retrieved November 11, 2016, from [http://www.esa.int/Our\\_Activities/Space\\_Science/Herschel/How\\_many\\_stars\\_are\\_there\\_in\\_the\\_Universe](http://www.esa.int/Our_Activities/Space_Science/Herschel/How_many_stars_are_there_in_the_Universe)
- Fanselow, G., & Frisch, S. (2006). Effects of processing difficulty on judgments of acceptability. In G. Fanselow, C. Fery, & M. Schlesewsky (Eds.), *Gradience in grammar: Generative perspectives* (pp. 291-316). Oxford: Oxford University Press.
- Featherston, S. (2009). Relax, lean back, and be a linguist. *Zeitschrift Für Sprachwissenschaft*, 28(1), 127-132. <https://doi.org/10.1515/ZFSW.2009.014>
- Gibson, E., & Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences*, 14(6), 233-234. <https://doi.org/10.1016/j.tics.2010.03.005>
- Gibson, E., & Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28(1-2), 88-124. <https://doi.org/10.1080/01690965.2010.515080>
- Gibson, E., Piantadosi, S. T., & Fedorenko, E. (2013). Quantitative methods in syntax/semantics research: A response to Sprouse and Almeida (2013). *Language and Cognitive Processes*, 28(3), 229-240. <https://doi.org/10.1080/01690965.2012.704385>
- Hagoort, P. (2017). Don't forget the neurobiology: An experimental approach to linguistic representation. *Behavioral and Brain Sciences*, 40, 26 (e292). <https://doi.org/10.1017/S0140525X17000401>
- Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints and island effects. *Language*, 86(2), 366-415. <https://doi.org/10.1353/lan.0.0223>
- Kizach, J., Christensen, K. R., & Weed, E. (2016). A Verbal Illusion: Now in Three Languages. *Journal of Psycholinguistic Research*, 45(3), 753-768. <https://doi.org/10.1007/s10936-015-9370-6>
- Kizach, J., Nyvad, A. M., & Christensen, K. R. (2013). Structure before Meaning: Sentence Processing, Plausibility, and Subcategorization. *PLoS ONE*, 8(10), e76326. <https://doi.org/10.1371/journal.pone.0076326>
- Kush, D., & Lindahl, F. (2011). On the escapability of islands in Scandinavian. In *85th Annual Meeting of the Linguistic Society of America*. Pittsburgh, Pennsylvania. Retrieved from [http://ling.umd.edu/~kush/KushLindahl\\_LSA\\_ScandinavianExtraction.pdf](http://ling.umd.edu/~kush/KushLindahl_LSA_ScandinavianExtraction.pdf)
- Kush, D., Omaki, A., & Hornstein, N. (2013). Microvariation in islands? In J. Sprouse & N. Hornstein (Eds.), *Experimental Syntax and Island Effects* (pp. 239-264). Cambridge: Cambridge University Press.
- Mahowald, K., Graff, P., Hartman, J., & Gibson, E. (2016). SNAP judgments: A small N acceptability paradigm (SNAP) for linguistic acceptability judgments. *Language*, 92(3), 619-635. <https://doi.org/10.1353/lan.2016.0052>
- Mastin, L. (2018). The Universe By Numbers – The Physics of the Universe. Retrieved May 3, 2018, from <https://www.physicsoftheuniverse.com/numbers.html>
- Montalbetti, M. M. (1984). *After binding : on the interpretation of pronouns*.

- Massachusetts Institute of Technology, Dept. of Linguistics and Philosophy.
- Müller, C. (2015). Against the Small Clause Hypothesis: Evidence from Swedish relative clause extractions. *Nordic Journal of Linguistics*, 38(01), 67-92. <https://doi.org/10.1017/S0332586515000062>
- Natsopoulos, D. (1985). A verbal illusion in two languages. *Journal of Psycholinguistic Research*, 14(4), 385-397. <https://doi.org/10.1007/BF01067882>
- Newmeyer, F. J. (2003). Grammar is grammar and usage is usage. *Language*, 79(4), 682-707. <https://doi.org/10.1353/lan.2003.0260>
- Nyvad, A. M., Christensen, K. R., & Vikner, S. (2017). CP-recursion in Danish: A cP/CP-analysis. *The Linguistic Review*, 34(3), 449-477. <https://doi.org/10.1515/tlr-2017-0008>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Phillips, C. (2009). Should we impeach armchair linguists? In S. Iwasaki, H. Hoji, P. M. Clancy, & S.-O. Sohn (Eds.), *Japanese/Korean Linguistics 17* (pp. 49-64). Stanford University: CSLI Publications. Retrieved from [http://www.colinphillips.net/wp-content/uploads/2014/08/phillips2010\\_armchairlinguistics.pdf](http://www.colinphillips.net/wp-content/uploads/2014/08/phillips2010_armchairlinguistics.pdf)
- Phillips, C. (2013). Some arguments and nonarguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes*, 28(1-2), 156-187. <https://doi.org/10.1080/01690965.2010.530960>
- Phillips, C., Wagers, M. W., & Lau, E. F. (2011). 5 Grammatical Illusions and Selective Fallibility in Real-Time Language Comprehension. In J. T. Runner (Ed.), *Experiments at the Interfaces* (pp. 147-180). Bingley: Emerald Group Publishing.
- Rizzi, L. (1990). *Relativized minimality*. Cambridge, Mass: MIT Press.
- Ross, J. R. (1967). *Constraints on variables in syntax*. Massachusetts Institute of Technology.
- Saddy, D., & Uriagereka, J. (2004). Measuring language. *International Journal of Bifurcation and Chaos*, 14(02), 383-404.
- Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago, Il: University of Chicago Press.
- Sprouse, J., & Almeida, D. (2013). The empirical status of data in syntax: A reply to Gibson and Fedorenko. *Language and Cognitive Processes*, 28(3), 222-228. <https://doi.org/10.1080/01690965.2012.703782>
- Sprouse, J., & Almeida, D. (2017). Setting the empirical record straight: Acceptability judgments appear to be reliable, robust, and replicable. *Behavioral and Brain Sciences*, 40, 43-44 (e311). <https://doi.org/10.1017/S0140525X17000590>
- Sprouse, J., & Hornstein, N. (Eds.). (2013). *Experimental syntax and island effects*. Cambridge; New York: Cambridge University Press.
- Sprouse, J., Schütze, C. T., & Almeida, D. (2013). A comparison of informal and

- formal acceptability judgments using a random sample from Linguistic Inquiry 2001-2010. *Lingua*, 134, 219-248. <https://doi.org/10.1016/j.lingua.2013.07.002>
- Townsend, D. J., & Bever, T. G. (2001). *Sentence Comprehension. The integration of habits and rules*. Cambridge, Mass: MIT Press.
- Vikner, S. (1995). *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford University Press.
- Vikner, S., Christensen, K. R., & Nyvad, A. M. (2017). V2 and cP/CP. In L. Bailey & M. Sheehan (Eds.), *Order and structure in syntax I: Word order and syntactic structure* (pp. 313-324). Berlin: Language Science Press. Retrieved from <http://langsci-press.org/catalog/view/159/1125/971-1>
- Wason, P. C., & Reich, S. S. (1979). A verbal illusion. *Quarterly Journal of Experimental Psychology*, 31(4), 591-597. <https://doi.org/10.1080/14640747908400750>

## An Experimental Approach to the Conrad Phenomenon

Camilla Søballe Horslund

University of Amsterdam<sup>1</sup>

### Abstract

This study adopts an experimental approach to the Conrad Phenomenon, i.e. the phenomenon that second language (L2) learners can perform remarkably better in some aspects of their L2 while performing poorly at others. L2 performance in syntax, phonetics and phonology, and the lexicon in four L2 learner groups differing in L2 experience and native language background was examined and correlations between L2 performance in the three domains revealed a general trend of positive relations between domains, thus suggesting that the Conrad Phenomenon is uncommon. The strongest between-domain relation was observed between the lexicon and phonetics and phonology, thus supporting the notion of lexical facilitation in L2 speech acquisition.

### 1. Introduction<sup>2</sup>

Second language acquisition (SLA) studies most often investigate learning of different linguistic domains, e.g. syntax, phonology, or the lexicon. Comparisons between these domains, which are the aim of this study, are rare. Anecdotal evidence suggests that second language (L2) learners may do remarkably better at some language aspects than at others. One famous example is the Polish-British author Joseph Conrad, who wrote English remarkably well and with an, in many respects, native-like mastery of English grammar (Morzinski, 1994:24), but spoke English with a strong,

<sup>1</sup> This article presents part of my PhD project, which was conducted at Aarhus University.

<sup>2</sup> Many thanks to Ocke-Schwen Bohn for help and advice in developing the framework and constructing the tasks and for comments and suggestions concerning the analysis.

apparently unintelligible, Polish accent (Lucas, 1998), suggesting that he had been successful in his acquisition of English morphosyntax and lexicon, but not in his acquisition of the English sound system. This study adopts an experimental approach to the *Conrad phenomenon* (Scovel, 1978) by investigating whether Joseph Conrad was a special L2 learner or whether it is common among L2 learners to perform well within one linguistic domain and poorly within another.

The Conrad phenomenon is in line with the results of Snow and Hoefnagel-Höhle's (1979) study of native (L1) English speaking learners of Dutch, in which two separate factors of L2 ability were identified, i.e. lexical and morphosyntactic ability on the one hand and phonological ability on the other hand. The present study differs from Snow and Hoefnagel-Höhle's study in a number of ways. First, the subjects in Snow and Hoefnagel-Höhle's study were fully immersed in the L2 country, whereas the participants of this study are learners whose L2 exposure is largely through formal instruction. Second, Snow and Hoefnagel-Höhle investigated L2 learners from various age groups, whereas this study is concerned with adult L2 learners only. Third, this study investigates another L1-L2 combination than the one investigated in Snow and Hoefnagel-Höhle, namely L1 Danish and L1 Finnish learners of L2 English. Finally, some of the tasks used to measure syntactic L2 ability in Snow and Hoefnagel-Höhle's study confound variables and hence measure more than L2 syntactic ability. The tasks used in the present study clearly separate different variables in L2 performance.

A number of more recent linguistic studies also point to interesting differences in the acquisition of different linguistic domains. Age of acquisition has been found to constrain the learning of L2 phonology to a greater extent than the learning of L2 morphosyntax (Flege, Yeni-Komshian, & Liu, 1999). Likewise, Granena and Long (2012) found that age of acquisition affects L2 performance differently within pronunciation, morphosyntax, and lexicon. Specifically, Granena and Long reported that the age effect starts earlier for pronunciation with a cut-off point for reaching native-like pronunciation at five years of age compared to nine years for lexicon and twelve years for morphosyntax. The authors take these results as evidence for the existence of multiple sensitive periods in second language acquisition.

Age effects have also been studied by Knightly, Jun, Oh, and Au (2003), who tested production benefits of overhearing normal conversation during childhood, comparing childhood overhearers and late L2 learners

with respect to phonology and morphosyntax. Their results suggest an advantage of childhood overhearing in phonology but not in morphosyntax. The results of a study on retention of L1 remnants in international adoptees, who had been exposed to their native language for the first three months of their life (Hyltenstam, Bylund, Abrahamsson, & Park, 2009), also point to a qualitative difference in the acquisition of phonology and morphosyntax, since an advantage in phonological relearning was observed for international adoptees compared to regular L2 learners, while such an advantage was not observed in morphosyntactic relearning.

Moreover, results from studies of neural processing within different linguistic domains for L1 and L2 speakers (e.g. Bowden, Steinhauer, Sanz, & Ullman, 2013) suggest a difference between L2 processing of syntax and lexicon. While L2 semantic/lexical processing relies on native-like neural cognitive mechanisms, L2 syntactic processing seems to depend on degree of L2 experience or L2 proficiency, with advanced L2 learners showing native-like processing and less advanced L2 learners relying on semantic processing for syntax (Bowden et al., 2013).

Neither of these studies, however, compared performance between linguistic domains directly, which is the aim of this study. The present study investigates the Conrad phenomenon experimentally by examining L2 performance within three linguistic domains, syntax, phonetics and phonology, and the lexicon, in order to examine whether L2 learners' performance within one linguistic domain is related to their performance within other linguistic domains or whether it is possible to perform well within one domain of one's L2 while performing poorly within others.

## **1.2 The modularity approach**

The modularity approach presents a theoretical perspective on the Conrad phenomenon by viewing linguistic domains as modules, i.e. as partly separate entities in line with Elsabbagh and Karmiloff-Smith's view that 'modularity concerns the degree to which cognitive domains can be thought of as separable, i.e., whether they function independently of one another' (2006, p. 218).

The modularity debate is in part based on a number of different definitions of the term *module*. Some of these definitional disagreements may stem from the fact that the modularity approach encompasses several academic disciplines. While there is general consensus regarding the existence of modularity in highly specialised areas of vision, for instance, the question of modularity for higher order cognitive functions, such as

language, is much more controversial. One important distinction is the one between functional (or cognitive) modularity on the one hand and anatomical (or neural) modularity on the other hand (Elsabbagh & Karmiloff-Smith, 2006). According to the Functional Modularity Assumption, human cognition consists of several cognitive modules, which, in line with Fodor (1984), are characterised as being domain-specific, innately specified, and informationally encapsulated. The Anatomical Modularity Assumption builds on the Functional Modularity Assumption and adds that cognitive modules each reside in specific brain areas (Bergeron, 2007). As this study deals with behavioural data only, the present discussion is limited to functional modularity. A modularity approach to language adopts this idea of separation of cognitive domains either as a separation between language and general cognition (Chomsky, 1986, p. xxvi; 1988, p. 161) or as a subdivision within the language module such that separate submodules deal with different linguistic domains (Chomsky, 1965: 16; Sharwood Smith, 1994, pp. 17-18). The former is called external modularity and the latter is called internal modularity. Since the topic of this study is second language performance in different linguistic domains, this study is concerned with internal modularity only.

A modularity approach to L2 performance thus predicts that an L2 learner's performance in one linguistic domain is independent from the learner's performance in other linguistic domains. According to this approach, the Conrad phenomenon is accounted for by independence between the modules within which Joseph Conrad performed well, i.e. syntax and the lexicon, on the one hand and the module within which he performed poorly, i.e. phonetics and phonology, on the other hand.

### **1.3 Relations between domains in first language acquisition**

In first language acquisition research, the relationship between linguistic domains is the topic of an ongoing debate. In particular, the relationship between the development of lexical and morphosyntactic knowledge has been widely debated within different linguistic frameworks. The debate is motivated by a strong positive correlation between lexical and morphosyntactic knowledge and centres on the relative autonomy or interdependence of these two linguistic domains, i.e. the degree of internal modularity in first language acquisition (Marchman, Martínez-Sussmann, & Dale, 2004). This strong positive correlation between lexical and morphosyntactic knowledge in children is explained by the hypothesis that

lexical knowledge is a prerequisite for morphosyntactic knowledge (e.g. Marchman & Bates, 1994). However, others argue that morphosyntactic knowledge facilitates word learning (e.g. Anisfeld, Rosenberg, Gasparini, & Hoberman, 1998).

The idea that lexical acquisition drives morphosyntactic acquisition is often presented within a Single Mechanism Account. Marchman and Bates (1994), for instance, argue that the correlation between lexical and morphosyntactic acquisition is due to both domains being acquired by the same learning mechanism, which starts out as a rote learning mechanism that handles individual mappings but develops into a system building mechanism that both handles individual mappings and organises these mappings according to general patterns, e.g. regular verbs and irregular verbs. Importantly, this qualitative change in the learning mechanism comes about when the vocabulary reaches a critical mass, since the child's "dataset" needs to reach a certain size to support the extraction of general classifications. Marchman and Bates' study shows a significant positive non-linear relationship between vocabulary growth (number of verbs in particular) and the appearance of correct past tense formations as well as the onset of overregularisation errors, which the authors take as evidence for the Single Mechanism Account. Once the vocabulary reaches a critical mass, incremental increases in the number of verbs acquired result in qualitative shifts in the treatment of both previously acquired forms and novel forms.

A Dual Mechanism Account is known from the Words and Rules Theory, developed by Prince and Pinker (e.g. Pinker, 2006). The Words and Rules Theory holds that language acquisition relies on two qualitatively different learning mechanisms, namely associative memory of arbitrary sound-meaning relationships (the principle underlying the lexicon) and symbol-manipulating rules (the principle underlying the mental grammar). Hence, words must be rote learned, while the acquisition of grammar is subject to rule learning (Pinker, 1998). Pinker argues that, as children's memory retrieval is less reliable than adults', overregularisation errors in child speech serves as a compensation strategy for children when their memory fails them. Importantly, overregularisation errors in past tense formation start when the child acquires the regular rule, which is evident from the observation that the onset of overregularisation co-occurs with the point at which the child starts inflecting past tense forms more often than not (Pinker, 2006).



Regarding the proposed morphosyntactic facilitation of vocabulary acquisition, Anisfeld et al. (1998) argue that the onset of combinatorial speech may facilitate vocabulary acquisition in two ways. First, combinatorial speech calls for specificity of expression, which motivates word learning. Specifically, when children stop using words holophrastically (using a single word to express a complex idea), a need for more words arises. The observation of a car and the request to go for a car ride, for instance, which were both earlier expressed with the single word 'car', may now elicit two words each and thus become distinguishable, e.g. 'car there' and 'Johnny car'. Second, grammatical context helps children identify the meaning of words, especially relational words such as verbs. Anisfeld et al. do not explicitly propose any theoretical account of their findings in terms of single or dual mechanisms, but their argumentation seems to be more compatible with a dual mechanism account than with a single mechanism account, as lexical and grammatical acquisition are presented as qualitatively different.

The modularity debate in L1 acquisition does not seem to be on the verge of settlement, which may in part be due to the lack of clear empirical results favouring either modularity or non-modularity. A factor contributing to this lack of empirical decisiveness may be the unavoidable confound in L1 acquisition between linguistic development and the development of world knowledge; a confound that seems particularly relevant in lexical development. This problem is not present in adult L2 acquisition, as adult L2 learners' world knowledge is highly developed before language acquisition even begins. Hence, an examination of modularity in L2 acquisition may inform the modularity debate in L1 acquisition.

#### **1.4 Domain interdependence in SLA: The Vocab Model**

To my knowledge, only one theoretical account of L2 performance in different linguistic domains exists, namely the Vocabulary-Tuning Model of L2 Rephonologisation (the Vocab Model) (Bundgaard-Nielsen, Best, & Tyler, 2011a). Interestingly, the Vocab Model presents a counter-hypothesis to the modularity assumption by claiming that L2 vocabulary performance affects performance in L2 phonetics and phonology. Specifically, the Vocab Model posits that the impact of L2 vocabulary acquisition on L2 speech perception is analogous to the impact of L1 vocabulary acquisition on L1 speech perception (Bundgaard-Nielsen et al., 2011a).

Developed within the framework of the Perceptual Assimilation Model (Best, 1995), the Vocab Model claims that the language learning

processes and mechanisms applied in L1 acquisition remain available at all points in life, making L1 and L2 acquisition essentially similar processes with different starting points. Both L1 and L2 learners must learn to attend to those phonetic differences that are phonemic in the language of acquisition (phonological distinctiveness) while ignoring those differences that are not phonemic (phonological constancy). However, whereas the starting point for L1 speech acquisition is the abstract organisation of phones, L2 acquisition takes prior linguistic experience as its starting point. Consequently, early L2 perception is based on the learner's native language (Bundgaard-Nielsen, Best, & Tyler, 2011b).

Central to the Vocab Model is the Lexical Growth Hypothesis claiming that initial lexical growth facilitates L2 rephonologisation in much the same way as the lexical spurt facilitates the establishment of phonological constancy in L1 acquisition. Infants show phonological distinctiveness for vowels (Kuhl, Williams, Lacerda, Stevens, & Lindholm, 1992) around the age of six months and for consonants around the age of 10-12 months (Best & McRoberts, 2003). However, they do not show phonological constancy until the age of 19 months (Best, Tyler, Gooding, Orlando, & Quann, 2009), i.e. around the onset of the lexical spurt, typically between the ages of 14 months and 22 months (Reznick & Goldfield, 1992). This argument may be extended to L2 acquisition; L2 comprehension requires L2 learners to differentiate between an increasing number of contrasting L2 words, some of which initially sound homophonous to the L2 learner, that is, the need for successful L2 comprehension drives the need to rephonologise.

In two studies of vowel perception and vocabulary size in L1 Japanese learners of Australian English, Bundgaard-Nielsen et al. (2011a, b) found empirical support for the Vocab Model. Specifically, L1 Japanese learners with vocabularies above 6,000 word families<sup>3</sup> were found to be more consistent in their assimilation of Australian English vowels to Japanese (Bundgaard-Nielsen et al., 2011a; 2011b) and more accurate in discriminating phonemic vowel contrasts in Australian English. Moreover, increased L2 exposure was not found to improve L2 vowel perception for L2 learners whose vocabularies were above 6,000 word families at the first point of testing (Bundgaard-Nielsen et al., 2011b), suggesting that increased vocabulary facilitates L2 rephonologisation up to the point of 6,000 word families, above which point L2 vocabulary size does not impact L2 speech perception further.

---

<sup>3</sup> A word family consists of a lexical root along with its derivations and inflections (Schmitt, 2010, p. 8).

This study extends the empirical test of the Vocab Model to L1 Danish and L1 Finnish learners of English, investigating identification of both vowels and consonants. Differences in the acquisition trajectories of vowels and consonants are expected, since perceptual attunement to vowels in L1 acquisition happens four months earlier than to consonants (Kuhl et al., 1992; Best & McRoberts, 2003).

### **1.5 Aim and scope**

The study adopts an experimental approach to the Conrad Phenomenon by examining the relationship between L2 performance within the three linguistic domains; syntax, phonetics and phonology, and the lexicon. These domains have been chosen for three reasons. First, syntax, phonetics and phonology, and the lexicon are considered crucial domains to master for L2 learners of English. For some languages, morphology would also be considered crucial for L2 learners, but for English, morphology is arguably less important than the other three domains. Second, syntax, phonetics and phonology, and the lexicon differ in a number of ways suggesting qualitative differences in processing; syntax and phonology are primarily rule-based while the lexicon is item-based, syntax and phonology are purely linguistic and present a finite set of entities, while the lexicon is related to world-knowledge and presents an open-ended learning task, and finally, phonetics and phonology, contrary to the other two domains, contains a physiological motor-aspect. Third, the prior research motivating this study all centres on two or three of the following domains: morphosyntax, phonology, and the lexicon. Yet, adopting a modularity approach, morphosyntax seems problematic as one domain, because linguistics traditionally views morphology and syntax as two separate though connected domains of language (e.g. McCabe, 2011, p. 169; Akmajian, Demers, Farmer, & Harnish, 2010, pp. 3-4; Morrish, 2015, p. 18). This study therefore examines syntax instead of morphosyntax.

As outlined above, the Modularity Account holds that there is modularity in L2 performance related to linguistic domains, so that an L2 learner's performance in one linguistic domain is independent from the learner's performance in other linguistic domains, hence accounting for the discrepancy between Joseph Conrad's written and spoken English by claiming independence between L2 performance within syntax and the lexicon on the one hand and phonetics and phonology on the other hand. However, one can also imagine an alternative account of the Conrad Phenomenon, one that I will call the Inverse Relation Account. Imagine

that L2 learners who perform well in domain X, tend to perform poorly in domain Y and vice versa. Following this line of thought, the Conrad Phenomenon can be accounted for by claiming an inverse relationship between L2 performance within syntax and the lexicon on the one hand and L2 performance within phonetics and phonology on the other hand. These two alternative accounts of the Conrad Phenomenon are investigated.

If domain-related modularity or inverse relationships are observed, this study examines whether it may be specific to the learners' native language or directly caused by the characteristics of the L2. A further question of interest is whether such modularity or inverse relationships, if existing, vary with degree of L2 experience, such that, e.g. domains for which performance is independent for Less Experienced L2 learners show related performance for More Experienced L2 learners. Moreover, the study investigates the possibility of an interaction between degree of L2 experience and L1 background, such that modularity or inverse relationships in L2 performance depend on the combination of second language experience and native language. The study moreover examines whether the present data support the Lexical Growth Hypothesis from the Vocab Model, a model claiming that the Conrad Phenomenon is uncommon among L2 learners.

The relationship between L2 performance in the different domains may show a number of different patterns. First, performance in the three domains may not be correlated, suggesting that performance within different domains is independent, i.e. suggesting modularity in L2 performance. However, the lack of a statistically significant correlation does not imply independence, since absence of evidence is not evidence of absence<sup>4</sup>, and clear non-correlational patterns must be observed in the data in order to argue for modularity. One such pattern could be a complete lack of systematicity, i.e. data showing a large number of different scores on domain X for any score on domain Y and vice versa. Alternatively, the score on domain Y could be almost constant for different scores on domain X, which would be the case if a ceiling or a floor effect is observed.

Second, performance within the three domains may be positively correlated, suggesting a positive relationship between linguistic domains in L2 acquisition, so that performing well in one domain positively affects performance in other domains. Hence, a statistically significant positive correlation between L2 performance in all three domains could be evi-

---

<sup>4</sup> The failure to reject a null hypothesis does not imply the acceptance of the null hypothesis. Hence non-significant results are inconclusive (Altman & Bland, 1995).

dence for some degree of interdependence between all three domains, i.e. evidence against modularity. Alternative accounts of positive correlations between domains include general intelligence or language aptitude, which have been found to be related but different constructs (Li, 2015). Studies on the effect of intelligence on second language learning are scarce and intelligence have been found to be a poor predictor of L2 performance (e.g. Sparks, Patton, Ganshow, & Humbach, 2009; Ganschow, Sparks, Javorsky, Pohlman & Bishop-Marbury, 1991). The very few existing studies that examine the effect of intelligence on different aspects of language show evidence that general intelligence affect some, but not all, aspects of foreign language learning. Genesee (1976) found that general intelligence is positively correlated with scores on academic L2 skills but shows no relationship with interpersonal communication skills. More recently, Sparks et al. (2009) found that, among a list of different L2 skills, general intelligence affected L2 word decoding only. Language aptitude, defined as ‘a number of cognitive factors making up a composite measure that can be referred to as the learner’s overall capacity to learn a foreign language’ (Dörnyei, 2005, pp. 33-34), is generally accepted to be componential rather than unitary (e.g. Dörnyei, 2005, p. 33) and research has found that different components of language learning aptitude impact L2 performance in different linguistic domains (e.g. Sparks, Patton, Ganschow & Humbach, 2011; Saito, 2017). Moreover, research (Li, 2015) shows that overall language aptitude has no impact on L2 vocabulary acquisition. Unfortunately, the roles of intelligence and language learning aptitude are outside the scope of this study.

Third, performance within two of the domains may be positively correlated but uncorrelated with performance in the third domain, suggesting some degree of interdependence between these two domains but providing no conclusion regarding the interdependence between the two correlated domains on the one hand and the third domain on the other hand. Such a finding would call for further research into the aspects that are shared between the two correlated domains but not shared with the third domain in order to better understand what drives the correlation.

Finally, performance within two domains may be inversely or negatively correlated, suggesting an inverse interdependence between these two domains, so that learners who perform well within one of the domains tend to perform poorly within the other domain and vice versa. Such a negative correlation is evidence against modularity, since it suggests some sort of interdependence between domains. However, such a result

might offer an alternative account of the Conrad Phenomenon, namely the Inverse Relation Account; if a negative correlation is observed between phonetics and phonology on the one hand and syntax and the lexicon on the other.

### **1.6. L1 Danish and L1 Finnish learners of English**

L1 Danish and L1 Finnish learners of English were chosen because Denmark and Finland offer similar learning environments while the linguistic differences between Danish and Finnish vis-à-vis English are considerable. Observed difference in L2 performance between L1 Danish and L1 Finnish learners are therefore likely to be due to language background rather than differences in learning environment.

When the L1 Danish participants went to school, English instruction was obligatory from 3<sup>rd</sup> to 9<sup>th</sup> grade of elementary school (Ministry for Children, Education and Gender Equality 2014) and in the first two years of upper-secondary school (Ministry for Children, Education and Gender Equality 2013). The Finnish participants were taken from the 91% of students who choose English as their first second language and received English instruction from 3<sup>rd</sup> to 9<sup>th</sup> grade in elementary school and in upper-secondary school (Leppänen, Pitkänen-Huhta, Nikula, Kytölä, Törmäkangas, Nissinen, and Kääntä, 2011). Moreover, the inhabitants in both countries are exposed to a fair amount of anglophone media on a daily basis, as foreign TV programs are interlingually subtitled rather than dubbed in both countries, and as anglophone soap operas, films, and pop music are pervasive, especially in youth culture (Preisler, 1999; Leppänen and Nikula, 2007).

Within all three domains of interest, Danish shows a fair amount of similarities with English, while Finnish has comparatively few similarities with English. This is in part due to historical relatedness. Old English and the ancestor of Danish, Old Norse, both descend from Proto-Germanic (Strang, 1970, p. 376; Herslund, 2002, p. 1). Consequently, Danish and English share a substantial number of common Germanic words, most of which are still alike in both meaning and form. As a Finno-Ugric language, Finnish shares no real cognates with English, and the only lexical similarities between English and Finnish are due to direct and indirect (primarily via Swedish) borrowings (Pulkkinen, 1989; Karlsson, 1999, p. 7). Broadly speaking, the syntax of Danish is very similar to that of English, as both are highly analytical languages (van Gelderen, 2006, pp. 214-220; Herslund, 2002, p. 79). Finnish, on the other hand, is a synthetic language with

an extensive case system (Karlsson 1999, pp. 4-6). With respect to the sound system, the difference in sheer size is noteworthy. At the phonemic level, English has 15 or 16 stressed vowels, of which 10 or 11 are monophthongs, depending on the variety (Ladefoged and Disner 2012, pp. 29-30, 133-134). Most varieties of English have 24 consonant phonemes, of which 23 occur in initial position (Cruttenden 2014, pp. 161, 211). The Danish vowel inventory is extensive and complex with at least 20 stressed phonemic monophthongs, organised into 10 short-long pairs, and extended allophonic variation, and Danish has 16 initial consonant phonemes (Bassbøll, 2005, pp. 50, 64; Grønnum, 1998). Finnish has 16 phonemic monophthongs, which can be organised into eight short-long pairs. (Wiik, 1965, pp. 40-44), while the reported range of initial consonants is between 11 and 17, depending on how loaned phones are treated (see Suomi, Toivanen, and Ylitalo 2008, pp. 24-25 for an overview).

## **2. Methods**

Modularity in L2 performance was examined by having More Experienced and Less Experienced L1 Danish and L1 Finnish learners of English and a group of L1 English speakers complete a set of tasks measuring performance in English syntax, phonetics and phonology, and lexicon.

### **2.1 Participants**

Three groups of participants were tested; 41 L1 Finnish learners of English (6 males, 35 females, mean age = 25.17 years), 41 L1 Danish learners of English (8 males, 33 females, mean age = 24.71 years), and 14 native English speakers functioning as a baseline group (2 males, 12 females, mean age = 20.65 years).

The L1 Finnish learners of English all lived in and around Jyväskylä, Central Finland. The L1 Finnish participants were divided into two groups: 1) 21 More Experienced Learners: students of English who had lived in an English-speaking country for a longer period (range: 2.5 months to 3 years, mean = 10.02 months), and 2) 20 Less Experienced Learners: students of Finnish who had not lived in an English-speaking country for a longer period.

The L1 Danish learners of English all lived in and around Aarhus, East Jutland, Denmark. The L1 Danish participants were also divided into two groups: 1) 20 More Experienced Learners: participants who had lived in an English-speaking country for a longer period (range: 4 months to

2.17 years, mean = 10.73 months), and 2) 21 Less Experienced Learners: participants who had not lived in an English-speaking country for a longer period. 14 of the L1 Danish More Experienced Learners and 15 of the L1 Danish Less Experienced Learners were students of English at Aarhus University. The remaining participants were students of other subjects at Aarhus University or non-students.

The native English speaker baseline group consisted of students at Bangor University, Wales, who were speakers of standard Southern British English.

None of the subjects reported any hearing problems.

## **2.2 Tasks**

The study consisted of five tasks: 1) a delayed repetition task, 2) a vowel identification test, 3) a consonant identification test, 4) a Grammaticality Judgement test, and 5) a vocabulary test. The aims and forms of the five tasks are briefly outlined below.

### **The delayed repetition task**

The delayed repetition task aimed at assessing the subjects' production of English. Subjects were asked to repeat five sentences spoken by a native speaker of Southern British English (SBE) in a question-answer framework, as illustrated in (1).

- (1) SBE speaker: *What did Paul eat?*  
SBE speaker: *Paul ate carrots and peas.*  
SBE speaker: *What did Paul eat?*  
**Subject:** *Paul ate carrots and peas.*

Recordings were rated twice for degree of foreign accent on a Likert-scale ranging from 1 (*No foreign accent*) to 9 (*Heavy foreign accent*), by six native speakers of English (2 male and 1 female speakers of American English, and 1 male and 2 female speakers of British English, mean age 26.2 years), who had no prior training in linguistics. Foreign accent was defined for the raters as *non-native accents of English*.

### **The vowel identification test**

The vowel identification test aimed at assessing the subjects' perception of SBE vowels. The TP stimulus presentation software (Rato, Rauber,



Kluge, & Santo, 2013) was used to present listeners with 2 randomizations of the 11 monophthongs of SBE in a /hVt/ context. Vowel stimuli were recorded from two male, native speakers of SBE. Subjects were asked to identify the vowel among the 11 options given by the 11 monophthongs of SBE, presented orthographically as <heat, hit, het (up)<sup>5</sup>, hat, heart, hoot, hUt, haught(y), hot, hurt, hut>. Since no real /hut/ word exists in English, participants were introduced to the non-word <hUt>.

### **The consonant identification test**

The consonant identification test is similar to the vowel identification test. The TP stimulus presentation software was used to present listeners with 2 randomizations of the 20 initial consonants of English in a /Ca/ context. Consonant stimuli were taken from a corpus of American English /Ca/ syllables made available by Shannon, Jensvold, Padilla, Robert, and Wang (1999). Three tokens of each consonant were selected from two male, native speakers of American English. Subjects were asked to identify the consonant among the 20 options given by the 20 English initial consonants, presented orthographically as <P, B, T, D, K, G, F, V, Think, Them, S, Z, Ship, Genre, Chin, Joke, W, L, R, Yes>.

### **The Grammaticality Judgement test<sup>6</sup>**

The grammaticality judgement accessed the participants' intuitions on English syntax in embedded and main clause negations, *wh*-questions, and *yes-no* questions. The test consisted of a corresponding set of grammatical and ungrammatical sentences, which the subjects were asked to judge as *Correct* or *Incorrect* with respect to grammar.

### **The vocabulary test**

The Vocabulary Size Test (Nation & Beglar, 2007; Nation, 2012) was used as an indicator of the subjects' vocabulary size. It is a multiple-choice definition test, in which the tested word is presented in a simple, non-defining context, and four different but semantically related definitions are supplied, of which one is correct. The subjects' task is to choose the right definition among the four options. (2) shows an item from the vocabulary test.

---

<sup>5</sup> *Het up* means *anxious, exited or slightly angry* (Deuter, Bradbery, & Turnbull, 2015).

<sup>6</sup> The results of the Grammaticality Judgement test are presented in Horslund (2016), which also outlines and motivates the structure of the test.

- (2) soldier: He is a **soldier**.
- a. person in a business
  - b. person who studies
  - c. person who uses metal
  - d. person in the army

Correct answer: d

### **2.3 Procedure**

For practical reasons, the five tasks were divided into two sessions, one consisting of the three sound-related tasks and another consisting of the Grammaticality Judgment task and the vocabulary test. The two sessions were conducted on different days or with a couple of hours in between for all participants in Jyväskylä and for the majority of participants in Bangor. The remaining participants in Bangor and all participants in Aarhus completed both sessions in one go with a short break between the two sessions. The order of the two sessions as well as of the tasks within them were counterbalanced across participants, except for the delayed repetition task, which always preceded the phoneme identification tasks in order to obtain speech recordings that were unaffected by the focus on segmentals possibly induced by the phoneme identification tests.

All participants participated voluntarily, and the participants in Jyväskylä and Bangor received lunch coupons, a movie ticket, or a monetary compensation for participating in the study. Subjects in Aarhus received no compensation for participating in the study.

### **2.4 Statistical analyses**

Relationship between performance in different domains was tested by means of Person correlation tests. The Vocab Model was tested by means of Mixed effect models. Mixed effect models are regression models that model the random variation between participants and items, thus dealing with the dependencies between observations in the model rather than by taking means. Mixed effect models constitute an alternative to both ANOVA and ordinary regression and offers a number of advantages to these models (see Jaeger, 2008; Cunnings, 2012). All p-values are Holm corrected (Holm, 1979) to avoid inflating the Type I error rate (the rate of false positives) by multiple comparisons.

All statistical analyses were conducted in the software program R (R Core Team, 2015). The R packages used were *lme4* (Bates, Maechler, Bolker, Walker, Christensen, Singmann, Dai, & Grothendieck, 2015) and *optimx* (Nash, 2014) for the construction of mixed effects models, *multcomp* (Hothorn, Bretz, & Westfall, 2008) for pairwise comparisons of parameters in mixed effects models, and *Hmisc* (Harrell, 2013) for correlations. Graphs were also constructed in R, by means of the package *ggplot2* (Wickham, 2009).

### 3. Results

This section first presents data on between-domain relations and subsequently a test of the Vocab Model.

#### 3.1 Between-domain relations

Between-domain relations are examined by means of Pearson correlation tests between Phoneme Identification scores (vowels and consonants combined) and mean Foreign Accent ratings representing L2 speech perception and production in the domain of phonetics and phonology, scores on the Grammaticality Judgement test representing L2 performance in the domain of syntax, and scores on the Vocabulary Size test representing L2 performance in the lexical domain. The scores for vowel and consonant identification are combined, since these are both measures of L2 speech perception. Foreign Accent ratings are kept separate from the perception scores, since Foreign Accent ratings measure production.

Pearson correlation tests on the L2 learner data for Foreign Accent, Phoneme Identification, Grammaticality Judgement, and Vocabulary revealed significant across-group correlations between all tasks. All correlations were positive except those with Foreign Accent, which were all negative, since high Foreign Accent scores indicate poor pronunciation and low Foreign Accent scores indicate good pronunciation. This suggests that all relationships between tasks are positive. Correlations across all L2 groups thus indicate that performance in one linguistic domain generally goes hand in hand with performance in other linguistic domains. However, there were considerable differences in the strength of the correlations between different tasks, and correlation tests within L2 groups did not always reach significance. Table 1 provides an overview over the between-domain correlations across and within L2 groups.

	L1 Danish learners		L1 Finnish learners		Across L2 groups
	More Experienced	Less Experienced	More Experienced	Less Experienced	
<b>Phon by Vocab</b>	0.6309 (0.0171)	0.7751 (0.0002)	0.3481 (0.7324)	0.5658 (0.0466)	0.6118 (<0.0001)
<b>FA by Vocab</b>	-0.3720 (0.3187)	0.1898 (0.3452)	-0.2860 (1.0000)	-0.4354 (0.1650)	-0.6075 (<0.0001)
<b>Phon by GJ</b>	0.6303 (0.0171)	0.3816 (0.3452)	-0.0148 (1.0000)	0.5397 (0.0562)	0.4896 (<0.0001)
<b>Phon by FA</b>	-0.1429 (1.0000)	-0.4409 (0.2271)	-0.0618 (1.0000)	-0.6030 (0.0293)	-0.4016 (0.0006)
<b>Vocab by GJ</b>	0.5073 (0.0897)	0.1898 (0.8198)	0.2069 (1.0000)	0.3385 (0.2887)	0.3559 (0.0021)
<b>FA by GJ</b>	-0.0774 (1.0000)	-0.1388 (0.8198)	-0.0324 (1.0000)	-0.3303 (0.2887)	-0.2191 (0.0480)

Phon: Phoneme Identification, FA: Foreign Accent, GJ: Grammaticality Judgment Test, Vocab: Vocabulary Test

Table 1. Between-task correlations for More Experienced and Less Experienced L1 Danish and L1 Finnish learners of English and native English speakers. Pearson correlation coefficients and Holm adjusted p-values in parenthesis. Significant (at the 0.05 level) correlations are highlighted in light blue. Marginally significant ( $p < 0.1$ ) correlations are highlighted in light pink.

The strongest across-group correlations were between Vocabulary and Phoneme Identification (Pearson's  $r=0.6118$ ,  $p < 0.0001$ ) and between Vocabulary and Foreign Accent (Pearson's  $r=-0.6075$ ,  $p < 0.0001$ ). Pearson correlation test for the separate L2 groups support the relationship between Phoneme Identification and Vocabulary. The within-group tests revealed significant, positive correlations between Phoneme Identification and Vocabulary for More Experienced L1 Danish learners (Pearson's  $r=0.6309$ ,  $p < 0.0171$ ), Less Experienced L1 Danish learners (Pearson's  $r=0.7751$ ,  $p < 0.0002$ ), and Less Experienced L1 Finnish learners (Pearson's  $r=0.5658$ ,  $p=0.0466$ ), suggesting that it is common among L2 learners to exhibit a positive relationship between L2 speech perception and L2 vocabulary. The correlation between Foreign Accent ratings and Vocabulary scores did not approximate significance within any of the L2 groups. Figure 1 shows a scatterplot of the relationship between Phoneme Identification scores and Vocabulary scores, separately for each group, and Figure 2

shows a scatterplot of the relationship between Foreign Accent ratings and Vocabulary scores, separately for each group.

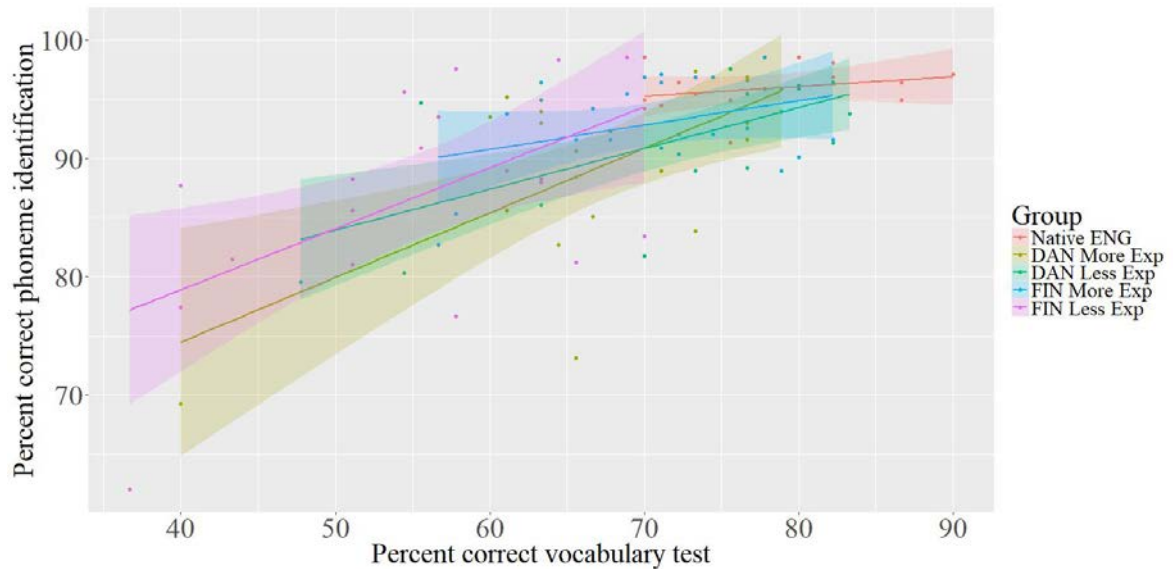


Figure 1. Scatterplot of percent correct in the Vocabulary Test and percent correct Phoneme Identification (vowels and consonants combined) for More Experienced and Less Experienced L1 Danish and L1 Finnish learners of English and the native speaker baseline with 95% confidence intervals (shaded areas) for each group.

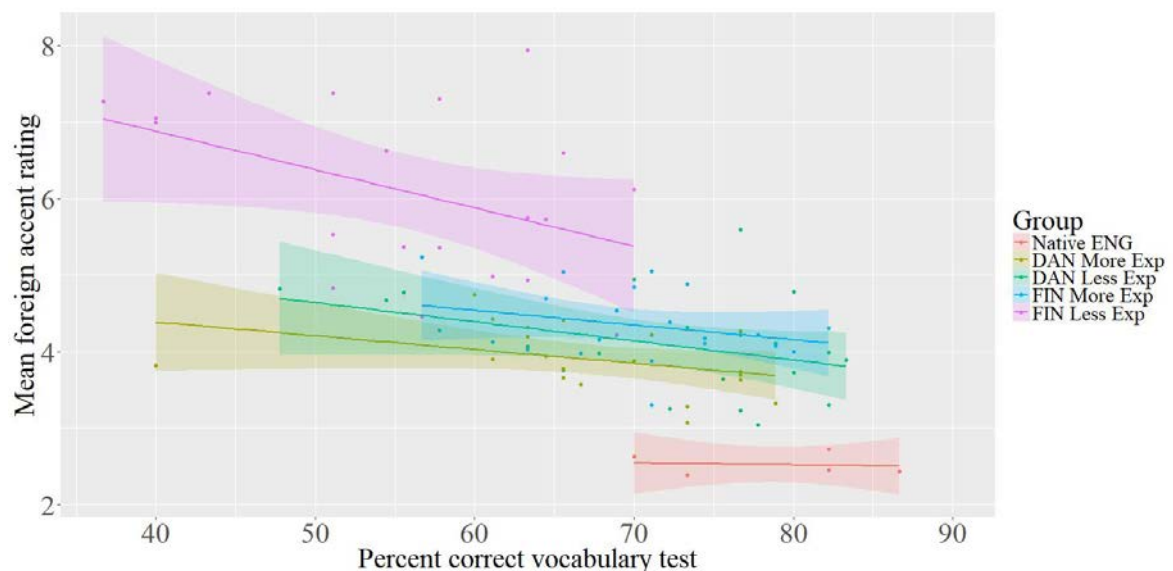


Figure 2. Scatterplot of percent correct in the Vocabulary Test and mean Foreign Accent rating for More Experienced and Less Experienced L1 Danish and L1 Finnish learners of English and the native English speaker baseline with 95% confidence intervals (shaded areas) for each group.

Interestingly the across-group correlations between Phoneme Identification and Foreign Accent on the one hand and Vocabulary on the other hand were both stronger than the correlation between Phoneme Identification and Foreign Accent (Pearson's  $r=-0.4016$ ,  $p=0.0006$ ), despite the fact that Phoneme Identification scores and Foreign Accent ratings represent tasks within the same linguistic domain. However, within-group tests revealed a significant, strong, negative correlation between Phoneme Identification and Foreign Accent for Less Experienced L1 Finnish learners (Pearson's  $r=-0.6030$ ,  $p=0.0293$ ), suggesting a positive relationship between L2 speech perception and production for Less Experienced L1 Finnish learners. Figure 3 shows a scatterplot of the relationship between Phoneme Identification scores and Foreign Accent ratings, separately for each group.

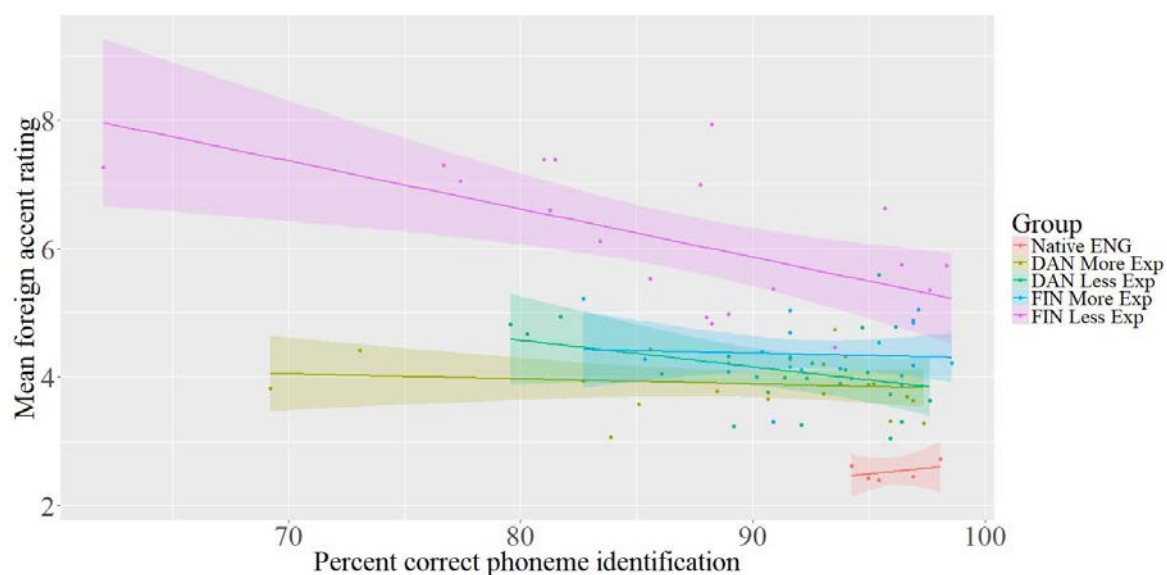


Figure 3. Scatterplot of percent correct Phoneme Identification (vowels and consonants combined) and mean Foreign Accent rating for More Experienced and Less Experienced L1 Danish and L1 Finnish learners of English and the native English speaker baseline with 95% confidence intervals (shaded areas) for each group.

Figure 4, Figure 5, and Figure 6 show scatterplots of the relationships between Grammaticality Judgement on the one hand and Phoneme Identification, Foreign Accent rating, and Vocabulary on the other hand, separately for each group. As is evident from Figure 4, Figure 5, and Figure 6 the amount of variation in the Grammaticality Judgement test is rather limited. The results show a ceiling effect with mean accuracy scores of 96.13% for

the native speaker baseline, 95.41% for L1 Danish learners, and 95.60% for L1 Finnish learners, which may in part explain why Grammaticality Judgement performance seems to be least related to performance on the other tasks. Across-group correlations between Grammaticality Judgement scores and performance on other tasks were moderate to weak (Pearson's  $r \leq 0.489$ ,  $p \leq 0.0480$ ). However, a significant, strong, positive correlation between Phoneme Identification and Grammaticality Judgement was observed for More Experienced L1 Danish learners (Pearson's  $r = 0.6303$ ,  $p = 0.0171$ ), suggesting that More Experienced L1 Danish learners exhibit a positive relationship between L2 speech perception and L2 syntax. Strong, marginally significant correlations were observed between Vocabulary and Grammaticality Judgement for More Experienced L1 Danish learners (Pearson's  $r = 0.5073$ ,  $p = 0.0897$ ), and between Phoneme Identification and Grammaticality Judgement for Less Experienced L1 Finnish learners (Pearson's  $r = 0.5397$ ,  $p = 0.0562$ ). Due to the ceiling effect in the GJ data, all correlations between performance on the GJ test and performance on other tests should be interpreted with caution and can only lead to preliminary conclusions.

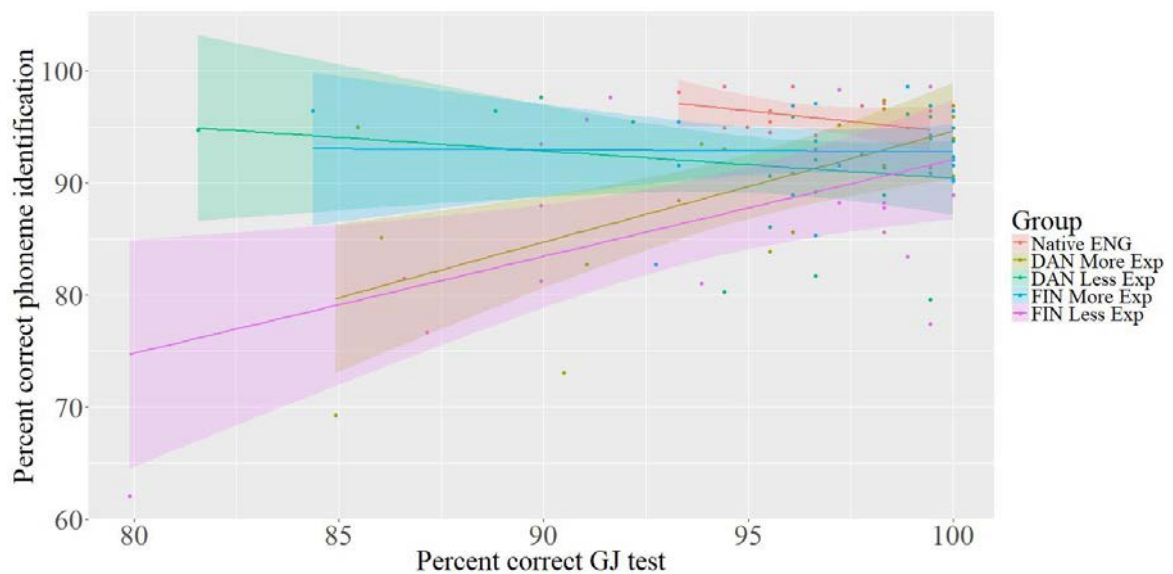


Figure 4. Scatterplot of percent correct Grammaticality Judgement and percent correct Phoneme Identification (vowels and consonants combined) for More Experienced and Less Experienced L1 Danish and L1 Finnish learners of English and the native speaker baseline with 95% confidence intervals (shaded areas) for each group.



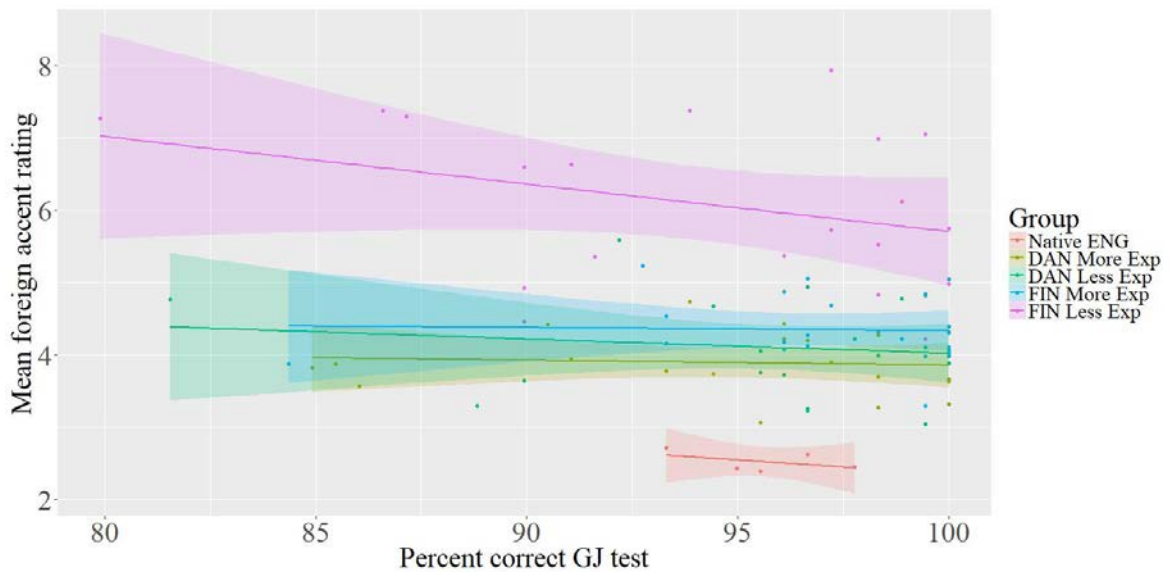


Figure 5. Scatterplot of percent correct Grammaticality Judgement and mean Foreign Accent rating for More Experienced and Less Experienced L1 Danish and L1 Finnish learners of English and the native speaker baseline with 95% confidence intervals (shaded areas) for each group.

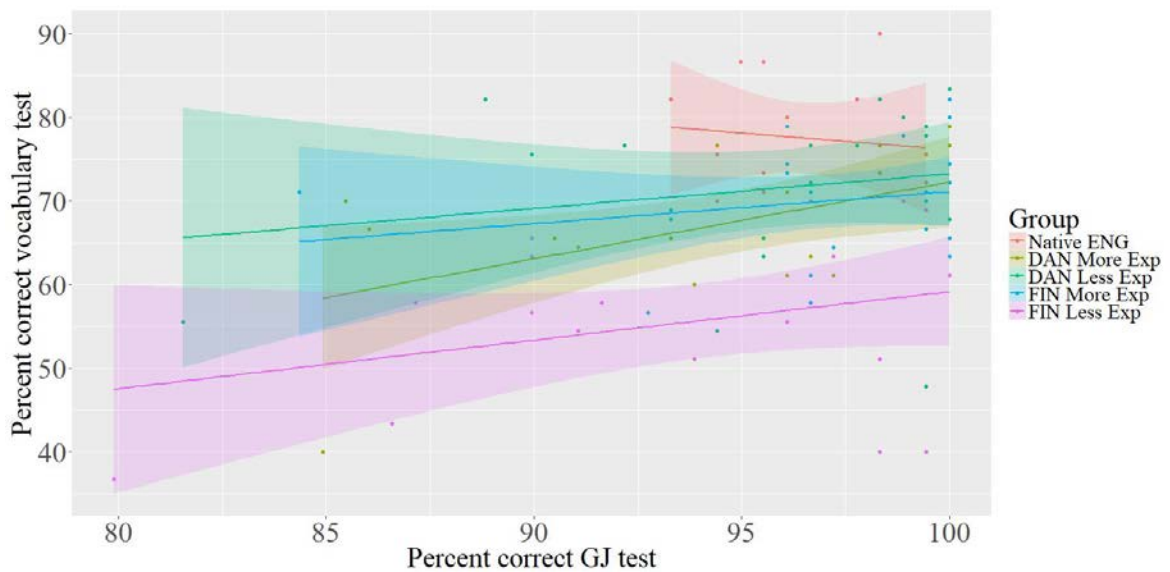


Figure 6. Scatterplot of percent correct Grammaticality Judgement and percent correct in the Vocabulary Test for More Experienced and Less Experienced L1 Danish and L1 Finnish learners of English and the native speaker baseline with 95% confidence intervals (shaded areas) for each group.



### 3.2 Test of the Vocab Model

In order to test the effect of vocabulary score on phoneme identification separately for consonants and vowels for L1 Finnish and L1 Danish learners respectively, a variable combining Category (Vowel/Consonant) and L1 was constructed. A logistic mixed effects model on the L2 learner data for Phoneme Perception with random intercepts for Item and Subject and with Vocabulary Score and the factor combining Category (Vowel/Consonant) and L1 as fixed effects<sup>7</sup> revealed significant Vocabulary effects for L1 Danish learners for both vowels and consonants ( $p \leq 0.0008$ ), and no significant Vocabulary effect for L1 Finnish speakers for either vowels or consonants. The model further revealed that the Vocabulary effect is significantly stronger for Vowels than for Consonants for L1 Danish learners ( $p < 0.001$ ) and L1 Finnish learners ( $p = 0.0226$ ). Table 2 shows an overview over the statistics of this model.

	Estimate	Std. Error	z-value	p-value
Vocab effect for L1 Danish for consonants	0.04631	0.01245	3.718	0.000803
Vocab effect for L1 Danish for vowels	2.84200	0.40936	6.942	2.31e-11
Vocab effect for L1 Finnish for consonants	-0.29299	1.01247	-0.289	1
Vocab effect for L1 Finnish for vowels	0.64531	1.04365	0.618	1
Difference in vocab effect for L1 Danish for consonants versus for vowels	-2.79569	0.41167	-6.791	5.56e-11
Difference in vocab effect for L1 Finnish for consonants versus for vowels	-0.93830	0.35121	-2.672	0.022645

Table 2. Estimates, standard error, z-values, and p-values (Holm adjusted) for the mixed effect model testing the effect of vocabulary score on phoneme perception. Significant effects (at the 0.05 level) are highlighted in light blue.

Figure 7 shows a scatterplot of the relationship between Vocabulary Scores and perception of vowels and consonants separately for L1 Danish learners and L1 Finnish learners.

<sup>7</sup> Model: `glmer (Performance ~ CategoryL1 * VocabScore + (1|Subject) + (1|Item), family = "binomial", data = VocabModel)`.

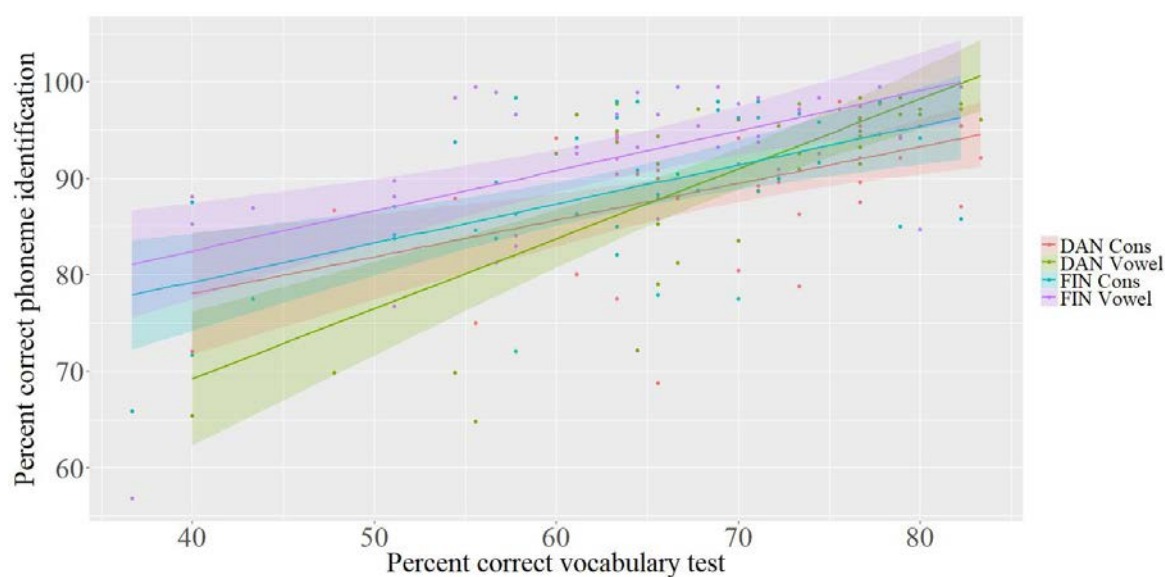


Figure 7. Scatterplot of percent correct Phoneme Identification and percent correct in the Vocabulary Test, divided by L1 and Category (consonant/vowel) with separate 95% confidence intervals (shaded areas) for L1 Danish learners' perception of consonants, L1 Danish learners' perception of vowels, L1 Finnish learners' perception of consonants, and L1 Finnish learners' perception of vowels.

#### 4. Discussion

This study examined relationships between linguistic domains in L2 performance. The main focus of this study was to test whether there is domain-related modularity in L2 performance, that is whether L2 learners generally perform well in one linguistic domain while performing poorly in other linguistic domains. The study moreover asked if the nature of the between-domain relationships depends on the learner's L1, the learner's degree of L2 experience, and/or the combination of these two variables.

Across L2 groups, all correlations between performance on tasks in different linguistic domains were significant, suggesting some degree of interdependence between domains in L2 performance. However, the strength of these between-domain correlations varied, suggesting that some domains are more closely related than others. Specifically, these across-group correlations suggest that the interdependence between the lexical domain and the domain of phonetics and phonology is stronger than the interdependence between the domain of syntax and the other two domains. Yet, the low degree of interdependence between Grammaticality Judgement and the other three tasks may be partly due to the low degree of inter-subject variation in the GJ data (i.e. the ceiling effect).

Interestingly, the strength and significance of between-task correlations varied considerably among the four L2 groups, suggesting that between-domain patterns in L2 performance are affected by the combination of L1 background and degree of L2 experience. For More Experienced L1 Danish learners, all three domains seem to be related to some degree, though the production aspect of phonetics and phonology does not seem to be related to other linguistic tasks. For Less Experienced L1 Danish listeners, there seems to be a relationship between phoneme perception and vocabulary. The data do not suggest any other relationships between tasks, suggesting that the domain of syntax is relatively independent from vocabulary as well as from phonetics and phonology in Less Experienced L1 Danish learners. However, the lack of a significant correlation between the syntax test and other tasks may in part be due to the ceiling effect in the GJ data. For More Experienced L1 Finnish learners the data do not suggest any relationships between tasks, suggesting that the four tasks are relatively independent in More Experienced L1 Finnish learners. Finally, for Less Experienced L1 Finnish learners, there seems to be a relationship between phoneme perception and foreign accent, between phoneme perception and vocabulary, and between phoneme perception and syntax, while syntax seems unrelated to vocabulary and foreign accent in Less Experienced L1 Finnish learners. Again, all interpretations involving the syntax test should be treated as tentative due to the ceiling effect in the syntax data. The observed pattern suggests that the amount of between-domain interdependence increases with L2 experience for L1 Danish learners and decreases with L2 experience for L1 Finnish learners. This difference in the relationship between L2 experience and between-domain interdependence may be related to the linguistic differences between Danish and Finnish vis-à-vis English, since learning context for the L1 Finnish and L1 Danish learners was similar. Perhaps some relationships between domains are not established until later stages of L2 acquisition, while other relationships dilute at later stages. This process may likely interact with the specific L1-L2 differences and similarities. Further studies are needed in order to confirm the observed group differences and further explore the interaction between L1 background and L2 experience in between-domain relationships. Importantly, such studies should include a syntax test that better distinguishes between different levels of performance in this domain. Figure 8 illustrates the observed between-task relationships across and within L2 groups.

As Figure 8 illustrates, there seem to be considerable differences among L2 groups with respect to the relationships between tasks with the exception of a general pattern suggesting some degree of interdependence between L2 Vocabulary and L2 Phoneme Identification and a less general pattern suggesting some degree of interdependence between Grammaticality Judgement and Phoneme Identification, which is uncertain given the ceiling effect. The observed relationships between the lexical domain and the syntactic domain on the one hand and the domain of phonetics and phonology on the other hand contradict the Modularity Account of the Conrad Phenomenon. The present data thus suggest that Joseph Conrad was an exceptional L2 learner in exhibiting such a low degree of interdependence between the lexical domain and the domain of phonetics and phonology and to a lesser extent in exhibiting such a low degree of interdependence between the syntactic domain and the domain of phonetics and phonology.

The study also considered an alternative account of the Conrad Phenomenon, i.e. the Inverse Relation account holding that the discrepancy observed in the level of Joseph Conrad's English syntax and vocabulary on the one hand and his pronunciation of English on the other hand is due to an inverse relation between L2 performance in the domains of syntax and the lexicon on the one hand and in the domain of phonetics and phonology on the other hand. The study therefore examined whether there are inverse relations between linguistic domains in L2 performance, and if so whether their nature depends on the learner's L1, the learner's degree of L2 experience, and/or the combination of these two variables. The present data revealed no inverse relations between L2 performance in different linguistic domains either across or within L2 groups, suggesting that domain-related inverse relationships are not the norm in L2 performance. However, as mentioned above, absence of evidence does not imply evidence of absence, and the present data cannot rule out the existence of domain-related inverse relations in L2 performance. Nevertheless, it is safe to assume that domain-related inverse relations are not common in L2 performance. Consequently, the Inverse Relation Account of the Conrad Phenomenon is not supported by the present data.

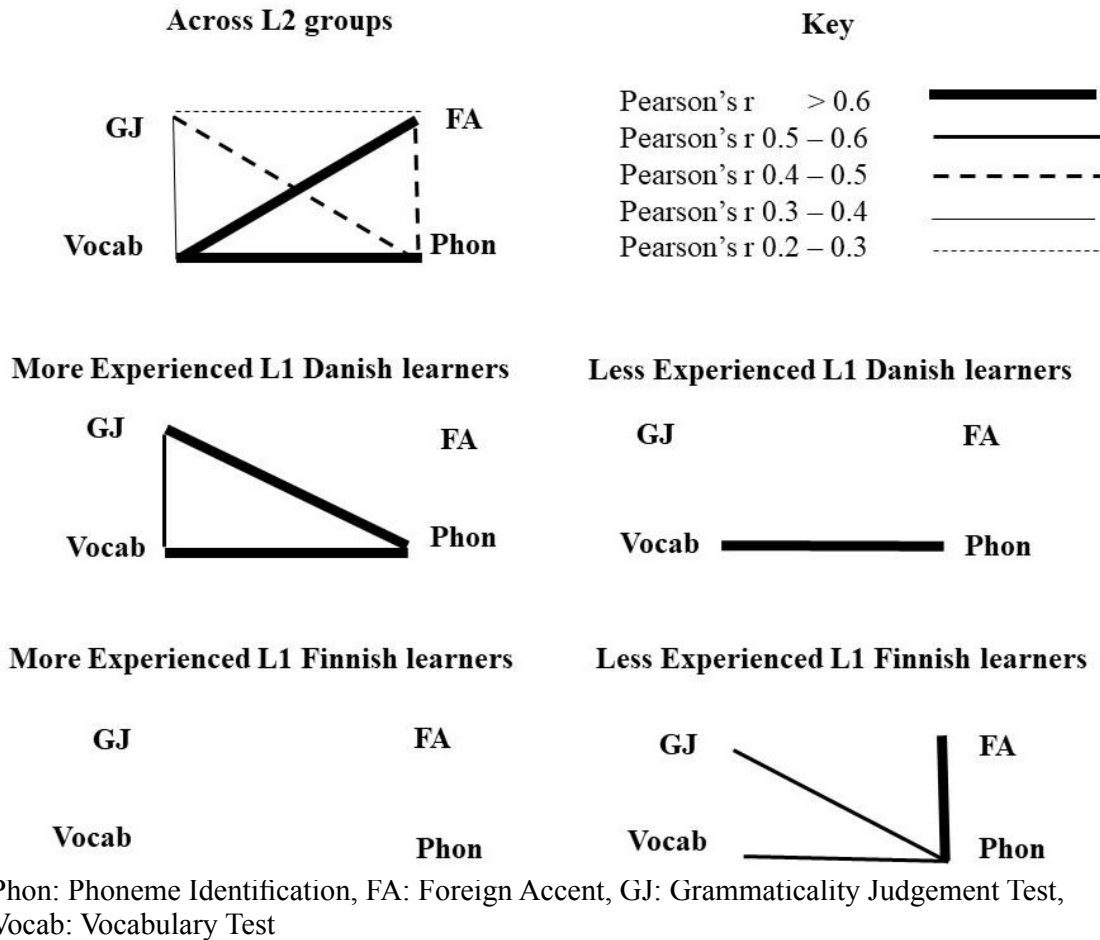


Figure 8. Illustration of between-task relationships. Significant (at the 0.05 level) and marginally significant ( $p < 0.1$ ) between-task correlations across and within L2 groups are marked with lines corresponding to the strength of the correlation as measured by the correlation coefficient.

#### 4.1. A test of the Lexical Growth Hypothesis of the Vocab Model

The study furthermore examined whether the present data support the Lexical Growth Hypothesis positing a positive relationship between L2 vocabulary size and L2 speech perception. The Lexical Growth Hypothesis is partly supported by the present data. A logistic mixed effects model revealed a significant effect of Vocabulary Score on Phoneme Identification in L1 Danish learners but not in L1 Finnish learners. For L1 Danish learners, the effect of Vocabulary Score on Phoneme Identification was significant for both Consonants and Vowels. In accordance with the Vocab Model, this effect of Vocabulary Score on Phoneme Identification may be interpreted

as lexical facilitation of L2 speech acquisition. However, the Vocab Model only predicts lexical facilitation in L2 speech acquisition for L2 learners with vocabularies below the cut-off point of approximately 6,000 word families, beyond which vocabulary size should no longer matter for L2 speech acquisition. Interestingly, it nevertheless seems to do so for the L1 Danish participants in the present study. All participants have estimated vocabulary sizes above approximately 6,000 word families. The present data thus suggest that lexical facilitation in L2 speech acquisition may persist beyond a vocabulary size of approximately 6,000 word families. Future research should investigate whether the cut-off point for lexical facilitation in L2 speech perception is dependent on the L1-L2 mapping, and if so how this variation may be explained, perhaps in terms of functional load, i.e. ‘a measure of the work which two phonemes (or a distinctive feature) do in keeping utterances apart’ (King, 1967, p. 831), of problematic contrasts. Moreover, a difference in strength of lexical facilitation was suspected for Consonants and Vowels, and this prediction was borne out by the present data. The logistic mixed effects model revealed a significantly stronger Vocabulary Score effect for Vowels than for Consonants in both L1 Danish and L1 Finnish learners.

#### **4.2. Relations between domains in first language acquisition revisited**

One of the motivations for investigating relations between linguistic domains in L2 performance was to inform the modularity debate in L1 acquisition, which is difficult to settle because language development and the development of world knowledge are naturally confounded in L1 acquisition. Since L2 acquisition does not suffer from this confound, L2 data on relations between linguistic domains may help illuminate whether the observed strong positive correlation between lexical and syntactic development in L1 acquisition is due to this confound. The present L2 data suggest that the observed correlation between syntax and lexicon in L1 acquisition may indeed be related to the co-development of language and world knowledge. A significant correlation between syntax and lexicon was observed in the across-group analyses, but this correlation was quite weak, and within-group analyses revealed a significant correlation between syntax and lexicon in one L2 group only, suggesting that a strong correlation between syntactic performance and lexical performance is not necessarily present in language acquisition, though the weakness of this correlation may in part be due to the low degree of inter-subject variability on the Grammaticality Judgement test (i.e. the ceiling effect). Consequently, in

order to argue for a linguistic account, in opposition to a world knowledge account, of the L1 correlation between syntax and lexicon, one has to claim a difference between L1 and L2 acquisition in this respect.

The weak relationship observed between L2 syntax and L2 lexicon may be interpreted as supporting a Dual Mechanism Account of L2 acquisition of these two domains. This would be in line with a previously mentioned study (Bowden, Steinhauer, Sanz, & Ullman, 2013) suggesting qualitative differences in neural processing of L2 syntax and L2 lexicon. Interestingly, the observed weak relationship between L2 syntax and L2 lexicon contradicts the results of Snow and Hoefnagel-Höhle (1979), who identified lexicon and morphosyntax as one single factor in L2 performance. This discrepancy between the present results and those of Snow and Hoefnagel-Höhle may be due to differences in L1-L2 mappings or to the ceiling effect in the present syntax test, but more research is required to settle this as the present study and that of Snow and Hoefnagel-Höhle are not directly comparable.

Along with the Vocab Model and its previous empirical support, (Bundgaard-Nielsen et al., 2011a; 2011b), the present data further suggests a relationship between L2 lexical development and the development of L2 speech acquisition. The Vocab Model is based on L1 acquisition patterns and supported for L2 acquisition in Bundgaard-Nielsen et al. (2011a; 2011b) and by the present data. Consequently, research on modularity in L1 and L2 acquisition might benefit from bringing phonetics and phonology into the equation.

## **5. Conclusion**

This study examined relationships between L2 performance in the domains of syntax, the lexicon, and phonetics and phonology in order to explore whether Joseph Conrad was an exceptional L2 learner in performing well in syntax and the lexicon and poorly in phonetics and phonology. Two competing accounts of the Conrad Phenomenon were tested. The Modularity Account claims independence between L2 performance in different linguistic domains, and the Inverse Relation Account claims an inverse relation between L2 performance in the domains of syntax and the lexicon on the one hand and the domain of phonetics and phonology on the other hand. The present data found support for neither account. Some degree of domain-interdependence was observed, though this interdependence varied considerably among L2 groups differing in L1 background and degree

of L2 experience. Across L2 groups, the strongest between-domain relationship was observed between the lexical domain and the domain of phonetics and phonology. Though syntax exhibited the weakest relationships with other domains, which may in part be due to the ceiling effect in the syntax data, the study found a strong relationship between L2 syntax and L2 speech perception in two L2 groups and a moderately strong relationship between L2 syntax and L2 vocabulary in one L2 group. No inverse relations between linguistic domains were observed in the data. Hence, the present data suggest that Joseph Conrad was indeed an exceptional L2 learner in performing well in syntax and the lexicon and poorly in phonetics and phonology. The general trend in L2 performance shows some degree of positive relation between linguistic domains. Future research should further examine between-domain relationships in L2 learners differing in L1 and degree of L2 experience in order to better account for the between-group differences in between-domain relationships.

The study further tested the Vocabulary Growth Hypothesis from the Vocab Model, which claims a lexical facilitation effect in L2 speech perception. The present study found significant lexical facilitation in L1 Danish learners only. Interestingly, lexical facilitation was stronger for vowels than for consonants in both L1 Danish and L1 Finnish learners. Lexical facilitation is predicted to occur only with vocabularies below 6,000 word families, but the results of the present study suggest that lexical facilitation effects may persist beyond 6,000 word families. The results from the current study along with results from previous studies on the Vocab Model suggest a strong positive relationship between L2 vocabulary and L2 speech perception, and future studies on between-domain relations in L2 performance may therefore benefit from including the domain of phonetics and phonology instead of examining only syntax and the lexicon.

## References

- Akmajian, A., Demers, R. A., Farmer A. K., & Harnish, R. M. (2010). *An Introduction to Language and Communication*. Sixth edition. Cambridge and London: MIT Press Linguistics.
- Altman, D. G., & Bland, J. M. (1995). Absence of evidence is not evidence of absence. *BMJ*, 311, 485.



- Anisfeld, M., Rosenberg, E. S., Gasparini, D. & Hoberman, M. J. (1998). Lexical acceleration coincides with the onset of combinatorial speech. *First Language*, 18(53), 165-184.
- Basbøll, H. (2005). *The Phonology of Danish*. Oxford, Oxford University Press.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B. Henrik Singmann, Dai, B., & Grothendieck, G. (2015). *lme4: Linear mixed-effects models using 'Eigen' and S4. R package version 1.1-10*. [cited 02/11 2015] <<http://CRAN.R-project.org/package=lme4>>.
- Bergeron, V. (2007). Anatomical and functional modularity in cognitive science: Shifting the focus. *Philosophical Psychology*, 20(2), 175-195.
- Best, C. T. (1995). A direct realist view on cross-language speech perception. In W. Strange, (Ed.) *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues* (pp. 171-204). Baltimore: York Press.
- Best, C. T., & McRoberts, G. W. (2003). Infant perception of non-native contrasts that adults assimilate in different ways. *Language and Speech*, 46(2-3), 183-216.
- Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of Phonological Constancy: Toddlers' Perception of Native- and Jamaican-Accented Words. *Psychological Science* 20, 539-542.
- Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia*, 51, 2492-2511.
- Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011a). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, 32, 51-67.
- Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011b). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, 33, 433-461.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: M.I.T. Press.
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York, Westport and London: Praeger.
- Cruttenden, A. (2014). *Gimson's Pronunciation of English*. Eight edition. London and New York: Routledge.
- Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28, 369-382.
- Deuter, M., Bradbery, J., & Turnbull, J. (2015). *Oxford Advanced Learner's Dictionary*. 9<sup>th</sup> edition. London: Oxford University Press.
- Dörnyei, Z. (2005). *The Psychology of the Language Learner – Individual differences in Second Language Acquisition*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

- Elsabbagh, M., & Karmiloff-Smith, A. (2006). Modularity of mind and language. In K. Brown (Ed.) *The Encyclopedia of Language and Linguistics* (2nd ed., pp. 218-224). Oxford: Elsevier.
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*, 78-104.
- Fodor, J. A. (1984). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Ganschow, L., Sparks, R., Javorsky, J., Pohlman, J., & Bishop-Marbury, A. (1991). Identifying native language difficulties among foreign language learners in college: A foreign language learning disability? *Journal of Learning Disabilities*, *24*, 530-541
- Gelderen, E. van. (2006). *A History of the English Language*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Genesee, F. (1976). The role of intelligence in second language learning. *Language Learning*, *26*, 267-280.
- Granena, G., & Long, M. H. (2012). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311-343.
- Grønnum, N. (1998). Danish. *Journal of the International Phonetic Association* *28*, 99-105.
- Harrell, F. E. Jr. (2013). With contributions from Charles Dupont and many others. *Hmisc: Harrell Miscellaneous*, R package version 3.10-1.1. [Cited 11.05.2015] <<http://biostat.mc.vanderbilt.edu/Hmisc>>, <<https://github.com/harrelfe/Hmisc>>
- Herslund, M. (2002). *Danish: Languages of the World/ Materials 382*. München: Lincom Europa.
- Horslund, C. S. (2016). I don't know why did they accept that: Grammaticality judgements of negation and questions in L1 Danish and L1 Finnish learners of English. In S. Vikner, H. Jørgensen, & E. van Gelderen (Eds.), *Let's have articles betwixt us – Papers in Historical and Comparative Linguistics in Honour of Johanna L. Wood* (pp. 221-260). Dept. of English, School of Communication and Culture, Aarhus University.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65-70.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346-363.
- Hyltenstam, K., Bylund, E., Abrahamsson, N., & Park, H-S. (2009). Dominant-language replacement: The case of international adoptees. *Bilingualism: Language and Cognition*, *12*(2), 121-140.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434-444.

- Karlsson, F. (1999). *Finnish: An Essential Grammar*. Translated by Andrew Chesterman. London (UK): Routledge.
- King, R. D. (1967). Functional load and sound change. *Language* 43(4), 831-852.
- Knightly, L. H., Jun, S-A., Oh, J. S., & Au, T.K-F. (2003). Production benefits of childhood overhearing. *Journal of Acoustical Society of America*, 114(1), 465-474.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindholm, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606-608.
- Ladefoged, P. & Ferrari Disner, S. (2012). *Vowels and Consonants*. Third edition. Chichester: Wiley-Blackwell.
- Leppänen, S. & Nikula, T. (2007). Diverse uses of English in Finnish society: Discourse-pragmatic insights into media, educational and business contexts. *Multilingua* 26, 333-380.
- Leppänen, S., Pitkänen-Huhta, A., Nikula, T., Kytölä, S., Törmäkangas, T., Nissinen, K., & Kääntä, L. (2011). National survey on the English language in Finland: Uses, meanings and attitudes. *Varieng: Studies in Variation, Contact and Change in English*, 5, 1-385.
- Li, S. (2015). The construct validity of language aptitude. *Studies in Second Language Acquisition* 38, 801-842.
- Lucas, M. A. (1998). Language acquisition and the Conrad phenomenon. *International Review of Applied Linguistics in Language Teaching*, 36(1), 69-81.
- Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, 21, 339-366.
- Marchman, V. A., Martínez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental science*, 7(2), 212-224.
- McCabe, A. (2011). *An Introduction to Linguistics and Language Studies*. Equinox textbooks and surveys in Linguistics. London and Oakville: Equinox.
- Ministry for Children, Education and Gender Equality. 2014. *Bekendtgørelse af lov om folkeskolen*. Nr. 665, of 20/06 2014.
- Ministry for Children, Education and Gender Equality. 2013. *Bekendtgørelse om uddannelsen til studentereksamen*. Nr. 776, of 26/6 2013.
- Morrish, L. (2015). Introduction: what is Language? What is Linguistics? In N. Braber, L. Cummings, & L. Morrish (Eds.) *Exploring Language and Linguistics* (pp. 1-23). Cambridge: Cambridge University Press.
- Morzinski, M. (1994). *Linguistic Influence of Polish on Joseph Conrad's Style*. Conrad: Eastern and Western Perspectives Vol. 3. New York: Columbia University Press.
- Nash, J. C. (2014). On best practice optimization methods in R. *Journal of Statistical Software*, 60(2), 1-14.

- Nation, P. (2012). *Vocabulary Size Test Instructions and Description*. <<http://www.victoria.ac.nz/lals/about/staff/paul-nation>>
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31, 9-13.
- Pinker, S. (1998). Words and rules. *Lingua*, 106(1), 219-242.
- Pinker, S. (2006). Whatever happened to the past tense debate? In E. Bankovic, J. Ito, & J. J. McCathy (Eds.), *Wondering at the Natural Fecundity of Things: Essays in honour of Alan prince* (pp. 221-238). University of California, Linguistics Research Center.
- Preisler, B. (1999). *Danskerne og det engelske sprog*. Frederiksberg: Roskilde Universitetsforlag.
- Pulkkinen, P. (1989). Anglicismerna i finska sproget. *Språk i Norden*, 89-93.
- R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rato, A., Rauber, A., Kluge, D., & Santo, G. (2013). *Designing Audio, Visual, and Audiovisual Perceptual Experiments with TP Software*. Poster presentation at NEW SOUNDS 2013, May 17-19, Concordia University, Montreal, Canada.
- Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology*, 28(3), 406-413.
- Saito, K. (2017). Effects of Sound, Vocabulary, and Grammar Learning Aptitude on Adult Second Language Speech Attainment in Foreign Language Classrooms. *Language Learning* 67(3), 665-693.
- Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Basingstoke: Palgrave Macmillan.
- Scovel, T. (1978). *The recognition of foreign accents in English and its implication for psycholinguistic theories of language acquisition*. Proceedings of the 5<sup>th</sup> International Association of Applied Linguistics (pp. 389-401). Montreal: Laval University Press.
- Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., & Wang, X. (1999). Consonant recordings for speech testing. *Journal of Acoustical Society of America*, 106(6), 71-74.
- Sharwood Smith, M. (1994). *Second Language Learning: Theoretical Foundations*. London: Longman.
- Snow, C. E., & Hoefnagel-Höhle, M. (1979). Individual differences in second language ability: A factor-analytic study. *Language and speech*, 22(2), 151-162.
- Sparks, R., Patton, J., Ganschow, L., & Humbach, N. (2009). Long-term relationships among early first language skills, second language aptitude, second language affect, and later second language proficiency. *Applied psycholinguistics* 30, 725-755.

- Sparks, R., Patton, J., Ganschow, L., & Humbach, N. (2011). Subcomponents of Second-Language Aptitude and Second-Language Proficiency. *The Modern Language Journal* 95(2), 253-273.
- Strang, Barbara M. H. (1970). *A History of English*. London and New York: Methuen.
- Suomi, K., Toivanen, J., & Ylitalo, R. (2008). *Finnish sound structure: phonetics, phonology, phonotactics and prosody*. Oulu: Oulu University Press.
- Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. New York: Springer.
- Wiik, K. (1965). *Finnish and English vowels: A Comparison with Special Reference to the Learning Problems Met by Native Speakers of Finnish Learning English*. Turku: Turun Yliopisto.

# Ungrammatical Sentences Have Syntactic Representations too

Johannes Kizach  
Aarhus University

## Abstract

A number of experiments have found structural priming effects for grammatical sentences but not for ungrammatical ones. This has led to the hypothesis that ungrammatical sentences do not have a syntactic representation, because this could explain the absence of a priming effect. In this article ungrammatical Danish sentences with heavy NP shift of the object to the right of the particle are investigated in an acceptability judgment study. A syntactic processing account predicts that the sentences should be easier to parse if the syntactic heads (the verb, the particle, and the head of the object) are as close as possible i.e., when the order is short-before-long. The result reveals that participants find the ungrammatical sentences more acceptable when the object is long. This is exactly what is predicted from a processing perspective and suggests that the ungrammatical strings indeed do have syntactic representations. Consequently, I argue that the hypothesis about structureless ungrammatical sentences should be abandoned.

## 1. Introduction

In this article I will present evidence suggesting that ungrammatical sentences have syntactic representations just like grammatical sentences do. The main finding (see section 2 below) is that the processing preference for sentences with short constituents preceding long constituents (henceforth short-before-long) can also be detected when comparing ungrammatical strings. Since the short-before-long preference is commonly assumed to

be related to the syntactic structure – it minimizes the distance between the syntactic heads and this facilitates processing (Hawkins, 1994, 2004) – the fact that the preference is still observed in ungrammatical sentences suggests that they have syntactic representations too, contrary to the suggestion in Sprouse (2007).

In head-initial languages, such as Danish and English, a strong tendency to place short constituents before long ones has often been observed, and this preference is usually ascribed to a processing advantage of the short-before-long order (Bresnan, Cueni, Nikitina, & Baayen, 2007; De Cuypere & Verbeke, 2013; Hawkins, 1994, 1998, 2004, 2014; Kizach, 2015; Kizach & Vikner, 2016; Seoane, 2009; Wasow, 1997). The syntactic heads of the constituents are simply closer together if the order is short-before-long in a head-initial language, as illustrated here with the particle construction in English (with the relevant syntactic heads in bold type):

- (1) a. Bill threw **[out]** **[the old suitcase]**.  
 b. Bill threw **[the old suitcase]** **[out]**.

In (1)a, the heads of the constituents, *out* and *the*, are adjacent and the shorter phrase (*out*) precedes the longer phrase (*the old suitcase*). In (1) b, on the other hand, the two heads are not adjacent and the longer phrase precedes the shorter one. If we accept the standard assumption that parsing is an incremental process where the structure is projected/built based on the incoming words (cf. Ferreira & Slevc, 2007; Frazier, 1987; Pritchett, 1992; Van Gompel & Pickering, 2007), then the parser can project both constituents after processing only two words in (1)a, but in (1)b four words have to be processed before the structure can be projected. If processing matters for how we order the strings of words, we would expect the short-before-long order in (1)a to be more frequent than the long-before-short order in (1)b. Indeed, a corpus study of the English particle construction demonstrated that 74% of 1,684 examples had the predicted short-before-long order, and the longer the DP was, the stronger the preference became (Lohse, Hawkins, & Wasow, 2004, p. 243).

In Danish there is no choice between orders in the particle construction: only the equivalent of the English (1b), i.e. (2)b, is grammatical (cf. Vikner, 1987):

- (2) a. \*Bent smed [ud] [den gamle kuffert].  
*Bent threw out the old suitcase*  
 ‘Bent threw out the old suitcase.’

- b. Bent smed [den           gamle   kuffert] [ud].  
*Bent threw the           old        suitcase out*  
 ‘Bent threw the old suitcase out.’

The question is whether the quite robust preference for short-before-long extends to ungrammatical sentences such as (2)a. That is, does the short-before-long order still give a processing advantage when we parse ungrammatical strings? If the short-before-long preference can also be observed in the processing of ungrammatical sentences, it would suggest that ungrammatical sentences also have syntactic representations.

Precisely the opposite was suggested by Sprouse (2007) who argued that strings that are not licensed by the grammar do not get a structural representation, which in turn explains the alleged lack of syntactic priming effects for ungrammatical sentences. Henceforth I will call this hypothesis the *No Structure Hypothesis* (abbreviated NSH). After being exposed to a specific syntactic structure, people are relatively faster when reading another sentence with the same structure (Balling & Kizach, 2015; Branigan, 2007; Kizach & Balling, 2013). The syntactic priming effect can also be measured in acceptability judgment experiments where a primed structure is judged more positively as a function of how much exposure it gets (Christensen, Kizach, & Nyvad, 2013; Luka & Barsalou, 2005). In other words – participants tend to rate a structure better and better the more they are exposed to it.

Sprouse (2007) investigated the subject, adjunct, *wh*-, and complex NP island constructions exemplified in (3), which are all considered ungrammatical in English, and found no priming effects for any of them. He argued that the explanation is that the ungrammatical strings are not assigned a syntactic structure and consequently, structural priming is not possible.

- (3) a. \*Who do you think the email from \_\_\_ is on the computer?  
 (subject island)  
 b. \*Who did you leave the party because Mary kissed \_\_\_?  
 (adjunct island)  
 c. \*Who do you wonder whether Susan met \_\_\_?  
 (*wh*-island)  
 d. \*Who did you hear the rumor that David likes \_\_\_?  
 (complex NP island)



However, Snyder (2000, p. 796) tested some of the same structures and reported priming effects for *wh*-islands and complex NP islands. These results have been partially replicated, but the reliability of these results have been debated (Crawford, 2012; Sprouse, 2009).

Christensen et al. (2013) found priming effects for grammatical strings in Danish, but not for ungrammatical strings, which supports Sprouse's (2007) NSH. However, Ivanova et al. (2012) examined sentences such as (4), where an intransitive verb is used as a ditransitive verb, and found priming effects despite the fact that the sentences were ungrammatical.

- (4) \*The waitress exists the book to the monk.

The NSH suggests that if the sentence is ungrammatical, the parser does not assign a structure to it. If this is indeed the case, then we would predict that the preference for short-before-long disappears in ungrammatical strings – there simply is no structure to project and consequently no word order can speed up the structure building process.

To test this prediction, I investigated the contrast between particles followed by pronominal DPs, one word nominal DPs, and DPs modified by a relative clause in Danish, as in (5) below.

- (5) a. \*Anita smed [væk] [den].  
*Anita threw away it*  
 'Anita threw away it.'
- b. \*Anita smed [væk] [banan-en].  
*Anita threw away banana.the*  
 'Anita threw away the banana.'
- c. \*Anita smed [væk] [den store kasse bananer  
*Anita threw away the big box bananas*  
*der stod i garag-en].*  
*which stood in garage.the*  
 'Anita threw away the big box of bananas which was standing  
 in the garage.'

All the examples in (5) are ungrammatical in Danish, so none of them should get a structural analysis according to NSH, and this means that the general preference for short-before-long word order should not affect the acceptability judgments of these sentences.

Heavy NP shift (Ross, 1967) is possible in other constructions in Danish (see examples in Drengsted-Nielsen, 2014, p. 166), and the strings in (5) have a word order that would in principle be derivable if the object was shifted to the right across the particle. However, it is ungrammatical to move the object in a particle construction to the right in Danish. But we already know from studies of English that heavy NP shift is more acceptable when the shifted object is longer than the constituent it moves across, and the acceptability increases as the length difference increases (Arnold, Wasow, Losongco, & Ginstrom, 2000; Hawkins, 1994; Wasow, 2002; Wasow & Arnold, 2003).

If the sentences in (5) have syntactic representations (even though they are ungrammatical), a processing theory such as Hawkins' (2004) would predict that the length/weight of the object DPs influence processing. In (5)a the object DP is pronominal (*den*) and contains just one maximal projection (a DP) – counting the number of XPs is a common way of quantifying the length/weight of constituents (Hawkins, 1994; Kizach, 2010, pp. 53-55; Szmrecsanyi, 2004; Wasow, 1997). In (5)b the object DP contains two XPs (a DP and an NP), and in (5)c the DP object contains more than five XPs. Hawkins (2004) predicts that the longer the DP is, the easier it becomes for the parser, and the higher the acceptability ratings should be. Notice that it is the relative weight that is important here: The benefit of displacing the long DP in (5)c is simply higher than it is in (5)b due to the greater relative weight difference. The grammatical DP-particle order, as in (2)b above, results in a long-before-short order which is difficult to process (and the longer the DP, the worse it gets) – the ungrammatical heavy NP shift, as in (5), will reduce the processing difficulty, but the price is ungrammaticality. The point is that this trade-off might be detectable in the processing of the sentences in (5), in which case we would expect an acceptability hierarchy such that (5)c is better than (5)b, which is better than (5)a – (5)c > (5)b > (5)a – precisely because of the processing benefit of short-before-long.

Note that Hawkins' (2004) theory is only used here to test Sprouse's (2007) hypothesis – if the NSH is right and ungrammatical sentences have no syntactic representations, Hawkins' (2004) theory would not predict anything either (the facilitating effect of having syntactic heads adjacent is only relevant for strings with a syntactic representation, not for e.g. shopping lists).

If any differences between the conditions in (5) are found, it would potentially be problematic for the NSH, but we know that the absolute length of a constituent affects acceptability negatively. Christiansen & MacDonald (2009) varied the length of DP constituents and compared sentences as those in (6).

- (6) a. The boss from the office says that the posters across the hall tell lies.  
 b. The boss says that the posters in the office across the hall tell lies.  
 c. The posters on the desk in the office across the hall tell lies.

Note that the underlined DP constituents are modified by one, two and three PPs respectively, and that (6)a and b contain embedded clauses. Christiansen & MacDonald (2009, pp. 141-142) report that the acceptability of the sentences in (6) is correlated with the length of the DPs. This means that (6)a is judged to be better than (6b) which is better than (6)c – the result suggests that even increasing the length of a DP with a single PP can decrease the overall acceptability.

So if the results show that there are differences between the sentences in (5) it may just be this absolute length effect and the NSH could still be right. The hierarchy predicted by Hawkins (2004) is in the opposite direction: The longer the DP, the *higher* the acceptability should be. If the results show this pattern it would lend further support to the idea that the NSH should be abandoned.

The evidence for the NSH is based on null-results – Sprouse (2007, p. 127) found no priming effects for various island-violations, and Christiansen et al. (2013, p. 58) found no priming effects for ungrammatical sentences. In the experiment presented below the NSH would again predict a null-result (or a slight preference for shorter sentences as mentioned above), but by introducing Hawkins' (2004) theory we have an alternative prediction that is the opposite of NSH's prediction.

In summary, NSH predicts no difference (or a preference for short sentences) in acceptability between the sentences in (5), but Hawkins' (2004) theory predicts the following acceptability hierarchy: – (5)c > (5) b > (5)a.

## 2. The experiment – the particle construction

For this experiment I chose the acceptability judgment task to test the predictions instead of a task that would give me a reaction time measure (RT) such as self-paced reading or eye-tracking. The reason was that the

prediction of Hawkins' (2004) theory is that the shortest sentences should be the hardest to process, and we know that RT increases with sentence length. This means that an RT measure might hide the increased processing load (the shorter sentences increase RT, but on the other hand they are of course faster to read than the longer ones, so the effect might be neutralized and undetectable). Previous research has shown that processing difficulty affects acceptability ratings, so even completely grammatical sentences, such as e.g. *wh*-questions as in (7) get a lower mean acceptability rating than similar sentences without *wh*-movement (8) (Christensen et al., 2013; Fanselow & Frisch, 2006).

- (7) Hvad ved hun godt at man kan leje dér?  
*what knows she well that one can rent there*  
'What does she know that one can rent there?'
- (8) Hun ved godt at man kan leje noget dér.  
*she knows well that one can rent something there*  
'She knows that one can rent something there.'

The acceptability judgment task was thus ideal for my purposes since I could measure the processing difference and avoid the confounding effect of total length.

## **2.1 Participants, materials and methods**

12 sets of sentences as in (5) were created and divided into three lists ensuring that each participant saw an equal number of items from each condition but never the same item in more than one condition. In addition to the experimental items each list contained 15 fillers which ranged from completely acceptable (9) to completely unacceptable (10) sentences. Google Forms on Google Drive was used to create the lists and collect the data.

- (9) Sonja talte i telefon med en veninde.  
*Sonya spoke in phone with a friend*  
'Sonya talked on the phone with a friend.'
- (10) \*Omend ham så gik det jo alligevel.  
*Although him so went it nevertheless anyway*  
'Even though him it went ok nevertheless anyway.'

Links to the lists were made available on-line on the Facebook site *Psycholab* (a forum for students at Aarhus University interested in syntax) and seventy people participated (18 males). The mean age was 24.3 with a range from 20 to 61.

An instruction was shown at the beginning of each list. The English translation of the instruction is: “*Judge the sentences on a scale from 1 (completely unacceptable) to 7 (completely acceptable). Try to follow your immediate intuition, and do not be affected by what you have been taught in school – there are no right or wrong answers here.*”

## 2.2 Results

As predicted, the results showed a (5)c > (5)b > (5)a acceptability hierarchy, as summarized in the table below:

Type of object	Example	Mean rating
Pronoun	(5)a	1.6
Nominal DP	(5)b	2.2
DP with a relative clause	(5)c	2.7

Table 1: Mean ratings across participants on a scale from 1 (completely unacceptable) to 7 (completely acceptable)

To see whether the mean ratings were statistically significant from each other, the data was analyzed with a linear mixed-effects model following the recommendations and practices common in the field (Gibson, Piantadosi, & Fedorenko, 2011; Sprouse, 2008). The software R and the R-package *lmerTest* were used to perform the analysis (Kuznetsova, Brockhoff, & Christensen, 2015; R Development Core Team, 2015).

The dependent variable was the acceptability score and the independent variable was condition – a factor with three levels as illustrated in (5) above (pronominal DP, nominal DP, and nominal DP modified by a relative clause). The so-called maximal model was fitted to the data (Barr, Levy, Scheepers, & Tily, 2013), and comparisons with the zero-correlation-parameter model did not justify a simpler model (Bates, Kliegl, Vasishth, & Baayen, 2015), and consequently the maximal model is reported. The reference level for the condition factor was set as the nominal DP because the question was whether the pronominal DP and the nominal DP modified by a relative clause were different from this reference level.

The results (see Table 2) showed that acceptability was significantly higher as a function of the length of the DP ( $p < 0.05$ ). In other words, the condition with pronominal DPs was judged to be less acceptable than the one with nominal DPs which was less acceptable than the one with DPs modified by a relative clause.

	Estimate	Std. Error	t-value	p-value
DP with a relative clause	0.451	0.197	2.288	0.045
Pronominal DP	-0.602	0.177	-3.396	0.004

Table 2: Results of the linear mixed-effect model – both rows show the comparison to the nominal DP condition

The analysis showed that the acceptability of the sentence types illustrated in (5) exactly followed the hierarchy predicted by Hawkins' (2004) model: (5c) > (5b) > (5a). The longer the DP, the higher the acceptability rating.

### 2.3 Discussion

The NSH is based on the absence of priming effects for ungrammatical strings in acceptability judgment experiments (Christensen et al., 2013; Sprouse, 2007), but as mentioned in the introduction, others have reported priming effects for ungrammatical sentences in English (Crawford, 2012; Ivanova et al., 2012; Snyder, 2000).

The prediction based on Hawkins' (2004) processing theory was fully borne out: the ungrammatical heavy NP shift resulting in the word orders we see in (5) is comparatively more acceptable with a longer DP. I interpret this as evidence for syntactic structure even in ungrammatical strings, since the prediction is based on the facilitating effect of having the syntactic heads close together.

Taken together the previous research and the experiment presented in this article seem to refute the NSH in its present form. One could, however, change the NSH to a universal version which would predict that there will be no priming effects for a structure only if it is disallowed by any possible grammar. In other words, only if the structure somehow violates universal principles will it fail to induce priming effects. In the following, I will briefly discuss this idea.

Three of the four island constraints investigated in Sprouse (2007) do not hold in Danish where there are grammatical examples with adjunct, *wh*-, and complex NP islands violations (Nyvad, Christensen, & Vikner,

2017, pp. 453-461). In Norwegian too there are grammatical examples with complex NP island violations (Áfarli & Eide, 2003, p. 268). Finally, Phillips (2006, p. 796) report that extraction from a subject island is acceptable in parasitic gap constructions in English as exemplified in (11):

(11) What did the attempt to repair \_\_\_ ultimately damage \_\_\_?

The ungrammatical Danish example that fails to induce priming effects reported in Christensen et al. (2013, p. 55) is shown in (12).

(12) \*Ved hun godt hvor hvad man kan leje?  
*knows she well where what one can rent*  
 ‘Does she know where what you can rent?’

In (13) a very similar but fully grammatical Czech construction is shown (Veselovská, 1993, p. 31; her (1c)):

(13) Zajímá mě kdo co přinese.  
*wonder me who what brings*  
 ‘I wonder who will bring what.’

Furthermore, the even more parallel (14) is perfectly grammatical, according to my two Czech informants.

(14) Zajímá mě kdy co Petr přinese.  
*wonder me when what Peter brings*  
 ‘I wonder when Peter will bring what.’

It seems that most of the structures examined in Sprouse (2007) and the ungrammatical one examined in Christensen et al. (2013) are all ungrammatical only because the English and Danish grammars happen to rule them out, not because they are in violation of what is possible in language as such. The only possible candidate among them for a universally ungrammatical structure is the subject island, but even extraction from this island type is possible in the right circumstances (namely in parasitic gap constructions as shown in Phillips, 2006). In summary, the examples investigated in Sprouse (2007) and Christensen et al. (2013) do not allow us to conclude anything about the universal version of the NSH. This means that it might still be true that sentences that somehow violate universal

principles may lack a structural representation, and as a result structural priming might not be possible with these structures. It is, however, not completely clear what structures this would concern. Given the flexibility of X-bar syntax it is difficult to imagine a sentence with a word order that is somehow against universal grammar. The reviewer pointed to a study by Musso et al. (2003) where participants attempted to learn artificial grammatical rules that were either natural, i.e. in correspondence with universal grammar (e.g. forming passive using a suffix on the verb), or unnatural (e.g. marking past tense with a suffix on the second last word in the sentence). An increased activation in Broca's area over time was observed in the learning sessions using the natural rules, but none was observed for the unnatural ones (Musso et al., 2003, p. 778), and this suggests that the unnatural rules simply cannot be learnt, and then maybe these sentences might not have a structural representation. Note, however, that this finding concerns rule types and not simply word order variation – so it seems as if the universal version of the NSH might possibly be true, but may have very little practical relevance (it may concern a very limited set of sentences).

### **3. Conclusion**

The results reveal two things. First, ungrammatical sentences appear to be subject to the same processing constraints on relative length as grammatical sentences. The ungrammaticality of the examined Danish sentences is due to the fact that heavy NP shift of the object across the particle is not allowed by the Danish grammar. Nevertheless, there is a positive correlation between the acceptability of ungrammatical heavy NP examples and the relative length (weight) of the DP immediately following the particle: the longer the better – precisely as is the case with grammatical examples of heavy NP shift in English (Arnold, Wasow, Losongco, & Ginstrom, 2000; Hawkins, 1994; Wasow, 2002; Wasow & Arnold, 2003). The same pattern is observed for grammatical and ungrammatical sentences, demonstrating the similarity between processing grammatical and ungrammatical strings.

Second, the NSH is not accurate. Previous studies have found priming effects for ungrammatical sentences, and the present results strongly suggest that the processing of ungrammatical sentences is subject to the same constraints as the processing of grammatical ones.

The conclusion is that we should simply abandon the idea that the absence/presence of structural priming effects in acceptability judgment experiments correlates with grammaticality in a straightforward way.



### Acknowledgements

I would like to thank an anonymous reviewer for very helpful comments and Sam Featherston for suggesting that I investigated the acceptability of ungrammatical sentences. I also wish to thank Sten Vikner for discussing the data and results with me, and Michaela Hejná and Kateřina Haušildová Graneberg for their judgments of the Czech examples.

### References

- Áfarli, T. A., & Eide, K. M. (2003). *Norsk generativ syntaks*. Novus-Verlag.
- Arnold, J. E., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering. *Language*, 76, 28-55.
- Balling, L. W., & Kizach, J. (2015). Surprised by locality: An eye-tracking study of Danish double object constructions. *Poster Session at Amlap 2015*, Poster.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv Preprint arXiv:1506.04967*. Retrieved from <https://arxiv.org/abs/1506.04967>
- Branigan, H. (2007). Syntactic priming. *Language and Linguistics Compass*, 1(1-2), 1-16.
- Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Bouma, I. Krämer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69-94). Amsterdam: Royal Netherlands Academy of Science. Retrieved from [http://www.ehu.es/seg/\\_media/morf/5/5/6/sak/predicting\\_the\\_dative\\_alternation.pdf](http://www.ehu.es/seg/_media/morf/5/5/6/sak/predicting_the_dative_alternation.pdf)
- Christensen, K. R., Kizach, J., & Nyvad, A. M. (2013). Escape from the Island: Grammaticality and (Reduced) Acceptability of wh-island Violations in Danish. *Journal of Psycholinguistic Research*, 42(1), 51-70.
- Christiansen, M. H., & MacDonald, M. C. (2009). A usage-based approach to recursion in sentence processing. *Language Learning*, 59(s1), 126-161.
- Crawford, J. (2012). Using syntactic satiation to investigate subject islands. In *Proceedings of the 29th West Coast Conference on Formal Linguistics* (pp. 38-45).
- De Cuypere, L., & Verbeke, S. (2013). Dative alternation in Indian English: A corpus-based analysis. *World Englishes*, 32(2), 169-184.
- Drengsted-Nielsen, C. (2014). *Grammatik på dansk* (2nd edition). Copenhagen: Hans Reitzels Forlag.

- Fanselow, G., & Frisch, S. (2006). Effects of Processing Difficulty on Judgements of Acceptability. In G. Fanselow, C. Fery, & M. Schlesewsky (Eds.), *Gradience in Grammar: Generative Perspectives* (pp. 291-316). Oxford: Oxford University Press.
- Ferreira, V. S., & Slevc, L. R. (2007). Grammatical encoding. In G. M. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 453-470). UK: Oxford University Press.
- Frazier, L. (1987). Sentence Processing: a Tutorial Review. In M. Coltheart (Ed.), *Attention and Performance XII*. Hove and London / Hillsdale: Lawrence Erlbaum Associates.
- Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using Mechanical Turk to Obtain and Analyze English Acceptability Judgments: Linguistic Acceptability on Mechanical Turk. *Language and Linguistics Compass*, 5(8), 509-524. <https://doi.org/10.1111/j.1749-818X.2011.00295.x>
- Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency*. Cambridge, UK: Cambridge University Press.
- Hawkins, J. A. (1998). A processing approach to word order in Danish. *Acta Linguistica Hafniensia*, 30(1), 63-101.
- Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Hawkins, J. A. (2014). *Cross-linguistic Variation and Efficiency*. Oxford: Oxford University Press.
- Ivanova, I., Pickering, M. J., Branigan, H. P., McLean, J. F., & Costa, A. (2012). The comprehension of anomalous sentences: Evidence from structural priming. *Cognition*, 122(2), 193-209.
- Kizach, J. (2010). *The function of word order in Russian, compared to English and Danish* (unpublished Ph.D. thesis). Aarhus University, Arts, Department of Aesthetics and Communication, English Degree Programme.
- Kizach, J. (2015). Animacy and the ordering of postverbal prepositional phrases in Danish. *Acta Linguistica Hafniensia*, 1-21.
- Kizach, J., & Balling, L. W. (2013). Givenness, complexity, and the Danish dative alternation. *Memory & Cognition*, 41(8), 1159-1171. <https://doi.org/10.3758/s13421-013-0336-3>
- Kizach, J., & Vikner, S. (2016). Head adjacency and the Danish dative alternation. *Studia Linguistica*. <https://doi.org/10.1111/stul.12047>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package "lmerTest" (Version 2.0-29). Retrieved from <http://CRAN.R-project.org/package=lmerTest>
- Lohse, B., Hawkins, J. A., & Wasow, T. (2004). Domain minimization in English verb-particle constructions. *Language*, 80(2), 238-261.
- Luka, B. J., & Barsalou, L. W. (2005). Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language*, 52(3), 436-459. <https://doi.org/10.1016/j.jml.2005.01.013>

- Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., & Weiller, C. (2003). Broca's area and the language instinct. *Nature Neuroscience*, 6(7), 774-781.
- Nyvad, A. M., Christensen, K. R., & Vikner, S. (2017). CP-recursion in Danish: A cP/CP-analysis. *The Linguistic Review*, 34(3), 449-477.
- Phillips, C. (2006). The real-time status of island phenomena. *Language*, 82(4), 795-823.
- Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. University of Chicago Press.
- R Development Core Team. (2015). R: A language and environment for statistical computing (Version R version 3.2.2). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Ross, J. R. (1967). *Constraints on variables in syntax*. MIT.
- Seoane, E. (2009). Syntactic complexity, discourse status and animacy as determinants of grammatical variation in Modern English. *English Language and Linguistics*, 13(3), 365. <https://doi.org/10.1017/S1360674309990153>
- Snyder, W. (2000). An experimental investigation of syntactic satiation effects. *Linguistic Inquiry*, 31(3), 575-582.
- Sprouse, J. (2007). Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics*, 1, 123-134.
- Sprouse, J. (2008). The differential sensitivity of acceptability judgments to processing effects. *Linguistic Inquiry*, 39(4), 686-694.
- Sprouse, J. (2009). Revisiting satiation: Evidence for an equalization response strategy. *Linguistic Inquiry*, 40(2), 329-341.
- Szmrecsanyi, B. (2004). On operationalizing syntactic complexity. *Jadt-04*, 2, 1032-1039.
- Van Gompel, R. P., & Pickering, M. J. (2007). Syntactic parsing. In G. M. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 289-307). Oxford: Oxford University Press.
- Veselovská, L. (1993). *WH Movement and Multiple Questions in Czech* (PhD Thesis). Durham University.
- Vikner, S. (1987). Case assignment differences between Danish and Swedish. In *Proceedings of the Seventh Biennial Conference of Teachers of Scandinavian Studies in Great Britain and Northern Ireland* (pp. 262-281).
- Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change*, 9(1), 81-105.
- Wasow, T. (2002). *Postverbal Behavior*. Stanford: CSLI Publications.
- Wasow, T., & Arnold, J. (2003). Post-verbal constituent ordering in English. *Topics in English Linguistics*, 43, 119-154.

## Why German is not an SVO-language but an SOV-language with V2

Sten Vikner  
Aarhus University

### Abstract

This paper<sup>1</sup> will take as its starting point the widely assumed distinction between SVO-languages and SOV-languages, with a particular focus on German as compared to English and to Danish. It will be argued that German (and Dutch, Frisian and Old English) is an SOV-language whereas Danish and English (and Icelandic) are SVO-languages, even though several orders may be found inside each of these languages. It will also be shown where the verb second (V2) property fits in, which is common to German and Danish (and Old English), but only found in (present-day) English to a much smaller extent.

The differences between this analysis and two other analyses will also be discussed, namely the analysis in Greenberg (1963) and Bohn (1983) that both German and English are SVO-languages, and the analysis in Bohn (2003) that German is SVO in main clauses but SOV in subordinate clauses.

### 1. Is German SVO or SOV?

I will take my starting point in Greenberg's (1963, p. 109) discussion of "basic word order", by which he means the "dominant" order of the

<sup>1</sup> Many thanks to Ken Ramshøj Christensen, Henrik Jørgensen, Anne Mette Nyvad, Ramona Römisch-Vikner, Carl Vikner, and Johanna Wood. A special thank you to Ocke-Schwen Bohn for always being ready to discuss and dispel linguistic misunderstandings, myths and prejudices.

**subject**, the **verb** and the **object**. Establishing the basic word order of a particular language is not as easy as it may sound. Danish, e.g., allows at least four different orders:<sup>2</sup>

- S     V     O
- (1) a. Hvis Ocke bruger det her program, ...  
*If Ocke uses this here programme*, ...  
 = ‘If Ocke uses this programme, ...’
- O             V     S
- b. Det her program bruger Ocke.  
*This here programme uses Ocke*.  
 = ‘This programme Ocke uses.’
- V     S     O
- c. Bruger Ocke det her program?  
*Uses Ocke this here programme?*  
 = ‘Does Ocke use this programme?’
- O             S     V
- d. Jeg ved ikke hvad for et program Ocke bruger.  
*I know not what for a programme Ocke uses*.  
 = ‘I don’t know which programme Ocke uses.’

Now the question is which of these four should be chosen as the basic order of Danish. Here I agree with Greenberg’s (1963, p. 109) suggestion that the basic order of Danish is **S**ubject-**V**erb-**O**bject, as in (1a). However, although I agree with Greenberg on what the basic order is, I do not agree with him as to why this should be so.

Greenberg (1963, p. 109) puts **all** the Germanic languages into the same group, i.e. **SVO**, and similarly Bohn (1983, p. 75) analyses both English and German as SVO-languages.<sup>3</sup>

<sup>2</sup> All examples in this paper have been constructed and checked with native speakers, with two obvious exceptions: Examples (3d) and (11d), which were constructed by Johanna Wood.

<sup>3</sup> I should hasten to add that Bohn (1983, p. 75) explicitly says that he is only concerned with main clauses with a finite main verb. This limitation will be discussed in more detail in section 4 below.

I find it more promising to classify only Scandinavian and English as SVO, (2), and to take the basic order of German, Dutch and Frisian (and by extension also Old English) to have the object before the verb, i.e. to classify these three languages as Subject-Object-Verb, SOV, (3):

(2) SVO			verb	object
a.	Danish	Jeg har	læst	bogen.
b.	Icelandic	Ég hef	lesið	bókina.
c.	English	I have	read	the book.

(3) SOV			object	verb
a.	Dutch	Ik heb	het boek	gelezen.
b.	Frisian	Ik ha	it boekje	lêzen.
c.	German	Ich habe	das Buch	gelesen.
d.	Old English	Ic habe	þa boc	gereded.

*I have the book read*

(The analysis of Dutch, Frisian, German and Old English as SOV-languages goes back to Bach, 1962, Bierwisch, 1963, and Koster, 1975).

Why does Greenberg (1963, p. 109) categorise German as SVO and why does Bohn (1983, p. 75) say that German has SVO order? Neither of them go into any great detail, but they both talk about the “dominant word order” (Greenberg, 1963, p. 76, 109; Bohn, 1983, p. 75).

Whaley (1997, p. 106), a textbook in descriptive comparative linguistics, is more explicit about why he also takes SVO to be the “basic constituent order” of German. He takes an order to be the basic constituent order if it tends to be “strongly felt to be the basic order by native speakers”, if it tends to be “the most frequent order”, “the least marked order”, or the “pragmatically most neutral order”. The reference is thus to tendency rather than to theory.

The classification of German as SOV that I want to advocate here has a theoretical basis: If one order is declared to be the basic order, then all other possible orders have to be accounted for relative to the basic order. The objective is thus to find the order from which all the actually occurring orders can be derived in the least complex way, i.e. necessitating the minimal number of additional rules and exceptions.

Consider therefore first the complications involved in deriving the various orders if we follow Greenberg (1963, p. 109), Bohn (1983, p. 75), and Whaley (1997, p. 103) in taking the basic order of German to be SVO:

## Taking the basic order to be SVO

(4) Main clauses (subject-initial)

a.	Ocke <u>h</u> ält gerade einen Vortrag. <i>Ocke gives just now a talk</i>	no movement required
b.	Ocke hat ___ gestern einen Vortrag <u>g</u> ehalten. <i>Ocke has yesterday a talk given</i>	past participle moved to the right
c.	Ocke wird ___ morgen einen Vortrag <u>h</u> alten. <i>Ocke will tomorrow a talk give</i>	infinitive moved to the right
d.	Ocke wird ___ ___ morgen einen Vortrag <u>g</u> ehalten <u>h</u> aben. <i>Ocke will tomorrow a talk given have</i>	past participle + infinitive moved to the right
e.	Ocke fängt ___ gerade einen Vortrag <u>a</u> n. <i>Ocke begins just now a talk PRT</i>	separable prefix moved to the right
f.	Ocke wird ___ ___ morgen einen Vortrag <u>a</u> nfangen. <i>Ocke will tomorrow a talk PRT-begin</i>	separable prefix + infinitive moved to the right

(5) Embedded clauses

a.	... weil Ocke ___ gerade einen Vortrag <u>h</u> ält. <i>... because Ocke just now a talk gives</i>	finite verb moved to the right
b.	... weil Ocke ___ ___ gestern einen Vortrag <u>g</u> ehalten <u>h</u> at. <i>... because Ocke yesterday a talk given has</i>	past participle + finite verb moved to the right
c.	... weil Ocke ___ ___ morgen einen Vortrag <u>h</u> alten <u>w</u> ird. <i>... because Ocke tomorrow a talk give will</i>	infinitive + finite verb moved to the right
d.	... weil Ocke ___ ___ ___ morgen einen Vortrag <u>g</u> ehalten <u>h</u> aben <u>w</u> ird. <i>... because Ocke tomorrow a talk given have will</i>	past participle + infinitive + finite verb moved to the right
e.	... weil Ocke ___ ___ gerade einen Vortrag <u>a</u> nfängt. <i>... because Ocke just now a talk PRT-begins</i>	separable prefix + finite verb moved to the right
f.	... weil Ocke ___ ___ ___ morgen einen Vortrag <u>a</u> nfangen <u>w</u> ird. <i>... because Ocke tomorrow a talk PRT-begin will</i>	separable prefix + infinitive + finite verb moved to the right

(6) Non-subject-initial main clauses

a.	<u>G</u> estern <u>h</u> at Ocke ___ ___ einen Vortrag <u>g</u> ehalten. <i>Yesterday has Ocke a talk given</i>	adverbial moved to the left + finite verb moved to the left + past participle moved to the right
b.	<u>E</u> inen Vortrag <u>h</u> at Ocke ___ ___ gestern ___ <u>g</u> ehalten. <i>A talk has Ocke yesterday given</i>	object moved to the left + finite verb moved to the left + past participle moved to the right

To get from a basic SVO order to the various word orders actually found in German, a considerable number of different movements would have to be assumed. (4) shows that not only do all non-finite verbal forms (and separable prefixes = separable verb particles) have to be moved to the right in main clauses (as stated explicitly in e.g. Lass 1987, p. 328), but it also has to be assured that all of these non-finite verbal forms (and separable prefixes) occur in the mirror image order of the one they would have had if they had not moved (as seen from their order in e.g. Danish or English: Danish *Lyset må være<sub>1</sub> gået<sub>2</sub> ud<sub>3</sub>* = English *The light must have<sub>1</sub> gone<sub>2</sub> out<sub>3</sub>*, = German *Das Licht muss aus<sub>3</sub>gegangen<sub>2</sub> sein<sub>1</sub>*).<sup>4</sup> The exact same is true of embedded clauses, except that here also the finite verb would have to be moved to the right, as seen in (5). Finally, notice also that even though the basic order has the verb before the object, it is nevertheless also necessary to assume a movement that moves a finite verb to the **left** to account for (6) in addition to a movement that moves a finite verb to the **right** to account for (5).

Consider now how much less complicated the derivation is if the basic order of German is taken to be SOV (adapted from Wöllstein-Leisten et al., 1997, pp. 28-32, see also Vikner, 2001, pp. 87-124; 2005, 2007):

**Taking the basic order to be SOV**

(7) Subject-initial main clauses - same data as (4)

a.	Ocke <u>hält</u> gerade einen Vortrag ____. ↑ <i>Ocke gives just now a talk</i>	finite verb moved to the left
b.	Ocke <u>hat</u> gestern einen Vortrag gehalten ____. ↑ <i>Ocke has yesterday a talk given</i>	finite verb moved to the left
c.	Ocke <u>wird</u> morgen einen Vortrag halten ____. ↑ <i>Ocke will tomorrow just now a talk give</i>	finite verb moved to the left
d.	Ocke <u>wird</u> morgen einen Vortrag gehalten haben ____. ↑ <i>Ocke will tomorrow a talk given have</i>	finite verb moved to the left
e.	Ocke <u>fängt</u> gerade einen Vortrag an ____. ↑ <i>Ocke begins just now a talk PRT</i>	finite verb moved to the left
f.	Ocke <u>wird</u> morgen einen Vortrag anfangen ____. ↑ <i>Ocke will tomorrow a talk PRT-begin</i>	finite verb moved to the left

<sup>4</sup> Given that a number of different verbal forms plus a separable prefix may move, under the SVO-analysis, an answer also has to be found why they all have to move in (4), with one notable exception: the finite verb. If it is possible to move only *gehalten* in (4b), why is it not possible to move only *gehalten* in (4d)? Similarly, if it is possible to move only *an* in (4e), why is it not possible to move only *an* in (4f)?



## (8) Embedded clauses - same data as (5)

a.	... weil Ocke gerade einen Vortrag <u>hält</u> . ... because Ocke just now a talk gives	no movement required
b.	... weil Ocke gestern einen Vortrag gehalten <u>hat</u> . ... because Ocke yesterday a talk given has	no movement required
c.	... weil Ocke morgen einen Vortrag halten <u>wird</u> . ... because Ocke tomorrow a talk give will	no movement required
d.	... weil Ocke morgen einen Vortrag gehalten haben <u>wird</u> . ... because Ocke tomorrow a talk given have will	no movement required
e.	... weil Ocke gerade einen Vortrag <u>anfängt</u> . ... because Ocke just now a talk PRT-begins	no movement required
f.	... weil Ocke morgen einen Vortrag anfangen <u>wird</u> . ... because Ocke tomorrow a talk PRT-begin will	no movement required

## (9) Non-subject-initial main clauses - same data as (6)

a.	<u>Gestern</u> <u>hat</u> Ocke ___ einen Vortrag gehalten ___. ↑     ↑     ↓     ↓ Yesterday has Ocke a talk given	adverbial moved to the left + finite verb moved to the left
b.	<u>Einen Vortrag</u> <u>hat</u> Ocke gestern ___ gehalten ___. ↑     ↑     ↓     ↓ A talk has Ocke yesterday given	object moved to the left + finite verb moved to the left

To get from a basic SOV order to the various word orders actually found in German, a relatively small number of different movements will have to be assumed. Notice e.g. that a finite verb is only ever moved to the left, (7) and (9), never to the right and notice also that no other verbal forms or separable prefixes need to move to account for the data in (7) and (9). Sound theoretical reasoning thus clearly supports the assumption that German (and Dutch, Frisian and Old English) are SOV-languages, not SVO.

## 2. Aux-VP vs. VP-aux the across Germanic languages

The advantage of making a distinction between Scandinavian and English as SVO and Dutch, Frisian, German and Old English as SOV is that it allows a number of further empirical generalisations to be made. One such empirical generalisation is that Germanic SVO-languages always put the finite auxiliary verb, e.g. *have* in (10)/(11), to the left of the verb phrase (VP)<sup>5</sup> in embedded clauses, (10), whereas Germanic SOV-languages most

<sup>5</sup> The assumption behind VP is that just like a preposition together with its complement, e.g. *mit seinem Betreuer* / *with his supervisor*, forms a preposition phrase (PP), a verb together with its complement, e.g. *dieses Buch lesen* / *read this book*, forms a verb phrase (VP). This is supported by the observation that VPs can occur in different positions in the clause:

- (i) Ich hätte [dieses Buch gelesen], wenn ich die Zeit gehabt hätte.  
I would-have this book read, if I the time had would-have
- (ii) [Dieses Buch gelesen] hätte ich, wenn ich die Zeit gehabt hätte.  
This book read would-have I, if I the time had would-have

often (but not exclusively) put the finite auxiliary verb to the right of the VP in embedded clauses, (11):

(10)	SVO			aux	VP
a.	Danish	... fordi	jeg	har	læst bogen.
b.	Icelandic	... af því að	ég	hef	lesið bókina.
c.	English	... because	I	have	read the book.

(11)	SOV			VP	aux
a.	Dutch	... omdat	ik	het boek gelezen	heb.
b.	Frisian	... om't	ik	it boekje lêzen	ha.
c.	German	... weil	ich	das Buch gelesen	habe.
d.	Old English	... forðan	ic	þa boc gereded	habe.

*... because I the book read have*

This empirical generalisation can be formulated as follows:

- (12) **SVO** languages only have **aux-VP**,  
 whereas only **SOV** languages may have **VP-aux**.

From this we can e.g. derive the prediction that if a Germanic language has VO order as in English (i.e. the main verb *read* before the object *the book*), it will **not** have **VP-aux** order (i.e. *read the book* before *have*). In other words, we predict that no Germanic language can have the order ... *because I read the book have*.

### 3. Verb second (V2)

A potential problem with this difference in basic word order between German and Danish is that it might now seem as if these two languages are much more different than they “really” are. However, even though this analysis says that they have different basic word orders (German is SOV, Danish SVO), they still have other central properties in common, e.g. verb second (V2): As shown for Danish in (13) and for German in (14) (see also (9) above), the finite verb in main clauses in both languages moves into the second position and some other constituent, e.g. an adverbial, the object

---

Notice furthermore that if the existence of VPs as constituents is assumed, this is not compatible with the existence as a constituent of a “verb group” consisting of only verbs (i.e. *hätte* and *gelesen* in (i) and (ii)), as assumed by Bohn (1983, p. 80). This point is discussed in more detail in Vikner (2016), in particular the abundant evidence for VP as a constituent (including (i) and (ii)) and the absence of evidence for the verb group as a constituent.

or the subject<sup>6</sup> moves into the first position. In generative linguistics, the first position is called CP-spec and the second position C°, cf. e.g. Vikner (1995, pp. 41-46) or Vikner & Jørgensen (2017, p. 163).

(13)	①	②		
a.	<u>I morgen</u> <sub>1</sub> <i>Tomorrow</i>	<u>vil</u> <sub>2</sub> <i>will</i>	Ocke ____ <sub>2</sub> holde et foredrag ____ <sub>1</sub> .	adverbial to ① + finite verb to ②
b.	<u>Et foredrag</u> <sub>1</sub> <i>A talk</i>	<u>vil</u> <sub>2</sub> <i>will</i>	Ocke ____ <sub>2</sub> holde ____ <sub>1</sub> i morgen.	object to ① + finite verb to ②
c.	<u>Ocke</u> <sub>1</sub> <i>Ocke</i>	<u>vil</u> <sub>2</sub> <i>will</i>	____ <sub>1</sub> ____ <sub>2</sub> holde et foredrag i morgen.	subject to ① + finite verb to ②

(14)	①	②		
a.	<u>Morgen</u> <sub>1</sub> <i>Tomorrow</i>	<u>wird</u> <sub>2</sub> <i>will</i>	Ocke ____ <sub>1</sub> einen Vortrag halten ____ <sub>2</sub> .	adverbial to ① + finite verb to ②
b.	<u>Einen Vortrag</u> <sub>1</sub> <i>A talk</i>	<u>wird</u> <sub>2</sub> <i>will</i>	Ocke morgen ____ <sub>1</sub> halten ____ <sub>2</sub> .	object to ① + finite verb to ②
c.	<u>Ocke</u> <sub>1</sub> <i>Ocke</i>	<u>wird</u> <sub>2</sub> <i>will</i>	____ <sub>1</sub> morgen einen Vortrag halten ____ <sub>2</sub> .	subject to ① + finite verb to ②

I would therefore like to suggest the typological classification that both Danish and German are V2, even though the basic word order in Danish is SVO and the one in German SOV. On the other hand, Danish and English have in common that they both have SVO as the basic word order, but English is not V2 (as opposed to Danish and to German and Old English), as can be seen from the fact that the English version of e.g. (13a)/(14a) does not have the finite verb in the second position left of the subject, but in the third position right of the subject:

(15) Tomorrow, Ocke will give a talk.

#### 4. Can a language be both SVO and SOV?

When Bohn (1983, p. 75) takes both English and German to be SVO-languages, he also says that he is only concerned with main clauses with a finite main verb. Furthermore, when Bohn (1983, p. 80) says that English and German have in common that “the auxiliary verb occurs before the

<sup>6</sup> In fact, (14c) shows that the analysis given in (7) above was strongly simplified. All main clauses in German, also the subject-initial ones in (7) and (14c), are the result of **two** movements: The finite verb moves to the second position and the subject (or object or adverbial or ...) moves to the first position. The same is true for all the other Germanic V2 languages, including Danish and Old English.

main verb in declarative structures”, he is presumably still only talking about main clauses with only one auxiliary verb (as otherwise (4d)/(7d) and (5a-f)/(8a-f) above would be counterexamples).

Explicitly or implicitly limiting the SVO-analysis of German to main clauses in this way raises the question of whether a language can have more than one basic word order, i.e. whether a language can have one basic word order in some circumstances and another basic word order in others.

Where Bohn (1983) simply says nothing about the basic word order in embedded clauses, this might seem to be different in Bohn (2003), a set of lecture notes from a course on the history of the English language, based to a large extent on Lass (1987). However, when Bohn (2003, p. 15) says that German is “verb-second in main clauses, verb-final in subordinate clauses”, it is not 100% clear that he commits himself to the account that German has one basic word order in main clauses (SVO) and another (SOV) in embedded clauses.

Even so, I find it worthwhile to briefly discuss such a view (i.e. that German has one basic word order in main clauses (SVO) and another (SOV) in embedded clauses), also in order to underline that it is no accident that for a language to have more than one basic word order is neither possible in Greenberg’s (1963, pp. 77, 108-110) analysis, nor in the analysis advocated in the present paper.

Admittedly, one advantage of an analysis that says that the basic word order in German (or Dutch or Frisian or Old English) is SVO in main clauses but SOV in embedded ones is that it would replace (5) with (8) above as the analysis of embedded clauses, which is clearly a simplification. However, such an analysis would retain the very non-uniform account of main clauses in (4) and (6), where not only all non-finite verbal forms (and separable prefixes) are moved to the right, but where it also has to be assured that all of these non-finite verbal forms (and separable prefixes) occur in mirror image order, as compared to their order in English or Danish. If instead, as was suggested in (7), (9) and (14) above, also main clauses were to be seen as SOV, the positions of all verbal forms (and separable prefixes) would follow assuming one single additional movement, that of the finite verb to the second position of the main clause. Put differently, compared to SVO-languages like Danish and English, German differs not only in embedded clauses, but crucially also in main clauses. In other words, a dual basic word order analysis of German (SVO in main clauses, SOV in embedded ones) would be an improvement over the analysis of German as

generally SVO only as far as embedded clauses are concerned, and not at all where main clauses are concerned.

Another argument against assuming a dual basic word order analysis of German is that it violates Occam's razor, as it allows more options than are necessary. To be more concrete, given that we need a way of deriving V2 (see section 3 above) independently of German (and of Dutch, Frisian and Old English), because such a derivation is needed for SVO V2 languages like Danish, (13), we might as well use that same derivation for V2 in German. As this will in turn obviate the need for allowing for a dual SOV/SVO option, so that we only need to allow for the simplex SOV option and the simplex SVO option, the preferable analysis has to be one that does not allow for the dual option to begin with, but only for the two simplex ones. Of course, Occam's razor only prohibits extra options (and extra assumptions) if they are not strictly necessary, and so it remains to be seen whether there are other languages in the world where the dual SOV/SVO option is the only possible analysis. I have merely argued here that this is not the case within the Germanic languages, but as the World Atlas of Language Structures Online (Dryer, 2013) lists a total of 189 "languages lacking a dominant order", only three of which are Germanic (viz. German, Dutch and Frisian), it remains to be seen whether any of the other 186 ones might require the dual SOV/SVO option and not be amenable to alternative and more restrictive derivations.

## **5. Conclusion**

The distinction between SVO-languages and SOV-languages made by Greenberg (1963), Bohn (1983) and many many others was argued to be a very useful distinction, even more so if it is put on a solid theoretical and empirical footing, rather than just being a tendency. More concretely, the SVO-SOV-distinction was used to account for a number of very basic and common differences between English and Danish (and Icelandic) on one hand and German (and Dutch, Frisian and Old English) on the other. The resulting account was argued to be clearly preferable to accounts where all of the Germanic languages are taken to be SVO-languages, such as Greenberg (1963, p. 109) and Bohn (1983, p. 75).

The verb second (V2) property was shown to play a crucial role in this account. It was also shown how other generalisations concerning the Germanic languages could make reference to the SVO-SOV-distinction. Finally, it was argued to be desirable not to allow any languages to be both SOV and SVO.

## References

- Bach, E. (1962). The Order of Elements in a Transformational Grammar of German. *Language*, 38, 263-269.
- Bierwisch, M. (1963). *Grammatik des Deutschen Verbs*, Berlin: Akademie-Verlag.
- Bohn, O.-S. (1983). *The L2 Acquisition of English Sentence Structure: The Early Stages – A Case Study of Four German Children*. PhD-dissertation, University of Kiel, published as *Arbeitspapiere zum Spracherwerb* 32.
- Bohn, O.-S. (2003). Present-day changes in English. Unpublished lecture notes from a course in the History of the English Language, Aarhus University.
- Dryer, M. S. (2013). Order of Subject, Object and Verb. In M. S. Dryer & M. Haspelmath (Eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://wals.info/chapter/81>>
- Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.): *Universals of Language* (pp. 73-113). Cambridge MA: MIT Press.
- Koster, J. (1975). Dutch as an SOV Language. *Linguistic Analysis*, 1, 111-136.
- Lass, R. (1987). *The shape of English: structure and history*. London: Dent.
- Vikner, S. (1995). *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford Studies in Comparative Syntax. New York: Oxford University Press.
- Vikner, S. (2001). *Verb Movement Variation in Germanic and Optimality Theory*, “Habilitationsschrift”, University of Tübingen. <[www.hum.au.dk/engelsk/engsv/papers/viknhabi.pdf](http://www.hum.au.dk/engelsk/engsv/papers/viknhabi.pdf)>.
- Vikner, S. (2005). Immobile Complex Verbs in Germanic. *Journal of Comparative Germanic Linguistics*, 8.1-2, 83-115. <[www.hum.au.dk/engelsk/engsv/papers/vikn05b.pdf](http://www.hum.au.dk/engelsk/engsv/papers/vikn05b.pdf)>
- Vikner, S. (2007). Teoretisk og komparativ syntaks. In H. Jørgensen & P. Widell (Eds.), *Det bedre argument – Festskrift til Ole Togeby, 7. marts 2007* (pp. 469-480). Aarhus: Wessel & Huitfeldt. <[www.hum.au.dk/engelsk/engsv/papers/vikn07a.pdf](http://www.hum.au.dk/engelsk/engsv/papers/vikn07a.pdf)>
- Vikner, S. (2016). English VPs and why they contain more than just verbs. In S. Vikner, H. Jørgensen & E. van Gelderen (Eds.), *Let us have articles betwixt us – Papers in Historical and Comparative Linguistics in Honour of Johanna L. Wood* (pp. 439-464). Aarhus: Dept. of English, Aarhus University. <[www.hum.au.dk/engelsk/engsv/papers/vikn16b.pdf](http://www.hum.au.dk/engelsk/engsv/papers/vikn16b.pdf)>
- Vikner, S. & Jørgensen, H. (2017). En Formel vs. En Funktionel Tilgang Til Dansk Sætningsstruktur. *Nydanske Sprogstudier – NyS* (pp. 52-53 & 135-68). <<https://doi.org/10.7146/nys.v1i52-53.24954>>
- Whaley, L. (1997). *Introduction to typology: The unity and diversity of language*. Thousand Oaks, CA: Sage Publications.
- Wöllstein-Leisten, A., Heilmann, A., Stepan, P & Vikner, S. (1997). *Deutsche Satzstruktur*. Tübingen: Stauffenburg.



# The Logical Problem of Language Acquisition Revisited: Insights from Error Patterns in Typical and Atypical Development<sup>1</sup>

Anne Mette Nyvad  
Aarhus University

## Abstract

A major impetus for understanding and building theories of language acquisition is the fact that children's grammars often deviate from adult-state grammars in intriguingly systematic ways, before converging on a grammatical system that is equivalent to that of the local linguistic community. This paper will focus on error patterns in children's non-adult structural configurations, particularly those found in a subpopulation of children diagnosed with Specific Language Impairment and Autism Spectrum Disorders. Data from language disorders may provide a prolonged window into primitives of grammar and suggest a mapping of certain genes to higher-level cognitive modules such as language. However, the heterogeneity along the developmental paths highlights the significance of the process of ontogenetic development, ultimately demonstrating that the relationship between genotype (the genetic code, i.e. the material encoding heritable traits) and phenotype (the expression of the genetic code, i.e. the observable characteristics or behavior) is quite indirect.

## 1. Introduction

It is astonishing that every typically-developing (henceforth TD) child acquires a natural language without formal instructions or scaffolding in the form of progressively sequenced linguistic input. Children thus

<sup>1</sup> Thank you to Ocke-Schwen Bohn for making linguistics feel like home to me, starting almost twenty years ago. The work presented here is funded by the Independent Research Fund Denmark (Danmarks Frie Forskningsfond, grant ID: DFF – 6107-00190).

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 449-473). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.



converge on a grammatical system parallel to that of the local linguistic community, in the face of significant variability in the linguistic input (Crain, 1991, p. 597). Considering how hard it is even for trained linguists to discern grammatical principles, it is remarkable that research on language acquisition has demonstrated that young children know them, often by the age of three.

At its core, generative grammar aims to understand the (finite) combinatorial system of rules that underlie the (infinite) range of possible sentences in the world's languages (past, present and future). The logical problem of language acquisition refers to the idea that the data that children are exposed to underdetermine what they wind up knowing about their native language, as there may not be conclusive evidence for it in the linguistic input, i.e. what is known as *poverty of the stimulus*. This raises the question of what exactly the language-acquiring child brings to this induction task (Crain, 1991; Crain and Pietroski, 2001; Thornton, 1990).

Brown (1973: 156) concludes that errors in language acquisition are "triflingly few". This paper will focus on grammatical, primarily syntactic, errors in typical and atypical language development. The errors discussed are "grammatical" in the sense that they conform to both grammar (i.e. language-specific rules) and Grammar (i.e. the underlying grammatical system common to all languages), contra Kizach's (this volume) interpretation. A further distinction needs to be clarified: This contribution will not deal with grammatical *mistakes*; in contrast to errors, a mistake is made by a learner who knows a language-specific grammatical rule, but neglects to employ it, due to performance-related or extralinguistic factors.

Language is a like an organism, a biological system, and the methods linguists use when we study it ought to reflect this. Investigations into the biology of language typically draw upon empirical data from either language acquisition, language breakdown (e.g. Broca's aphasia and Wernicke's aphasia), neuroscience (in relation to neurologically intact individuals, using fMRI, EEG, MEG, etc.) or molecular biology (scrutinizing the relation between gene expression and language). By focusing on genetic developmental disorders in language-acquiring children, this contribution combines data from all four areas. Investigations into the grammatical nature of these language disorders have previously tended to be descriptive and not rely on theoretical linguistic principles. This limits their interdisciplinary potential, as explorations of the grammatical phenomena at the interface between e.g. linguistics and neurobiology require hypotheses built on underlying principles that can be tested (or

falsified). One of the aims of this paper is to highlight a new avenue of evidence and point to a theoretical platform that can integrate language disorders into the theory of the biological underpinnings of language, as the errors made by children with Autism Spectrum Disorders (ASD) and Specific Language Impairment (SLI) are so systematic in their deviance from the target that they may reflect universal properties of grammatical structure (e.g. Universal Grammar).

Just like the TD peers, children with language disorders such as SLI and ASD are in a linguistic learning environment that is characterized by idiosyncrasies and finiteness, but even though their grammars may never reach full convergence with those of the surrounding speech communities, their productions nonetheless exhibit universal properties in the face of selective and underdetermined input.

This contribution is a review of the state-of-the-arts in language acquisition research as it relates to language disorders. It is primarily a theoretical story that calls for extensive future research on language disorders, employing the fine-grained theoretical apparatus provided by decades of research in theoretical linguistics. Section 2 discusses constraints on the hypothesis space in language acquisition, including a probabilistic model, while Section 3 and 4 examine typical development in order to put the language disorders (ASD and SLI) in a relevant context. Two linguistic arguments for innate constraints (the phenomena of medial-*wh* and structure-dependence) are summarised and evaluated in the process. Based on these considerations, section 5 goes into depth with the syntactic profiles of SLI and ASD and discusses the extent to which the phenomena found here supports the idea of innate constraints in language acquisition. Finally, section 6 debates the implications of the data for neural networks and whether it is appropriate to map the linguistic phenotypes found in SLI and ASD with specific genotypes.

## **2. Constraints on the hypothesis space in language acquisition**

According to Chomsky (1965, 1986), children are born equipped with Universal Grammar (henceforth UG), i.e. innate, biologically determined information about language. UG is envisaged as “a distinct system of the mind/brain” (Chomsky, 1986, p. 25), separate from general intelligence, and it is typically regarded as a two-tier system present *ab initio*: In the Principles & Parameters framework, restrictions on the learning space in language acquisition consist of both a hard-wired basic layer of universal *principles*, applicable to all languages, and a second layer, only partially

wired-in and subject to parametric variation, referred to as *parameters*, to which structural variation between languages are to a large extent attributed. However, the innate knowledge cannot be information about any particular language, because babies can learn all natural languages with equal ease: A Danish baby brought up in England will learn English just as easily as an English baby in Denmark will learn Danish!

Innate constraints are negative in the sense that they sanction certain constructions and hence restrict the hypothesis space that children have to contend with. Hence, rapid language acquisition would not be surprising. It is widely believed that the data for grammar construction available to the child does not include negative evidence (information about which sentences are unacceptable or ungrammatical). Negative evidence could be used by a child to avoid constructing an overly general grammar, but parents usually do not correct their children's errors, and when they do, their feedback is typically disregarded, as illustrated by Vikner's (2005, p. 3) interaction with his 5-year-old son:

- (1) Child: Ved du hvor meget jeg drikkede?  
*Know you how much I drink-ED*
- Parent: Nej, hvor meget drak du?  
*No, how much drank you*
- Child: Først drikkede jeg en hel kop te og så drikkede jeg et glas juice, og så ...  
*First drink-ED I an entire cup tea and then drink-ED I a glass juice, and then*
- Parent: Drak du så meget?  
*Drank you that much*
- Child: Ja, så meget drikkede jeg  
*Yes, so much drink-ED I*

Genuinely conservative item-based learning in the sense of MacWhinney (2004) would result in children simply parroting back what they hear, and not making the classic errors found cross-linguistically with irregular verbs where children overgeneralize the regular tense-marking, as in the example above with Danish *drikkede*. Thus, “children generalize along some dimensions but not others” (Pinker, 2004, p. 951), but given innate constraints, positive evidence should suffice for language acquisition. Even children with language disorders go through the logical stages of language acquisition, even if they do not attain full linguistic competence by adulthood (see e.g. Gernsbacher, Morson & Grace, 2015). Hence, the

grammatical errors made in e.g. SLI and ASD are not unique to those specific populations, but generally follow a trajectory similar to that found in typical development, qualitatively if not quantitatively.

## 2.1 The interplay between constraints and statistical learning

The relative contributions of biological endowment and learning in the process of language acquisition is a controversial issue. Chomsky's UG is a theory that relates to the part of language acquisition that hinges on the biological endowment. Infants have been demonstrated to employ statistics in language acquisition (Saffran et al., 1996), and these findings have been employed by Tomasello (2000, 2003) to argue against innateness. However, there is no inherent opposition between the existence of UG and the use of statistical learning (demonstrably based in part on transitional probabilities), as an effective learning algorithm requires a proper representation of the relevant learning data (cf. Yang, 2004, p. 451).

There has been a general consensus in the generative literature that parameter setting proceeds on the basis of "triggering", such that the grammar of the child (or learner, for that matter) is identified with specific parameter values, which are then modified by the input (see e.g. Gibson & Wexler, 1994). However, this "triggering" model faces problems on multiple counts: First, because the linguistic evidence that children encounter in the process of language acquisition is so variable, there is a theoretical possibility that convergence on the target grammar of the local speech community might not happen. Second, one would expect abrupt changes to the child's syntactic production when she switches between grammars. However, this is not what the empirical data suggest; instead, children appear to settle on a specific parameter quite gradually (Yang, 2004, p. 453).

This led Yang (2004) to suggest an account in which the idea of innateness is combined with a model of probabilistic learning, which he calls *the variational model* and, based on a hypothesis space built on UG-defined grammars, principles and parameters, it proceeds as follows (from Yang, 2004, p. 453):

- (a) For an input sentence,  $s$ , the language-acquiring child:
  - (i) with probability  $P_i$  selects a grammar  $G_i$ ,
  - (ii) analyzes  $s$  with  $G_i$ ,
  - (iii) if successful, rewards  $G_i$  by increasing  $P_i$ , otherwise punishes  $G_i$  by decreasing  $P_i$ .

In this model, there is selectionist competition between grammars, and only the grammar that best fits the target grammar will survive, eventually eliminating all the other possible grammars made available by UG. How long it takes for a specific parameter value to become dominant is related to the incompatibility of its competitors with the input data, its “fitness value” (Yang, 2004, p. 454). Hence, Yang argues that the triggering model of children’s language development must be abandoned and replaced with an account that conjoins the domain-specific space of UG’s principles and parameters with domain-general probabilistic mechanisms.

To sum up, according to Yang’s (2004) account of the basic mechanisms in language acquisition, variational learning hinges on the cumulative effect of language input on the one hand and UG constraints on the hypothesis space on the other. In addition, his probabilistic approach to parameter setting can be extended to account for mechanisms in language change, as the latter typically proceeds gradually diachronically and offers a foundation for variation synchronically. Thus, Yang’s (2004) model comes with the added benefit of accounting for Labov’s (2001) “enigma” in sociolinguistics, namely that speakers tend to display great uniformity in the structural aspects of language (including the error patterns), while varying greatly when it comes to other levels of linguistics.

### **3. UG-compatible errors in typical language development**

There is general agreement about the necessity of innate constraints but not about their exact nature and source (Crain, 1991, p. 597). One proclaimed source of evidence for innateness is based on children’s non-adult (but UG-compatible) question formation. Crain and Pietroski (2002, pp. 177-182) consider this type of phenomenon a genuine poverty of the stimulus argument. Employing an elicited production task, Thornton (1990) found that about one-third of the 3-4-year-old English-speaking children she studied consistently inserted an “extra” *wh*-word in their long-distance questions, as illustrated in (2):

(2a) What do you think what pigs eat? (Object WH)

(2b) Who did he say who is in the box? (Subject WH)

The emergence of the medial-*wh* in the language of children learning English cannot be explained as a response to the input, as English-speaking adults (who provide the primary linguistic data to children) do not produce medial-*wh* constructions. Although these constructions are not

grammatically well-formed in English, structures like (3) are attested in colloquial (adult) German (see also Müller, 2000, p. 54; Thornton & Crain, 1994):

(3) Wer glaubst du wer nach Hause geht? (Subject WH)

*Who think you who to house went*

“Who do you think went home?”

The acquisition-related data in (2) and the variational example in (3) are viewed as evidence supporting the idea of successive-cyclic movement, the “stopping over” of a filler undergoing long-distance movement at the left edge of the clause. This is assumed to be a universal property of language, a basic computational principle (Chomsky, 1973, 1986). As suggested by Thornton (1990) and Crain and Thornton (1998), the extra *wh*-phrase in children’s questions may be an overt manifestation of a process that appears in French when extraction occurs from subject position. In French, the alternation from *que* to *qui* takes place in subject relative clauses and subject extraction questions. An example of a subject relative demonstrating the necessity of a *qui* complementizer is given in (4) from Rizzi (1990, p. 56):

(4) L’homme que je crois \*que/qui viendra (Subject REL)

*The man who I think who will come*

“The man who I think will come”

The complementizer *que* and its alternating form *qui* both also function as *wh*-words in French. This fact is important in the account given by Crain and Thornton (1998) because their claim is that the medial-*wh* in Child English is also a complementizer, although it is similar to a *wh*-phrase in appearance (see Rizzi, 1990 for a full analysis of the phenomenon in French, and Crain & Thornton, 1998 for an analysis of the English language acquisition data).

The similarity of Child English to a foreign language extends even further. Investigation has shown that lexical (full) *wh*-phrases cannot be repeated in the medial position for both adult Germans and English-speaking children. Finally, children never employed a medial-*wh* when extracting from infinitival clauses, so they never asked questions like (5), and it is not permissible in languages that allow the medial-*wh* either (Thornton, 1990):

## (5) # Who do you want who to win?

This complex pattern of linguistic behaviour suggests that many children of English-speakers go through a stage at which they speak a language that is like adult English in many respects, but also one that is analogous to other languages in allowing for the medial-*wh* (cf. Crain and Pietroski, 2002, pp. 177-182). As pointed out by Crain and Pietroski (2001, p. 179), “similarities between child-English and adult-German are as unsurprising as similarities between cousins who have never met”. Children acquire a native language by testing wide range of the linguistic options that exist in human languages. However, they do not appear to entertain syntactic structures that would violate the constraints enforced by UG. This is known as the Continuity Hypothesis (cf. Crain, 1991; Crain and Thornton, 1998; Pinker, 1984). English-speaking children make grammatical errors that may exhibit German or Romance syntax in the absence of any evidence for these structures in the primary linguistic data.

These systematic mismatches between child and target adult language are at the core of the theoretical backbone of the stimulus poverty argument and may be the strongest argument for UG, as they demonstrate that children do not simply parrot their input or are inductively determined by it, but instead project beyond their linguistic data. Relating these data to Yang’s (2004) variational model, the presence of non-target grammars in the hypothesis space ensures a gradual syntactic development before children settle on specific parameter settings, and this would explain why children appear to only make “principled” errors that correspond to potential grammars (i.e. UG-compatible), such as medial-*wh* in Child English.

A growing body of research suggests that there are many parameter-driven plateaus in domains of syntactic development, apart from the medial-*wh* constructions (see Pierce, 1991 for an overview). In a certain sense, then, children’s errors should not merely be viewed as failures to match the target language; at any given time, they are in effect speaking a foreign language (cf. Crain and Pietroski, 2001, pp. 178-181), or a possible, natural human language, rather like the interlanguage in foreign language acquisition. The same appears to apply to children with language disorders (see section 5). These systematicities across typical development and language disorders are not only consistent with the theory of UG but may in fact be considered evidence for it.

#### 4. The non-occurrence of UG violations

At all levels of language, it is hierarchically organized, and the fact that syntactic structure operates on specific types of linguistic representations, namely constituents and phrases, rather than linear strings of words is a classic argument for the innateness of language. Chomsky (1995) proposed that the operation Merge was the Basic Property of language (Berwick & Chomsky 2016), at the core of the formation of linguistic structures. It is basically a principle of recursion, in that it combines two linguistic units,  $x$  and  $y$ , forming the composite  $(x, y)$ , which may in turn merge with  $z$  producing  $((x, y), z)$ , a hierarchical structure. However, logically speaking, even if recursive Merge is indeed a Basic Property of language, this does not mean that it is necessarily employed as an option in all languages. The theory of UG predicts that language-acquiring children do not make errors that violate innate principles and parameters, and it is a basic tenet of UG that grammatical rules are structure-dependent (cf. Chomsky, 1971, p. 1975). The structure-dependence constraint demands that syntactic derivations operate on hierarchical structure (not linear order) and hence it restricts the hypothesis space of language-acquiring children (cf. Crain and Thornton, 1998, p. 165).

Thus, one of the strongest cases of learning from inadequate evidence discussed in the literature concerns verb-initial positioning in *yes/no*-questions, e.g. *Er han tysker?* “Is he German?”, the *yes/no*-question corresponding to the declarative *Han er tysker* “He is German”. The formation of such sentence structure is structure-dependent, as it hinges on hierarchical relations: the finite verb (auxiliary or main verb *be*) in the matrix clause is assigned initial position. Chomsky (1971, pp. 29-33) gives these examples:

- (6a) The dog in the corner is hungry
- (6b) Is the dog in the corner hungry?
- (6c) The dog that is in the corner is hungry
- (6d) Is the dog that is in the corner hungry?
- (6e) \*Is the dog that in the corner is hungry?

When transforming the declarative in (6a) into an interrogative question, (6b), main verb *be* is placed in sentence-initial position. Two hypotheses regarding the formation of *yes/no*-questions can be formed on this basis: one, the first (finite) verb in the declarative is fronted, and two, the first



(finite) verb in the matrix clause is fronted. The first hypothesis would incorrectly yield (6e) on the basis of the declarative clause in (6c), while the second hypothesis, based on the structure-dependence constraint, would result in (6d) (cf. Pullum and Scholz, 2002, p. 36).

Employing an elicited production technique, Crain and Nakayama (1987) tested children's knowledge of the structure-dependence constraint. If the structure-dependence constraint is not part of children's innate knowledge and their ungrammatical productions instead constitute misgeneralizations of a structure-independent hypothesis, their errors would be expected to be random. This turns out not to be the case, and the conclusion reached by Crain and Nakayama (1987) was that children's questions provide no evidence that can be incontrovertibly employed as evidence representing violations of the structure-dependence constraint, which they thus assume to be part of UG (cf. Crain and Thornton, 1998, pp. 171-175).

If children initially formed a structure-independent hypothesis when encountering complex examples like (6c), positive evidence would not suffice to prohibit non-local movement, as it could co-exist alongside the local movement option in children's grammars (cf. Yang's variational model). Nonetheless, every language – irrespective of impairments – appears to be imposed with a restriction on non-local movement of the heads of phrases (cf. Travis' 1984 Head Movement Constraint, cf. Crain and Pietroski, 2001, p. 166). Structure-dependence is thus likely an innate constraint, a negative principle that bars certain structures (both in comprehension and production), and children do not appear to adopt grammatical analyses that are not made available by UG.

In sum, a number of language acquisition studies indicate that language-acquiring children do not make errors relating to a range of syntactic structures and dependency relations, not just structure-dependence (Crain, 1991), but also Subjacency (Newmeyer, 1991; Pinker & Bloom, 1990), *that*-trace effects (Chomsky & Lasnik, 1977), the Empty Category Principle (Chomsky, 2001), inter alia (however, see MacWhinney, 2004 for a critical review). In the presumed absence of sufficient evidence in the child's input (the poverty of the stimulus), these linguistic phenomena might hence be assumed to be innate principles. What might be perceived as even more remarkable is the fact that in a variety of language disorders, children make systematic error patterns that match the performances of TD children at an earlier stage in the process of language acquisition.

## **5. Language impairments in ASD and SLI**

The errors that language impaired children make are not random, but are constructed in a manner that appears to follow the basic architecture of the language system (see Fromkin, 1997). Thus, Levi and Kavé (1999, p. 138) suggested that language deficits may be regarded as “a natural laboratory in which linguistic theories may be tested”. The performance data that we can gather from genetic developmental language disorders such as those found in SLI and ASD may provide an extended window into both the neurobiological and computational system of language (perhaps even UG), by reflecting primitives of grammar and some of its core properties, and they have the potential of revealing important aspects of syntactic representations in the brain. In addition, data from this field can advance concepts in learnability.

According to the *Diagnostic and Statistical Manual of Mental Disorders 5* (DSM-V, American Psychiatric Association [APA] 2013), ASD and SLI, also known as developmental dysphasia, share the diagnostic trait of poor communication skills. However, in SLI, linguistic deficits and delays are at the core of the symptomatology, whereas language-acquiring children with ASD exhibit immense variability in their language abilities, ranging from absence of functional verbal abilities to fluent speech (cf. Lord et al., 2006). Pragmatic impairments, however, are ubiquitous in ASD and are thus found at both ends of the spectrum (Tager-Flusberg, 2004). Roberts, Rice & Tager-Flusberg (2004), Kjelgaard & Tager-Flusberg (2010) Zebib et al. (2013) have suggested that a subset of ASD children exhibit grammatical impairments that are reminiscent of those found in SLI.

### **5.1 The selective nature of (morpho)syntactic errors in ASD and SLI**

The exact nature of the grammatical impairments in ASD in general is largely undetermined. Early speech production-based studies carried out by Bartak, Rutter & Cox (1975) and Pierce & Bartolucci (1977) indicated that the grammatical competencies of ASD children are parallel to those of typically developing (TD) peers when the two groups are matched on mental age (see Durrleman & Delage, 2016, p. 362). However, later work (e.g. Roberts, Rice & Tager-Flusberg 2004; Zebib et al. 2013) has revealed domain-specific grammatical impairments in the ASD population that appear to be independent of domain-general cognitive deficits. SLI

is a heterogeneous family of language impairments which affects 7% of children (Lely & Pinker, 2014). Recently, the claim has been put forth that a subset of ASD children have a syntactic profile akin to that found in SLI (see e.g. Kjelgaard & Tager-Flusberg, 2010 and Zebib et al., 2013).

Children tend to leave out and/or substitute bound inflectional morphemes in SLI (Levi & Kavé, 1999) and ASD (Tager-Flusberg, 2002); speech in SLI (Leonard, 1995) and ASD (Bartolucci, Pierce & Streiner, 1980) is also characterized by omissions of free function words (e.g. articles, auxiliary verbs and conjunctions). Sentence length and complexity may also be reduced in SLI and ASD (Tager-Flusberg et al., 1990). All of these error types have parallels in typical language development, e.g. as described by Radford (1990) for English, and in Broca's aphasia (Grodzinsky, 2000). Overall, then, children with SLI and ASD (and individuals with other types of language disorders) mirror typically developing children in terms of the error patterns that they exhibit.

Lely (1996) identified a subtype of SLI relating specifically to certain aspects of syntax, morphology and phonology. She termed it Grammatical-SLI (henceforth G-SLI) and Lely & Pinker (2014, p. 586) define it as having "greater impairments in 'extended' grammatical representations, which are non-local, hierarchical, abstract, and composed, than in 'basic' ones, which are local, linear, semantic, and holistic". Lely & Pinker (2014) suggest that G-SLI is related to abnormalities in the left hemisphere. This would fit recent models of the neurobiology of language making a distinction between dorsal and ventral processing streams. As the name suggests, G-SLI does not affect language globally, but locally (or specifically) in certain properties of language, while leaving others intact (Pinker & Lely, 2014, p. 586). More specifically, it has been found cross-linguistically that children with G-SLI have both production and comprehension problems relating to syntactic dependencies in hierarchical structures, e.g. *wh*-questions, relative clauses, passive structures and syntactic embedding, especially if they involve non-canonical word orders (Lely and Battell, 2003; Hamann, 2006). In addition, they omit tense-marking on verbs (Bishop, 1979). In what Lely & Pinker (2014, p. 587) term Basic syntax (or lexical semantics), words are "inserted directly from the lexicon", whereas they have to be "computed by operations such as movement and feature checking" in Extended syntax (see Lely & Pinker, 2014, p. 587 for an extensive overview of studies that have found children in G-SLI having a contrast in performance between Extended syntax and Basic syntax).

Interestingly, a subgroup of ASD children with language impairment exhibit the same pattern, as can be gleaned from the following spontaneous productions of an 11-year-old boy with language delay and low-functioning autism from the Østergaard corpus on the *Child Language Data Exchange Systems* (CHILDES, 2015, see Østergaard, 2016 for more details):

- (7) \*ASD: Anker (...)skyde tungen derover op i fryseren.  
*Anker shoot tongue overthere up in freezer-the*
- \*ASD: <og så> [//] indtil da, <så blev> [/] så blev savl frysede.  
*and then until then then became then became saliva froze*
- \*ASD: og så sidde tunge fast.  
*and then sit tongue stuck*
- \*ASD: <og så er det nemlig sådan at så har de ehm> [//] og så er det  
*and then is it right so that then have they uhm and then is it  
 that they have got an uhm*  
 <at de har fået en ehm>  
 [//] <at de så> [//] <at ham> [//] at kommer snart med en bil.  
*that they then that him that comes soon with a car*

As exemplified in (7), this ASD child has consistent problems with irregular tense-marking (e.g. *frysede* used as a past participle instead of “frosset”). The infinitival forms *skyde* and *sidde* appear to be inserted directly from the lexicon and are uninflected for past tense (targets would be the irregular forms *skød* and *sad*). In addition, this child does not produce any structures with non-canonical word-order in this example (characteristic of his syntactic profile) and he encounters serious problems with embedding, as is evident from his multiple retracings of the complementizer *at* “that” and the fact that he ends up omitting the subject. The example in (7) is just an illustration of the (morpho)syntactic profile discussed, but it certainly warrants further investigation into the parallels between SLI and ASD (see Nyvad, 2016 for more details), as only a few studies have examined this.

Among these, Riches et al. (2010) found that adolescents with ASD and SLI perform significantly less accurately than TD peers in a sentence-repetition task involving subject and object relatives, such as:

- (8a) The thief that \_\_\_ robbed the granny (Subject REL)  
 (8b) The granny that the thief robbed \_\_\_ (Object REL)

Both groups (ASD and SLI) tended to make performance errors because they wanted to avoid structures with non-canonical word-order (complex and part of Extended syntax in Lely and Pinker's 2014 sense). In (8a), there is canonical word-order, as the subject of the main verb *robbed* precedes the direct object *the granny*, whereas it follows it in (8b), resulting in increased syntactic complexity. Complex (morpho)syntax requires more processing capacity as it involves more working memory load. For instance, in the object relative in (8b), the filler (the relative element) has to be held in working memory longer than is the case for the subject relative in (8a). However, when matched with a control group in terms of working memory capacity, individuals with G-SLI still appear to experience more problems relating to Extended syntax, according to Lely and Pinker (2014) (see Tager-Flusberg, 1981 and Van der Lely, 1996).

However, an asymmetrical pattern in the performance on subject and object relatives, cf. (8), is by no means unique to SLI and ASD. A great variety of individuals with language impairment have been demonstrated to have a better comprehension of sentences with canonical word-order than those where elements have been displaced. This is also true for Broca's aphasia (Grodzinsky, 2000), Wernicke's aphasia (Bastiaanse & Edwards, 2004), Alzheimer's disease (Grober & Bang, 1995), Down's syndrome (Ring & Clahsen, 2005) and for children who sustain focal brain damage (Dick et al., 2004), especially when it is localized in the left hemisphere (Dennis & Whitaker, 1976) (see Penke, 2015 for an excellent overview).

The neural organization of language can be gleaned through new technologies such as functional magnetic resonance imaging (fMRI), electroencephalography (EEG) and magnetoencephalography (MEG), and so far they indicate that the neural networks supporting Extended syntax is different from the ones that form the basis of Basic syntax. New models of the neural organization of language outlined in Lely & Pinker (2014) offer a more fine-grained picture by transcending the basic distinction between Broca's and Wernicke's areas. Three distinct fronto-temporal networks appear to be related to the processing of syntax. A dorsal pathway seems to be particularly related to Extended syntax, namely one that connects Broca's area (specifically Brodmann area 44) to Wernicke's area (in the posterior superior temporal gyrus) via the arcuate fasciculus. This neural pathway does not mature fully until the child reaches the age of approximately 7. As pointed out by Lely & Pinker (2014, p. 590), "the dorsal pathways in human brains differ substantially from those in other primates, suggesting that phylogenetic changes to the dorsal pathway may have been a key driver of the evolution of language".

SLI and ASD are thought of as separate disorders with distinct aetiologies (cf. Bishop, 2003, p. 214). However, impairments in Extended syntax appear to be common to a subpopulation of both groups, and the dissociation between language and cognition is also found in SLI, which has led a number of researchers to consider whether SLI and ASD are on a continuum (Tager-Flusberg, 2004; Bishop, 2003, 2010).

### **5.2 Are the linguistic deficits in SLI and ASD on a continuum?**

Interestingly, SLI and ASD were considered mutually exclusive diagnoses on the DSM-IV, but they no longer are on the DSM-V (see Durrleman & Delage, 2016, p. 361). This illustrates how the linguistic impairments of SLI and ASD may be considered deficits on a continuum of severity, such that milder cases would only involve problems in syntax whereas both syntax and pragmatics are affected in more severe cases. However, such a view would predict that pragmatic deficits should be manifest in those with the most severe syntactic impairments, and this prediction does not square with the facts: Pragmatic difficulties are ubiquitous in ASD, whereas syntactic deficits are only present in a subgroup (cf. Kjeldgaard & Tager-Flusberg, 2000). In other words, there is a double dissociation between syntax and pragmatics in the two types of language disorder: In ASD, syntax is not uncommonly unaffected while pragmatics is impaired, and in SLI, syntax may be impaired and pragmatics unaffected. This lack of logical dependency between the two levels of linguistics in SLI and ASD suggests that they have “distinct neurological bases” (Bishop, 2003, pp. 219-220). In other words, if you view SLI and ASD as being on the same spectrum, you have to do so for each linguistic level of description in isolation, as the deficits in pragmatics phenotypically are not continuous with the observed impairments in syntax (cf. Bishop, 2003, p. 224).

The symptom-overlap in terms of parallels in syntactic impairment in SLI and ASD may hint at a shared aetiology, but these surface correspondences become even more striking viewed against the backdrop of genetic studies involving relatives of people with autism for whom it is common to exhibit subthreshold symptomatology which resembles SLI (Bishop, 2003, pp. 218-219). The neurological bases for ASD and SLI can thus be envisaged as being distinct, but common aetiological factors may be implicated. Bishop (2003, p. 222) further proposes that there may be genes that “disrupt processes of neuronal migration, leading to abnormal brain structure”. The effect of this disruption will be dependent upon which neural networks are involved, and the correlations found in

the symptomatologies of SLI and ASD may thus reflect overlaps in the implicated neural networks. The specificity found in the syntactic profiles of ASD and SLI, reviewed in this section, may be a reflection of an underlying division in the neural and genetic substrates of language (cf. Lely & Pinker, 2014, p. 586).

## **6. On the mapping of genotypes and linguistic phenotypes**

Chomsky argues for the biological model of language development. Given that only humans can acquire grammatical rules, language must partly derive from the genome. Indeed, genetics appear to be able to interfere with language (e.g. genetic region SPCH1, chromosome 7q31). Smith & Tsimpli (1995, p. 31) suggested that the human mind “is equipped with a body of genetically determined information specific to Universal Grammar”, and Chomsky (2012) proposes that human language originates from a single genetic mutation (and hence, that it did not evolve gradually through natural selection). However, this theory is complicated by the fact that hundreds of genes (out of a total of approximately only 24,000 in humans) contribute to the development and functioning of the neural substrate of language (Benítez-Burraco, 2009).

It is a truism that our genes “code for a brain that can learn language” (Karmiloff-Smith et al., 2002, p. 312), but the suggestion that there are “grammar genes” in the sense that information specific to the domain of grammar is pre-wired in the genes is highly contentious. The temptation to map genes and cognitive modules 1:1 largely comes from genetic developmental disorders such as SLI, where syntax can be selectively impaired, while other domains of language are seemingly preserved, as described above. However, based on recent research, one might need to be wary of suggesting that there is a gene or even a set of genes for e.g. syntax. The relation between genotype and phenotype is far too indirect and complex, and mapping between specific genes and higher-order cognitive modules such as language in general or grammar in particular is (still) untenable. Ontogenetic development (e.g. biochemical, nutritional and social experience) plays a crucial role at this complex interface, and it may sometimes lead to SLI, sometimes to ASD, and sometimes to an intermediate clinical picture (cf. Bishop, 2003, p. 224). In fact, Karmiloff-Smith et al. (2002, p. 318) point out that, even the discovery of “a gene (*y*) for *x*” (where absence of gene *y* correlates with phenotype *x* in a developmental disorder) does not entail that the absence of *y* is the sole

cause of phenotype *x*. More likely, according to Karmiloff-Smith et al. (2002, p. 318), gene *y* is a component part of a collection of genes coding for molecular processes responsible for constructing the brain.

So far, the development of SLI has been associated with at least four candidate genes and it is believed to be exceedingly heritable, as is the case for ASD. These facts suggest that research into language disorders such as SLI and ASD may provide information about the intricate relationship between nature and nurture on the one hand and the biological underpinnings of language on the other. However, a simple 1:1 mapping with the phenotypic outcome is implausible, because the genes in question may have a number of both cognitive and physical effects (see Karmiloff-Smith et al., 2002). Further, several of these so-called “language genes” are polymorphic in the sense that they may or may not lead to language impairment, depending on which variant is present in the genome. In addition, the same pathogenic allele can lead to different developmental disorders (cf. Benítez-Burraco & Boeckx, 2014). This heterogeneity (both with respect to genetic make-up and symptomatology), also found in connection with ASD, has led Lely & Pinker (2014, p. 586) to recommend that instead of trying to find a direct link between genotype and linguistic phenotype, it would be more fruitful to search for links between “genetic variants with alterations in the neural substrates of subcomponents of language processing”. Anatomically speaking, the neural underpinnings of language are difficult to pinpoint, as the functional areas of the brain vary from person to person. These variations are astounding in light of the fact that, in the normal population, and to a more limited extent in the impaired population, they converge phenotypically on the same grammatical system, and the errors made along the developmental path adhere to universal principles.

Talk of a linguistic genotype that is equated with UG conflates nativism with geneticism (cf. Benítez-Burraco & Boeckx, 2014). What is striking, nonetheless, is that language pathologies such as SLI and ASD (but also Broca’s aphasia) do not hit random areas of the linguistic system. Quite systematically, they appear to affect inflectional morphology and complex syntax. It may be the case that the neural network implicated in these processes (the neural substrate and the dorsal pathway that supports it) is so spread-out and complex that its sheer intricacy makes it vulnerable, as any disturbance in the system would break it down (for an analysis, see e.g. Nyvad, 2018). These dorsal pathways may, however, also be engaged in computations relating to other high-level cognitive functions. Thus, for



example, Broca's area and its vicinity may be specialized for computations of a hierarchical, complex nature, but these computations may not only be relevant for syntax, but also other cognitive functions. As pointed out by Benítez-Burraco & Boeckx (2014, p. 6), "it seems that it is only their basic architecture that is genetically encoded, while their functional specificities are environmentally driven". The functional variability is, however, fairly confined. As pointed out by Grodzinsky (2010), the areas of the brain that are activated when processing language appear to be relatively uniform across individuals. Benítez-Burraco (2009) states that, at the molecular level, "a core set of genetic cues" are responsible for "the initial wiring of the linguistic brain...in all subjects" (Benítez-Burraco & Boeckx, 2014, p. 6). The specific parallels in linguistic phenotypes can thus emerge from quite diverse brain architectures and genotypes.

To sum up, only certain structures appear in grammars, be they normal or impaired, delayed or broken down. This parallels what is found in language variation and change where only particular elements of the grammar are subjectable to variability. All of this points to the existence of a *genotype grammar* (an underlying grammar, common to all languages, referred to as UG) which leads to *phenotypic grammars* (the observable variations in grammars – typical or atypical - in the world's languages). The study of language disorders is a new frontier of research which can be a powerful tool to help us understand the biological underpinnings of language. The linguistic description of SLI has been advanced significantly by theories grounded in UG, and the next step is to carry out investigations into ASD while applying the same theoretical apparatus. As pointed out by Lely & Pinker (2014, p. 593), future research has to take an interdisciplinary approach that takes full advantage of the fine-grained analyses offered in linguistics, instead of coarsely mapping genotype (genetic variants) directly unto phenotypes (overall language impairment), as is largely the case in the extant literature on language disorders.

## 7. Conclusion

Gene expression and experience (both linguistic and non-linguistic) interact in the development of grammar. By adulthood typically-developing language users have mastered a rich and complex linguistic system. However, while adhering to e.g. structure-dependence, children's grammars deviate from adult grammars in intriguingly systematic and constrained ways – both in typical and atypical development, which would

be surprising if there are no underlying restrictions in the hypothesis space forming the patterns. Any theory of language acquisition, be it relating to typical development or language disorders, should thus provide an account of why children project beyond their experience in certain ways but not in others. While these deviances from adult grammar and the selective nature of syntactic impairments across language pathologies is not indisputable evidence for the existence of UG, it is surely a strong case for it. Especially in light of the fact that certain genetic mutations in language disorders are associated with these specific linguistic patterns.

That children use statistical information should be viewed against backdrop of them knowing what linguistic units are relevant and important for cracking the code of the specific language that they are acquiring (be it stress patterns, segments, word classes, syntactic dependency relations, etc.). An explanation of the data from ASD and SLI in terms of statistical learning in the absence of innate constraints would have to (implausibly) assume that these children (and adults) have a deficiency in their ability to extract knowledge from statistical regularities in their data.

There is a continuum concerning the degree to which individuals make errors in performance that cuts across the divide between impaired and unimpaired language (cf. Penke, 2015). It thus appears that the grammatical errors in language disorders like SLI and ASD are gradable and not qualitatively different from those found in typically-developing children. Without taking away the importance of UG, the variability in the performance data from language disorders may in part be explained with reference to processing capacities, such that limitations on working memory or short-term memory can impede the extraction of (morpho-) syntactic information when the latter is complex. This type of account may be able to capture the gradience in performance within and across language disorders (as well as within and across typical language), as variability in processing load may engender variability in performance. In addition, we expect people with e.g. ASD to make significantly more mistakes than neurotypical children, simply due to cognitive constraints that are domain-general, rather than domain-specific.

However, while this may help explain the gradience observed, it cannot answer for the systematicity in the error patterns without appealing to syntactic representation and perhaps ultimately innate constraints. The characteristics of the errors found in the performance of language-acquiring children and individuals with a wide range of acquired and developmental deficit syndromes (like ASD and SLI) all appear to involve the uppermost

projections of the syntactic tree (i.e. the left periphery of the clause). The latter has a crucial role in the comprehension and production of syntactically complex sentences with e.g. non-canonical word-order or embedding. If G-SLI and the subset of ASD children exhibiting syntactic impairments have a shared aetiology, the linguistic phenotype would be expected to be more or less identical. Demarcating what exactly it comprises might not only lead to improvements in training methods, but more fundamentally strengthen our understanding of how genes affect brain circuitry in healthy and pathological language profiles. Today, a lot of research is dedicated to integrating the formal/theoretical approach with behavioral/experimental studies, and future research must employ a fine-grained analysis of how distinctive linguistic components correlate with anatomical and functional aspects of the human brain. Whether there is an underlying neural deficit that manifests itself in this relative uniformity across disorders is still an open question.

## References

- American Psychiatric Association (APA) (2000). *The diagnostic and statistical manual of mental disorders* (4<sup>th</sup> ed. Revised). Washington.
- American Psychiatric Association (APA) (2013). *The diagnostic and statistical manual of mental disorders* (5<sup>th</sup> edition). Arlington, VA: American Psychiatric Publishing.
- Bartak, L., Rutter, M. & Cox, A. (1975). A comparative study of infantile autism and specific developmental receptive language disorder: I. The children. *British Journal of Psychiatry* 126(2), 127-145.
- Bartolucci, G., Pierce, S. J. & Streiner, D. (1980). Cross-sectional studies of grammatical morphemes in autistic and mentally retarded children. *Journal of Autism and Developmental Disorders* 10(1), 39-50.
- Bastiaanse, R. & Edwards, S. (2004). Word order and finiteness in Dutch and English Broca's and Wernicke's aphasia. *Brain and Language* 89, 91-107.
- Benítez-Burraco, A. (2009). *Genes y lenguaje*. Barcelona: Reverté.
- Benítez-Burraco, A. & Boeckx, C. (2014). Universal Grammar and Biological Variation: An EvoDevo Agenda for Comparative Biolinguistics. *Biological Theory*. DOI 10.1007/s13752-014-0164-0.
- Berwick, R. C. & Chomsky, N. (2016). The biolinguistics program: the current state of its development. In A. M Di Sciullo, & C. Boeckx (Eds.), *The biolinguistics enterprise* (pp. 19-41). Oxford: Oxford University Press.

- Bishop, D. V. M. (1979). Comprehension in Developmental Language Disorders. *Developmental Medicine & Child Neurology* 21(2), 225-238.
- Bishop, D. V. M. (2003). Autism and Specific Language Impairment: Categorical distinction or continuum? In G. Bock & J. Goode (Eds.) *Autism: neural basis and treatment possibilities: Novartis Foundation Symposium* (pp. 213-226). Chichester, UK: John Wiley & Sons.
- Bishop, D. V. M. (2010). Overlaps between autism and language impairment: Phenomimicry or shared etiology? *Behavior Genetics* 40, 618-629.
- Brown, R. (1973). *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.
- Child Language Data Exchange Systems (CHILDES), <http://childes.psy.cmu.edu/manuals/CLAN.pdf> (October 2015 version).
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky, N. (1971). *Problems of knowledge and freedom*. Pantheon Books.
- Chomsky, N. (1973). "Conditions on transformations". In S. Anderson & P. Kiparsky (Eds.), *A festschrift for Morris Halle* (pp. 232-286). New York: Holt, Reinhart & Winston.
- Chomsky, N. (1975). *Reflections on language*. Pantheon Books.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin and Use*. Praeger.
- Chomsky, N. (2001). Beyond Explanatory Adequacy. *MIT Occasional Papers in Linguistics* 20, 1-28.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, N. (2012). *The Science of Language*. Cambridge University Press.
- Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences* 14, 597-650.
- Crain, S. and Nakayama, M. (1987). Structure dependence in grammar formation. *Language* 63, 522-543.
- Crain, S. and Pietroski, P. (2001). Nature, nurture and universal grammar. *Linguistics and Philosophy* 24, 139-186.
- Crain, S. and Pietroski, P. (2002). Why language acquisition is a snap. *The Linguistic Review* 19, 163-183.
- Crain, S. and Thornton, R. (1998). *Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge Mass.: MIT Press.
- Dennis, M. & Whitaker, H. (1976). Language acquisition following hemide-ortication. *Brain and Language* 3, 404-433.
- Dick, F., Wulfeck, B., Krupa-Kwiatkowski, M., & Bates, E. (2004). The development of complex sentence interpretation in typically developing children compared with children with specific language impairments or early unilateral focal lesions. *Developmental Science* 7(3), 360-377.

- Durrleman, S. & Delage, H. (2016). Autism Spectrum Disorder and Specific Language Impairment: Overlaps in syntactic profiles. *Language Acquisition* 23(4), 361-386.
- Fromkin, V. (1997). Some thoughts about the brain/mind/language interface, *Lingua* 100, 3-27.
- Gernsbacher, M. A., Morson, E. M. (2015). Language development in autism. In G. Hickok & S. Small (Eds.), *Neurobiology of language* (pp. 879-886). Academic Press.
- Gibson, E. & Wexler, K. (1994). Triggers. *Linguistic Inquiry* 25, 355-407.
- Grober, E. & Bang, E. (1995). Sentence comprehension in Alzheimer's disease. *Developmental Neuropsychology* 11, 95-107.
- Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences* 23, 1-71.
- Hamann, C. (2006). Speculations About Early Syntax: The Production of Wh-questions by Normally Developing French Children and French Children with SLI. *Catalan Journal of Linguistics* 5, 143-189.
- Karmiloff-Smith, A. et al. (2002). Different Approaches to Relating Genotype to Phenotype in Developmental Disorders. *Developmental Psychobiology* 40(3), 311-322.
- Kjelgaard, M. & Tager-Flusberg, H. (2010). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes* 16(2-3), 287-308.
- Labov, W. (2010). *Principles of linguistic change*. Oxford: Blackwell.
- Lely, H. K. (1996). Specifically language impaired and normally developing children: verbal passive vs adjectival passive sentence interpretation. *Lingua* 98, 243-272.
- Lely, H. & Battell, J. (2003). Wh-movement in children with grammatical SLI: a test of the RDDR hypothesis. *Language* 79, 153-181.
- Lely, H. & Pinker, S. (2014). The biological basis of language: insight from developmental grammatical impairments. *Trends in Cognitive Sciences* 18(11), 586-595.
- Leonard, L. B. (1995). Functional categories in the grammars of children with specific language impairment. *Journal of Speech and Hearing Research* 38, 1270-1283.
- Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.
- Levi, Y. & Kavé, G. (1999). Language breakdown and linguistic theory: a tutorial overview. *Lingua* 107, 95-143.
- Lord, C. et al. (2006). Autism from two to nine. *Archives of General Psychiatry* 63(6), 694-701.
- MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language* 31(4), 883-914.

- Müller, G. (2000). *Elemente der optimalitätstheoretischen Syntax*. Linguistik 20. Stauffenburg.
- Newmeyer, F. J. (1991). Functional explanation in linguistics and the origins of language. *Language & Communication* 11(1-2), 3-28.
- Nyvad, A. M. (2016). Syntaksens rolle i sprogtilægnelsen hos autistiske børn. In: I. Schoonderbeek Hansen, T.T. Hougaard & K.T. Petersen (Eds). 16. *Møde om Udforskningen af Dansk Sprog*, 287-299.
- Nyvad, A. M. (2018). Deficit or Delay in the Acquisition of Complex Syntax in Autism?. *Acta Neuropsychiatrica* 30(2), 38.
- Penke, M. (2015). Syntax and language disorders. In: T. Kiss & A. Alexiadou (Eds.), *Handbooks of Linguistics and Communications Science* 42(3), 1833-1874.
- Pierce, A. E. (1991). Acquisition errors in the absence of experience. *Behavioral and Brain Sciences* 14, 628-629.
- Pierce, S. & Bartolucci, G. (1977). A syntactic investigation of verbal autistic, mentally retarded and normal children. *Journal of Autism and Childhood Schizophrenia* 7, 121-134.
- Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.
- Pinker, S. (1990). Language Acquisition. In D.N. Osherson & H. Lasnik (Eds.), *Language: An Invitation to Cognitive Science*. Cambridge Mass.: Bradford Books: MIT Press.
- Pinker, S. (2004). Clarifying the logical problem of language acquisition. *Journal of Child Language* 31, 949-953.
- Pinker, S. & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences* 13(4), 707-727.
- Pullum, G. and Scholz, B. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19, 8-50.
- Radford, A. (1990). *Syntactic Theory and the Acquisition of English Syntax: The Nature of Early Child Grammars of English*. Cambridge, Mass.: Blackwell.
- Riches, N. G. T. et al (2010). Sentence repetition in adolescents with specific language impairments and autism: An investigation of complex syntax. *International Journal of Language & Communication Disorders* 45(1), 47-60.
- Rizzi, L. (1990). *Relativized Minimality*. Cambridge, Mass.: MIT Press.
- Roberts, J. A., Rice, M.L. & Tager-Flusberg, H. (2004). Tense marking in children with autism. *Applied Psycholinguistics* 25, 429-448.
- Saffran, J. R. et al. (1996). Statistical learning by 8-month old infants. *Science* 274, 1926-1928.
- Sampson, G. (2002). Exploring the richness of the stimulus. *The Linguistic Review* 19, 73-104.
- Scholz, B. and Pullum, G. (2002). Searching for arguments to support linguistic nativism. *The Linguistic Review* 19, 185-223.

- Smith, N. & Tsimpli, I.-M. (1995). *The Mind of a Savant. Language Learning and Modularity*. Oxford: Blackwell.
- Tager-Flusberg, H. (1981). Sentence comprehension in autistic children. *Applied Psycholinguistics* 2, 5-24.
- Tager-Flusberg, H. (1996). Current theory and research on language and communication in autism. *Journal of Autism and Developmental Disorders* 26, 169-172.
- Tager-Flusberg, H. (2002). Language impairment in children with complex neurodevelopmental disorders: the case of Autism. In Y. Levi & J. Schaeffer (Eds.), *Language Competence across Populations – Towards a Definition of Specific Language Impairment in Children* (pp. 297-321). Mahwah NJ: Lawrence Erlbaum.
- Tager-Flusberg, H. (2004). Do autism and specific language impairment represent overlapping language disorders? In M.L. Rice & S.F. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 31-52). Mahwah, NJ: Lawrence Erlbaum.
- Tager-Flusberg, H., Calkins, S., Nolin, T., Baumberger, T., Anderson, M. & Chudwick-Dias, A. (1990). A longitudinal study of language acquisition in autistic and Down syndrome children. *Journal of Autism and Developmental Disorders* 20(1), 1-21.
- Thornton, R. (1990). Adventures in long-distance moving; the acquisition of complex Wh-questions. Unpublished Ph.D. dissertation, University of Connecticut.
- Thornton, R. and Crain, S. (1994). Successful cyclic movement. In T. Hoekstra and B. Schwartz (Eds.) *Language Acquisition Studies in Generative Grammar* (pp. 215-253). John Benjamins Publishing Company.
- Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics* 11(1/2), 61-82.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA.: Harvard University Press.
- Travis, L. (1984). Parameters and effects of word order variation. Unpublished Ph.D. dissertation. Massachusetts Institute of Technology.
- Vikner, S. (2005). Generativ lingvistik: Optimalitetsteori og komparativ syntaks. *Fakultetets ForskningsFredage*. Københavns Universitet, 08.04.2005.
- Yang, C. (2004). Universal Grammar, statistics, or both?. *Trends in Cognitive Sciences* 8, 451-456.
- Yang, C. et al (2017). The growth of language: Universal Grammar, experience, and principles of computation. *Neuroscience and Biobehavioral Reviews* 81, 103-119.

- Zebib, R. et al (2013). Formal language impairment in French-speaking children with ASD: A comparative ASD/SLI study. In S. Stavrakaki, M. Lalioti & P. Konstantinopoulou (Eds.), *Advances in language acquisition* (pp. 472-480). Newcastle: Cambridge Scholars Publishers.
- Østergaard, J. S. (2016). *Retelling strange stories: An examination of language and socio-cognition for children with autism spectrum disorders*. Unpublished Ph.D. dissertation, Aarhus University.





# **SECOND LANGUAGE ACQUISITION**

Handling editor: Anders Højen



## Contributions of Cognitive Attention Control to L2 Speech Learning

Joan C. Mora and Ingrid Mora-Plaza  
Universitat de Barcelona

### Abstract

This study examined the effect of cognitive attention control on L2 phonological development from an individual differences perspective. L1-Catalan/Spanish learners of L2-English were trained on the perception and production of English /æ/-/ʌ/ and /i:/-/ɪ/ through AX discrimination, identification immediate repetition tasks. Learners' gains in L2 phonological development were assessed through L2 perception (ABX discrimination and lexical decision) tests. Additionally, we obtained individual measures of auditory selective attention, auditory attention switching and auditory inhibition and a measure of overall L2 proficiency. Results revealed robust gains in L2 perception for both target contrasts. Auditory selective attention scores were significantly related to learners' gains in /æ/-/ʌ/ perception, and attention switching skills to performance in the ABX discrimination tests. Overall the results highlight the role of cognitive attention control in L2 speech learning.

### 1. Introduction

Speaking in a second language (L2) requires listeners and speakers to efficiently switch their focus of attention between competing linguistic cues as required by the context of communicative interaction. Exercising successful control of attention in L2 use is therefore essential for communication, but the human attentional system is of limited capacity (Petersen & Posner, 2012) and individuals vary in how efficiently they can shift, focus and maintain their attention when using a L2 (Wager, Jonides, &

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 477-499). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.

Smith, 2006). In addition, whereas the use of attention control in language processing is highly automatized in one's first language, it appears to be relatively effortful and inefficient in a L2 (Segalowitz, 2010), especially at lower levels of proficiency. This suggests that individual differences in cognitive attention control, as well as in other cognitive components of executive functioning (e.g. inhibitory control and working memory), may have a substantial impact on second language acquisition and may therefore constitute relevant sources of variability that can help explain the large inter-learner variability commonly associated with L2 phonological development.

Previous research has shown that inter-learner differences in attentional capacity may have an overall impact on L2 learning either enhancing or impairing lexical, grammatical or phonological development (Segalowitz & Frenkiel-Fishman, 2005). However, cognitive attention control may not impact all linguistic domains to the same extent. For example, it may impact efficiency of L2 processing by regulating the shift of focus between form and meaning when processing utterances, or by selectively attending to various linguistic dimensions (phonology, morphology, grammar or semantics), or by focusing attention on a specific relevant feature within a linguistic dimension (e.g. duration differences in phonological encoding). In addition, languages also differ in what linguistic features speakers' attentional resources need to be allocated to and in how attention is variably allocated to these linguistic features. In L2 phonological processing, inefficient use of this attentional skill may cause perceptual difficulties for adult L2 learners who may fail to apply some L2-specific cue-weighting when phonologically encoding L2 sounds (Bohn, 1995; Flege, 1995).

Attention control has also been shown to be related to general mechanisms involved in the perception and production of speech by guiding auditory processes during speech perception. It allows listeners to focus their processing resources on the relevant acoustic information and to select the acoustic information that is critical for appropriately interpreting the auditory input during oral communication (Akeroyd, 2008; Astheimer, Berkes, & Bialystok, 2016; Baese-Berk et al., 2015). The use of such attention control mechanisms, which requires both attention-switching skill and the ability to selectively attend to a single dimension or feature during speech processing (Astheimer & Sanders, 2009; Bialystok, Craik, & Luk, 2012), facilitates perceptual learning and L2 learners' skill in processing L2 phonological contrasts (Ou, Law & Fung, 2015).

In the domain of L2 phonological acquisition, better attention switching skills have been associated to enhanced performance in L2 phonological processing tasks (Darcy, Mora & Daidone, 2014; Mora & Darcy, 2016). Learners with better attention control may thus be better able to make use of the phonological features embedded in L2 speech input to guide their perceptual learning process. Few studies to date have examined the relationship between cognitive attention control and L2 perception gains obtained through phonetic training, with apparently mixed findings. For example, Ghaffarvand Mokari and Werner (2018) recently examined the role of attention control (measured through the Stroop task) in vowel learning after two weeks of perceptual training on English vowels through discrimination and identification tasks and found no significant associations between attention scores and perceptual learning. However, Hazan and Kim (2010), investigated predictors of phonetic training benefits in the context of phonetic training and found that attention switching, as measured through one of the components of the Test of Everyday Attention (TEA) (Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994), correlated significantly with gains in word identification. Altogether these findings suggest that the impact of cognitive attention control in L2 speech learning is still not well understood. In addition, the role of the various components of attention control (attention switching, selective attention and inhibition) in L2 phonological acquisition are largely under-researched, especially as measured in the auditory domain. In the present study we measure L2 learners' efficiency in the use of their attentional resources through attention control tasks that require participants to recruit their attentional resources in the processing of L1 and L2 speech.

The goal of the present study is to explore the relationship between cognitive attention control and L2 phonological development. We examined individual differences in three subcomponents of cognitive attention control in the auditory domain (selective attention, attention switching and inhibition) and related these scores to L1-Spanish learners' performance and gains in perceptual sensitivity to two difficult L2-English vowel contrasts (/æ/-/ʌ/ and /i:/-/ɪ/) on which they had been trained through discrimination and identification tasks.

## **2. Methods**

L1-Spanish English learners were tested on their ability to accurately perceive and produce two difficult L2 vowel contrasts (/æ/-/ʌ/ and /i:/-/ɪ/) before and after four 45-minute phonetic training sessions. We

assessed L2 perception through ABX discrimination and lexical decision tasks. L2 production was assessed through delayed repetition tasks. The L2 phonetic training consisted of AX discrimination and identification tasks (perception) and two immediate repetition tasks (production). All the tasks were administered in *DmDx* (Forster & Forster, 2003) on laptop computers using noise-cancelling headphones. We used three auditory attention control tasks involving the learners' L1 and L2 to assess individual differences in their attentional skills: an auditory selective attention task, an auditory attention switching task, and an auditory inhibition task. In addition, we obtained a measure of overall L2 proficiency through an elicited imitation task. In the present chapter we report on the learners' perceptual performance only.

## 2.1 Participants

The participants in the study ( $N=17$ , 14 female) were Catalan-Spanish bilingual undergraduate learners of English who participated in this research for course credit. They had learnt English mainly through formal instruction at school and had limited weekly exposure to English (Table 1). They could all speak Catalan and Spanish but varied in degree of dominance (6 Catalan-dominant, 4 Spanish-dominant, 7 balanced), which was not expected to affect their perception and production of the target vowel contrasts, as both /æ/-/ʌ/ and /i:/-/ɪ/ are mapped onto the same L1 Spanish and Catalan vowels (/a/ and /i/, respectively). They reported having no speech or hearing pathologies.

Measure	<i>M</i>	<i>SD</i>
Age at testing (years)	22.06	9.33
Age of onset of L2 learning (years)	7.35	5.02
L2 instruction (years)	14.53	2.66
Spoken L2 input / output (hours per week) <sup>1</sup>	22.61 / 11.14	11.29 / 6.87
Self-estimated proficiency (1=very poor-9=native-like) <sup>2</sup>	6.32	1.11

<sup>1</sup>L2 use with native and non-native speakers in hours per week.

<sup>2</sup>Averaged self-estimated ability to speak spontaneously, understand, read, write and pronounce English.

Table 1. Participants' demographics

## **2.2 Materials**

The materials used in the training and testing consisted of 32 nonwords and 16 words for each one of the 4 vowels in the target vowel contrasts (/æ/, /ʌ/, /i:/, /ɪ/). They were elicited in carrier phrases (*I say X, I say X again*) read by 3 female (F1, F2, F3) and 3 male (M1, M2, M3) native speakers of Southern British English. Carrier phrases were digitally recorded in a soundproof booth and the best of the two target items in each carrier phrase was excised and normalized for amplitude. Half of the nonwords (16) were 3-syllable nonwords containing the target vowels in stressed position (*fadattick* /fə'dætɪk/, *faduttick* /fə'dʌtɪk/, *fadeetick* /fə'di:tɪk/, *fadittick* /fə'dɪtɪk/) and half of them were 1-syllable nonwords consisting of the same stressed syllable (*datt* /dæt/, *dutt* /dʌt/, *deet* /di:t/, *ditt* /dɪt/). The consonants preceding and following the stressed vowel in the target CVC syllables varied in place of articulation and voicing, and so did the consonants in the initial and final unstressed syllables, which always had a weak unstressed vowel (either /ə/ or /ɪ/; e.g. C/ə/-CVC-/ɪ/C). Half of the words were 1-syllable common English words (*cap, cup, feet, fit*) and half were common 2-syllable words (*ankle, uncle, feeling, filling*). Stimuli from 4 of the 6 speakers (F1, F2, M1, M2) were used in the training, whereas the remaining two voices (F3, M3) were used in the testing only. During the training participants were exposed to 1- and 3-syllable nonwords only, whereas the testing included both nonwords and words. This was done to test whether the phonetic training based on nonwords (and therefore void of lexical meaning) was effective in improving the perception of the target vowel contrasts in known lexical items.

## **2.3 Phonetic training**

The phonetic training sessions consisted of 4 45-minute training sessions, 2 sessions per week with a day in between. The order of the training tasks was consistent across all 4 sessions: AX discrimination, identification, and immediate repetition. Participants were trained on the production and perception of the two target vowel contrasts separately in two blocks within every session. The blocks were counterbalanced across participants. The stimuli were distributed across the training sessions in order of increasing linguistic complexity, so that participants were exposed to 1-syllable nonwords only in sessions 1 and 2 and to 3-syllable nonwords only in sessions 3 and 4. However, participants were exposed to all 4 speakers (F1,



F2, M1, M2) within a single session in every training task to ensure exposure to speaker variability. The order in which participants were trained on the two target contrasts (/æ/-/ʌ/ and /i:/-/ɪ/) was counterbalanced within participants across sessions. The same nonwords by the same speakers were used for perception and production training.

### 2.3.1 AX Discrimination

In the AX discrimination tasks participants were exposed to a total of 1152 minimal pair trials, 576 trials per vowel contrast, that is, 144 trials per contrast in each session. All AX trials consisted of two nonwords produced by two different voices presented with a 500ms inter-stimulus interval (ISI). Participants were exposed to the same number of *same* (AA, BB) and *different* (AB, BA) trials, where the 4 speakers' voices appeared equally frequently in all positions producing the two members of each vowel contrast in all possible orders. Out of the 144 trials per contrast in one session (6 minimal pair nonword pairs x 4 trial orders = 24 trials), half of the nonword pairs, 72 trials (6 voice combinations x 3 minimal pair nonword pairs x 4 trial orders) corresponded to trials with an initial female voice combination (F1-F2, F1-M1, F1-M2, F2-F1, F2-M1, F2-M2), whereas 72 trials corresponded to initial male voice combinations (M1-M2, M1-F1, M1-F2, M2-M1, M2-F1, M2-F2). The nonword pairs participants were exposed to were different in each training session so that they were exposed to a total of 12 different 1-syllable nonword pairs in sessions 1 and 2 and a total of 12 different 3-syllable nonword pairs in sessions 3 and 4.

During training participants were instructed to decide, as fast and as accurately as they could, whether the two (non)words they heard were the *same* (i.e. contained the same stressed English vowel) or *different* by pressing a designated labelled key on the computer keyboard. These instructions were provided in English orally by the researchers and in written form on the computer screen. Participants performed 6 practice trials to ensure they understood the task, after which they could ask questions if they had doubts. The 144 trials were presented in fully randomized order. After 72 trials participants could take a break if they wished. Participants received visual feedback for error (“**Correct!**” or “**Wrong!**”) and response latency (the RT in milliseconds: “**1056**”).

### **2.3.2 Identification**

The identification tasks were performed immediately after the discrimination tasks in each one of the training sessions. They were constructed to provide identification training on the same nonwords participants had previously been trained on through discrimination. A total of 192 identification training trials were included in the training, 48 per session (6 nonword pairs, i.e. 12 trials x 4 voices = 48 trials). Participants performed 4 practice trials, after which they were asked to identify 48 nonwords presented randomly. Each nonword was presented auditorily only as two pictures containing labels for each one of the target vowels in the contrast appeared simultaneously on the left and right side of the screen. The labels included orthographic, phonetic and visual semantic representations (standardized line drawings) of the words *cap*, *cup*, *feet*, *fit*. Participants selected the label corresponding to their selected response option by pressing a designated labelled key on the computer keyboard and received the same type of feedback they had received during the discrimination training.

## **2.4 Pre- and post-tests**

### **2.4.1 ABX Discrimination**

In order to assess learners' L2 perception, a speeded categorical ABX discrimination test was administered. Trials were created by combining (non)words into ABX triads with a 500ms ISI (e.g. A=*fadattick*-B=*faduttick*-X=*fadattick*). Participants were instructed to decide, as accurately and as fast as they could, whether the last (non)word in the triad contained the same stressed vowel as the first (A) or the second word (B) by selecting a key labelled as A or B on the computer keyboard. A, B and X were always produced by a different speaker. A and B were always produced by the speakers who had provided the stimuli for the training (F1, F2, M1, M2). These speakers' voices appeared the same number of times in all A and B positions. The last word in the trial was always produced by two speakers participants had not been exposed to during the training (F3 or M3). Different voices were used within each trial to ensure that participants made a decision based on the phonological categorization of the stimuli while disregarding indexical phonetic variability between nonwords coming from the speakers' voices.

For each contrast participants were presented with 64 experimental trials testing the target contrasts (/æ/-/ʌ/ and /i:/-/ɪ/) and 16 control

trials testing vowel contrasts that were not expected to pose perceptual difficulty to learners (/æ/-/i:/ and /u:/-/ɒ/). The 64 experimental trials per contrast consisted of 8 3-syllable nonword pairs and 8 real word pairs (4 1-syllable and 4 2-syllable words) presented in all 4 possible orders (ABB, ABA, BAA, BAB). Half of the nonwords were trained in the discrimination and identification tasks and half were untrained. None of the real English words had previously appeared in the training. Untrained nonwords were included to test for generalization to new nonword items. Untrained real words were included to test whether phonetic training based on non-lexical items (nonwords) was effective in modifying sensitivity to the same target vowel contrasts in a lexical context and consequently had the power of modifying already existing phono-lexical representations where the target vowel contrasts might have been previously misrepresented or not properly encoded phonologically.

Before doing the test participants performed 8 practice trials during which they received visual feedback for error and response latency, as explained above. The 80 test trials were presented in fully randomized order. If a participant made no response within 2500 milliseconds, the next trial was initiated. The response latencies in milliseconds measured from the onset of the third nonword in the triad were used as a measure of speed. Both the accuracy and speed measures were meant to reflect perceptual sensitivity to the contrasts being tested.

### 2.4.2 Lexical decision

A lexical decision test (Darcy, 2018) was used to obtain a measure of L1-Catalan learners' perceptual sensitivity to the L2-English contrasts /æ/-/ʌ/ and /i:/-/ɪ/ in a lexical context, reflecting the extent to which L2 learners had accurately encoded these phonological contrasts lexically. Participants were asked to decide, as accurately and as fast as possible, whether a sequence of sounds presented auditorily (as spoken by a male native speaker of English) constituted an English word or not. Control trials were distractor sound sequences consisting of 34 monosyllabic and disyllabic English words (*cake, jumping*) and 34 English nonwords (*peef, sagreem*). Test trials consisted of 28 English words containing the target test vowels (*map, sun, clean, gift*) and 28 English nonwords created by substituting the target test vowels by their contrasting counterparts (*mup, san, clin, geeft*). Native-like sensitivity to the /æ/-/ʌ/ and /i:/-/ɪ/ contrasts would therefore be reflected in correctly identifying both test

words and nonwords. We calculated, for every participant and testing time (pre-test and post-test), average accuracy rate and RT scores per contrast (including both words and nonwords) separately for test and control trials, as well as two individual measures of perceptual sensitivity to the contrast based on accuracy rates for words and nonwords. The first one is a *d-prime* ( $d'$ ) score,  $d'=(z(H))-z(FA)$ , where H (hit rate) is the proportion of test words correctly identified as words and FA (false alarm rate) is the proportion of test nonwords incorrectly identified as words. The second is an adjusted accuracy measure called *delta* ( $\delta$ ), which we computed as the average difference between performance on control trials and test trials (test-control accuracy rates).

## **2.5 Cognitive attention control tasks**

### **2.5.1 Auditory selective attention**

The English learners performed auditory selective attention tasks in their L1 (Catalan) and in their L2 (English) based on single-talker competition (Humes, Lee, & Coughlin, 2006). Each task consisted of 64 trials of pairs of sentences (target vs. competitor). The two sentences in a pair were always different, one spoken by a male voice and one by a female voice, and presented simultaneously (e.g. male: *Ready CHARLIE go to BLUE SIX now*; female: *Ready TIGER go to RED EIGHT now*). All of the Catalan and English sentences were normalized for duration to 1700ms. In every trial, a call signal (e.g. TIGER) appearing on the screen previous to the auditory presentation of the sentence, cued the voice participants had to attend to for correctly identifying 1 of 4 colours and 1 of 8 digits visually presented on the screen. Individual ASA scores were computed by adding up all correctly identified colours and digits in each one of the two tasks up to a maximum score of 128.

### **2.5.2 Auditory attention switching**

A measure of L1 attention switching skill (RT and accuracy switching costs) was obtained through a task that required participants to attend to either the duration (quantity) or the voice (quality) of L1 (Catalan) vowels presented in isolation (Safronova, 2016; Safronova & Mora, 2013). This task was designed as an auditory version of Segalowitz & Frenkiel-Fishman's (2005) linguistic version of the task-switching paradigm (Monsell, 2003) and aimed at providing a measure of attentional flexibility for speech dimensions. Participants were required to shift focus of attention from

segmental duration (long vs. short) to voice quality (female vs. male) in the perception of vowel sounds. Several tokens of the Catalan vowels /i e ε a ɔ o u/ produced by a male and a female speaker on a falling pitch were manipulated using the PSOLA algorithm in Praat (Boersma & Weenink, 2009) to create long (500ms) and short (200ms) versions of the 7 vowels ( $7 \times 2 \times 2 = 28$  stimuli). Eight identical copies of each stimulus ( $28 \times 8 = 224$  trials) were randomly presented to participants over headphones for categorization (*long, short, female, male*) after three separate practice blocks (long vs. short duration; female vs. male voice; duration + voice in alternating runs). Participants used designated labelled keyboard keys to categorize a vowel sound as *long* or *short* when a speaker icon appeared in any of the two top boxes of a framework of 4 square boxes and to label a vowel sound as *female* or *male* when a speaker icon appeared in any of the two bottom boxes of the framework. Speaker icons appeared predictably in clockwise fashion around the framework at the onset of the auditory stimuli. Trials alternated predictably between duration (D) and voice quality dimensions (V) creating a sequence of *repeat* (same dimension as preceding stimulus) and *switch* (different dimension from preceding trial) trials. Participants were expected to obtain lower accuracy and speed scores on switch than on repeat trials due to the cost associated with having to refocus attention on a different acoustic dimension. The switching cost (the difference between switch and repeat RTs) was used as a measure of attention control, so that the smaller the switching costs, the stronger the attention control.

### 2.5.3 Auditory inhibition

An auditory inhibition task based on Filippi, Leech, Thomas, Green, & Dick (2012) and Filippi, Karaminis, & Thomas (2014) was used to obtain a measure of auditory inhibition. Participants were presented with 72 pairs of sentences binaurally over headphones, one was always produced by a male voice (e.g. *the dog is chasing the cat*) and the other one by a female voice (e.g. *the dog is chased by the cat*). The sentences, which could be in English or in Catalan, were produced by 4 speakers, 1 male and 1 female native speaker of each language. They were recorded in a sound-proof booth and normalized for amplitude and duration (2000ms). The 72 trials were presented in two 36-trial blocks. In block 1, participants were instructed to attend to the female voice only, and in block 2 to attend to the male voice. Blocks 1 and 2 were counterbalanced across participants. The participants' task was to decide which of two animals in the sentence that

was being attended to (*bird, bull, cat, cow, dog, frog, goat, horse, parrot, seal, snake, wolf*) did the action (*bite, chase, eat, grab, scare, scratch*) by selecting one of two response keys corresponding to one of the two animal pictures appearing on the screen. Twenty-four trials consisted of pairs of L1-L1 (12) or L2-L2 (12) sentences whereas 48 trials consisted of L1-L2 sentence pairs. In the L1-L2 trials the voice that had to be attended to (*target*) was always a voice speaking in English, so that participants were forced to inhibit the sentence in their L1 (competitor) in order to correctly identify the animal doing the action (e.g. target: *the dog is chased by the cat* vs. competitor: *el gos persegueix el gat*). Half of the target English sentences in the L1-L2 trials were produced by a male voice and half by a female voice, half were in the active and half in the passive, and half of the correct responses corresponded to animals appearing on the right side of the screen and half on the left. Trials where the L1 had to be attended to and the L2 inhibited were not included in order to keep the task short. We obtained two measures of auditory inhibition accuracy and RT from this task, overall *general* measures based on all 72 trials in the test, and measures based on L1-L2 trials only (those where the L1 had to be inhibited) to measure L1 inhibition.

## **2.6 L2 proficiency**

Overall L2 proficiency was assessed through an elicited imitation task and a receptive vocabulary size test. The elicited imitation task was originally designed by Ortega, Iwashita, Rabie and Norris (2002) for a cross linguistic study on syntactic complexity measures. It includes 30 test sentences ranging from 7-17 syllables constructed to include high frequency vocabulary items, a range of syntactic complexity, and typical grammatical features known to challenge instructed learners. The sentences were produced by a female native speaker of English and were presented auditorily only over headphones for delayed repetition. Participants were instructed to repeat each sentence as accurately as they could (and as much of the sentence as they could) after a 250ms *beep* signal, which occurred 2000ms after the sentence end. Participants had 6.8 seconds to repeat the sentence after the *beep*. The learners' productions were recorded onto a digital recorder and assessed for accuracy following Ortega *et al's* (2002) rubric, where each sentence received a score from 0 to 4 as a function of how much of it was repeated and the type of inaccuracies and missing unrepeated material. Individual scores could therefore range 0-120 points.

### 3. Results

We first present the results of the pre- and post-tests for perception (ABX discrimination and Lexical Decision) and then those of the cognitive attention control tasks. When response latencies (RTs) are reported, these correspond to RTs screened for accuracy (only including correct responses) and extreme values (2.5 standard deviations below or above each subject's mean).

#### 3.1 Perceptual learning

The results of the ABX discrimination tests showed robust improvement from pre-test to post-test for the two test vowel contrasts, both in response accuracy and speed (see Table 2 for overall results and Table 3 for results by word type). A series of ANOVAs with *Trial Type* (Test, Control) and *Testing Time* (T1=pre-test, T2=post-test) as within-subjects factors revealed, for accuracy, significant main effects of *Trial Type* (/æ/-/ʌ/:  $F(1, 16)=298.14$ ,  $p<.001$ ,  $\eta^2=.949$ ; /i:/-/ɪ/:  $F(1, 16)=90.93$ ,  $p<.001$ ,  $\eta^2=.850$ ) and *Testing Time* (/æ/-/ʌ/:  $F(1, 16)=4.56$ ,  $p=.048$ ,  $\eta^2=.222$ ; /i:/-/ɪ/:  $F(1, 16)=11.89$ ,  $p=.003$ ,  $\eta^2=.426$ ) for both vowel contrasts, suggesting that control contrasts, as expected, were significantly easier to discriminate than test contrasts, and that correct discrimination rates improved from pre- to post-test. The *Trial Type* x *Testing Time* interaction, however, was significant (/æ/-/ʌ/:  $F(1, 16)=5.79$ ,  $p=.028$ ,  $\eta^2=.266$ ; /i:/-/ɪ/:  $F(1, 16)=9.46$ ,  $p=.007$ ,  $\eta^2=.372$ ), as gains from pre- to post-test did not reach significance for control trials (/æ/-/ʌ/:  $t(16)=-2.10$ ,  $p=.837$ ; /i:/-/ɪ/:  $t(16)=-2.10$ ,  $p=.837$ ). A similar pattern of results was obtained for response speed, with significant main effects of *Trial Type* (/æ/-/ʌ/:  $F(1, 16)=298.14$ ,  $p<.001$ ,  $\eta^2=.949$ ; /i:/-/ɪ/:  $F(1, 16)=90.93$ ,  $p<.001$ ,  $\eta^2=.850$ ) and *Testing Time* (/æ/-/ʌ/:  $F(1, 16)=4.56$ ,  $p=.048$ ,  $\eta^2=.222$ ; /i:/-/ɪ/:  $F(1, 16)=11.89$ ,  $p=.003$ ,  $\eta^2=.426$ ) for both vowel contrasts, suggesting that control contrasts, as expected, could be discriminated faster than test contrasts, and participants were significantly faster at doing so at post-test than at pre-test. Again, a significant *Trial Type* x *Testing Time* interaction arose, as gains in speed were much smaller for control than for test items.

<i>Trial Type</i>	<i>Contrast</i>	<i>Test</i>	<i>Accuracy</i>				<i>RT</i>			
			<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>
<i>Test</i>	/æ/-/ʌ/	T1	.644	.066	.53	.75	966	153	657	1160
		T2	.730	.077	.56	.91	799	172	597	1284
	/i:/-/ɪ/	T1	.605	.112	.42	.81	1002	159	724	1252
		T2	.732	.114	.55	.95	832	173	635	1278
<i>Control</i>	/æ/-/ʌ/	T1	.915	.112	.56	1.00	899	137	650	1114
		T2	.922	.084	.69	1.00	780	172	529	1262
	/i:/-/ɪ/	T1	.911	.143	.44	1.00	885	162	589	1109
		T2	.963	.058	.81	1.00	729	179	540	1228

Table 2. Mean accuracy (proportion of correct responses) and response latencies (RT) in the ABX discrimination test at pre-test (T1) and post-test (T2) by trial type and vowel contrast.

In order to assess whether these general learning outcomes were generalizable to untrained test nonwords and words, we examined trainees' performance at pre-test and post-test for untrained test nonwords and words. As shown in Table 3 below, gains in accuracy and speed were consistent across all item types. We submitted the accuracy and RT scores to a series of ANOVAs with *Testing Time* (T1=pre-test, T2=post-test) and *Word Type* (nonword, word) as within-subjects factors. These analyses revealed significant main effects of *Testing Time* (/æ/-/ʌ/:  $F(1, 16)=43.72, p<.001, \eta^2=.732$ ; /i:/-/ɪ/:  $F(1, 16)=29.31, p<.001, \eta^2=.647$ ) and *Word Type* (/æ/-/ʌ/:  $F(1, 16)=15.21, p=.001, \eta^2=.487$ ; /i:/-/ɪ/:  $F(1, 16)=11.89, p=.003, \eta^2=.426$ ) on accuracy, and significant main effects of *Testing Time* (/æ/-/ʌ/:  $F(1, 16)=175.52, p<.001, \eta^2=.911$ ; /i:/-/ɪ/:  $F(1, 16)=15.73, p=.001, \eta^2=.496$ ) and *Word Type* (/æ/-/ʌ/:  $F(1, 16)=15.73, p=.001, \eta^2=.496$ ; /i:/-/ɪ/:  $F(1, 16)=23.57, p<.001, \eta^2=.596$ ) on speed. None of the interactions reached significance. This showed that participants were more accurate and faster at discriminating the target vowel contrasts in untrained real English words than in untrained nonwords and that they improved significantly from pre-test to post-test both in discrimination accuracy and speed, confirming the effectiveness of the treatment.



Word Type	Contrast	Test	Accuracy				RT			
			M	SD	Min.	Max.	M	SD	Min.	Max.
Trained nonwords	/æ/-/ʌ/	T1	.591	.130	.25	.81	1094	189	753	1363
		T2	.720	.103	.50	.88	861	154	669	1294
	/i:/-/ɪ/	T1	.536	.166	.19	.81	1120	185	798	1431
		T2	.702	.176	.38	.94	922	174	707	1287
Untrained nonwords	/æ/-/ʌ/	T1	.518	.078	.38	.69	1060	176	657	1338
		T2	.577	.140	.25	.81	897	210	663	1414
	/i:/-/ɪ/	T1	.562	.134	.25	.81	1079	185	775	1370
		T2	.683	.153	.44	1.00	922	206	689	1459
Untrained words	/æ/-/ʌ/	T1	.733	.107	.53	.88	884	140	605	1071
		T2	.812	.115	.63	.97	736	175	528	1211
	/i:/-/ɪ/	T1	.661	.113	.47	.88	930	146	665	1119
		T2	.772	.096	.63	.94	753	169	557	1214

Table 3. Mean accuracy (proportion of correct responses) and response latencies (RT) in the ABX discrimination test at pre-test (T1) and post-test (T2) by word type and vowel contrast.

Given the consistency of the overall improvement in discrimination accuracy and speed of the target contrasts for both words and nonwords we computed accuracy and speed gain scores based on all test trials in the ABX discrimination task (Table 4) to be able to relate individual differences in attention control to individual gains in discrimination accuracy and speed.

Contrast	Accuracy				RT			
	M	SD	Min.	Max.	M	SD	Min.	Max.
/æ/-/ʌ/	.086	.067	-.05	.17	-167	143	-369	170
/i:/-/ɪ/	.126	.103	-.08	.36	-170	116	-391	25

Table 4. Mean accuracy (proportion of correct responses) and response latency (RT) gains in the ABX discrimination test by vowel contrast.

L2 learners' performance on the lexical decision task showed that, as expected, test words (*map*) were identified correctly at much higher accuracy rates (79-85%) than test nonwords (*mup*; 39-50%), whereas control words (86%) and nonwords (76%) were identified at similar accuracy rates. Similarly, test words were identified faster (1260-1304ms) than test nonwords (1387-1453 ms). Large differences between control and test items were obtained for test nonwords (76% vs. 39-50%, respectively), whereas for words differences between control and test items were very small (86% vs. 79-85%). Improvement in accuracy and speed between pre-test and post-test, however, was relatively small and only observable for test nonwords (5% for /æ/-/ʌ/ and 4.2% for /i:/-/ɪ/). The measures of perceptual sensitivity to the contrasts obtained through this task ( $d'$  and  $\delta$ ) showed a similar pattern of results (Table 5).

Contrast	Measure	$d$ -prime ( $d'$ )				delta ( $\delta$ )			
/æ/-/ʌ/	T1	.97	.81	-.57	2.91	.16	.08	.00	.33
	T2	.93	.47	.18	2.03	.18	.09	-.01	.33
	Gain	.016	.08	-.14	.14	-.007	.09	-.17	.17
/i:/-/ɪ/	T1	.91	.84	-.40	3.27	.16	.07	.02	.30
	T2	1.10	.91	.00	3.27	.18	.08	.00	.28
	Gain	.010	.10	-.25	.21	-.005	.10	-.12	.29

Table 5. Mean  $d$ -prime ( $d'$ ) and delta ( $\delta$ ) pre-test and post-test scores and gains by vowel contrast.

### 3.2 Cognitive attention control

Participants obtained slightly higher accuracy scores in the Catalan version of the auditory selective attention task (*AudSelAtt*) than they did in the English version (Table 6). This difference did not reach significance ( $t(16) = .968$ ,  $p = .348$ ), but both scores were only moderately correlated ( $r = .442$ ,  $p = .075$ ), suggesting that individual differences in auditory selective attention were not consistent across the two tasks within participants.

In the auditory attention switching task (*AudAttSw*), as expected, participants were less accurate and slower at identifying the duration (*long* or *short*) and voice quality (*male* or *female*) in the vowels on switch trials (86%, 865ms) than on repeat trials (90%, 726ms). The overall error rate

was low (10-13%), suggesting that these perceptual dimensions posed no difficulty to listeners (Table 6). Because RTs were measured from stimulus onset, participants took longer to respond to a duration trials than to voice trials, as whereas voice quality could be immediately identified from the beginning of the stimulus, the decision on duration required participants to wait for the duration of a short vowel (200ms). Consequently, we used an adjusted RT measure obtained by subtracting 200ms from the original RTs. We submitted the accuracy and adjusted RT scores to a series of ANOVAs with *Dimension* (duration, voice) and *Trial Type* (switch, repeat) as within-subjects factors. These analyses yielded a significant main effect of *Trial Type* ( $F(1, 16)=12.31, p=.003, \eta^2=.435$ ) and a non-significant main effect of *Dimension* ( $F(1, 16)=.317, p=.581, \eta^2=.019$ ) on accuracy, suggesting that participants were equally accurate on both dimensions but made significantly more errors on switch than on repeat trials. For response speed (RTs), the ANOVA revealed significant main effects of both *Dimension* ( $F(1, 16)=20.72, p<.001, \eta^2=.564$ ) and *Trial Type* ( $F(1, 16)=45.65, p<.001, \eta^2=.741$ ), because participants were slower at deciding on the duration of a vowel than on whether it was produced by a male or a female speaker. None of the interactions reached significance. We used the switch cost measure as an index of attention switching skill.

In the auditory inhibition task (*AudInh*), the results showed that target sentences were processed slightly less accurately in L1-L1 sentence pairs (69-77%) than in L2-L2 (78-86%) or L2-L1 sentence pairs (80-89%), especially when the voice of the competing sentence was male. In L2-L2 and L2-L1 sentence pairs accuracy was lower when the voice of the competing sentence was female (78-80%) than when it was a male voice (86-89%), indicating that, when the L2 is attended to, a female voice is harder to inhibit than a male voice (irrespective of whether the female voice is speaking in the participants' L1 or L2). We submitted the aggregated scores for accuracy (proportion of correct responses) and RTs to a series of ANOVAs with *Language* (L2-L1, L2-L2, L1-L1) and *Target Voice* (Male, Female) as within-subject factors. The results of these analyses showed, for accuracy, a significant main effect of *Language* ( $F(2, 15)=6.29, p=.010, \eta^2=.456$ ), a non-significant main effect of *Target Voice* ( $F(1, 16)=0.38, p=.546, \eta^2=.023$ ) and a significant *Language x Target Voice* interaction ( $F(2, 15)=6.29, p=.044, \eta^2=.340$ ). The interaction arose because whereas for L2-L1 sentence-pair trials with competitor sentences spoken by a female speaker obtained lower accuracy rates than those spoken by a male speaker ( $t(16)=-2.56, p=.021$ ), such a difference did not reach significance

for L2-L2 ( $t(16)=-1.07$ ,  $p=.299$ ) and L1-L1 sentence pairs ( $t(16)=1.03$ ,  $p=.316$ ). Also, whereas *Language* had a significant main effect on response accuracy when attending to a female voice ( $F(2, 15)=8.92$ ,  $p=.003$ ,  $\eta^2=.543$ ), this effect did not reach significance when attending to a male voice ( $F(2, 15)=.440$ ,  $p=.652$ ,  $\eta^2=.055$ ). For response latencies, however, the ANOVA yielded significant main effects of *Language* ( $F(2, 15)=18.24$ ,  $p=.001$ ,  $\eta^2=.709$ ) and *Target Voice* ( $F(1, 16)=7.36$ ,  $p=.015$ ,  $\eta^2=.315$ ), and a non-significant *Language* x *Target Voice* interaction ( $F(2, 15)=1.41$ ,  $p=.274$ ,  $\eta^2=.159$ ). We calculated *general* accuracy and RT inhibition scores across all language combinations and voices and a more specific score based on L2 learners' performance on L2-L1 trials only (Table 6), that is, trials where the L2 had to be attended to and the L1 had to be inhibited (*L1 inhibition*).

<i>Task</i>	<i>Conditions</i>	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>
<i>AudSelAtt</i>	Catalan (L1)	101.18	10.90	86	118
	English (L2)	98.47	10.932	79	114
<i>AudAttSw</i>	Switch	865	224	485	1338
	Repeat	726	221	445	1356
	Switch Cost	139	85	-17	283
<i>AudInh</i>	General (accuracy)	.803	.100	.54	.95
	L1 inhibition (accuracy)	.850	.102	.54	.94
	General (RT)	2392	386	1786	2989
	L1 inhibition (RT)	2338	314	1829	2861

Table 6. Mean scores in the attention control tasks: *AudSelAtt* (accuracy score 0-128), *AudAttSw* (adjusted RT in milliseconds) and *AudInh* (proportion of correct responses and RT in milliseconds)

### 3.3 Relationship between cognitive attention control and perceptual learning

Perception scores at pre-test, as expected, were related to the overall proficiency measure. L2 learners with higher scores in the elicited imitation task were better able to discriminate the target vowels /æ/-/ʌ/ ( $r=.434$ ,  $p=.082$ ) and /i:/-/ɪ/ ( $r=.590$ ,  $p=.013$ ) and also showed higher sensitivity to these contrasts ( $d'$  scores) in the lexical decision task (/æ/-/ʌ/:  $r=.597$ ,  $p=.011$ ; /i:/-/ɪ/:  $r=.551$ ,  $p=.022$ ), but proficiency was unrelated to gain scores.

Before assessing the contribution of cognitive attention control skills to L2 speech learning we explored the relationship between the various attention control measures (*AudSelAtt*, *AudAttSw*, *AudInh*). These analyses revealed an association between learners' auditory selective attention and auditory inhibition skills (Table 7), suggesting that the stronger their ability to focus their attention on a target voice in the presence of a competing voice in the *AudSelAtt* task, the better they could inhibit a competing voice in their first and second language in the *AudInh* task (Table 7). Thus, both these tasks appear to require participants to resort to the same underlying attentional resources. Interestingly, learners' switching costs in the *AudAttSw* task were strongly related to their ability to inhibit a voice in the L1 while attending to a voice speaking in the L2 (L1 inhibition), suggesting that learners with better auditory attentional flexibility (i.e. attention switching skills) were better able to inhibit their L1 when attending to the L2.

		<i>AudAttSw</i>		<i>AudInh</i>							
		RT		Accuracy		RT					
		Switch Cost		General		L1 Inhibition		General L1 Inhibition			
		<i>r</i> =	<i>p</i> =	<i>r</i> =	<i>p</i> =	<i>r</i> =	<i>p</i> =	<i>r</i> =	<i>p</i> =		
<i>AudSelAtt</i>	Catalan	-	.115	.578	.015	.649	.005	-	.543	.069	.794
		.396						.159			
	English	-	.308	.670	.003	.475	.054	-	.015	-	.033
		.263						.576		.520	
<i>AudAttSw</i>	Switch Cost			-	.102	-.627	.007	.311	.224	.252	.328
	Switch Repeat			.410				.807	<.001	.842	<.001
								.698	.002	.757	<.001

Table 7. Pearson-r correlation coefficients between the attention control measures (shaded cells indicate significance).

Both ABX discrimination accuracy and RT gains of the two target contrasts were correlated with one another ( $r=.549$ ,  $p=.022$  and  $r=.723$ ,  $p=.001$ , respectively), as they were in the lexical decision task ( $r=.541$ ,  $p=.025$ ), indicating that individual gain sizes were of similar magnitude for the /æ/-/ʌ/ and /i:/-/ɪ/ contrasts. We next assessed the relationship between the attention control measures and L2 learners' perception scores. We ran these analyses both for T1 perception scores and T1-T2 gains. For ABX

discrimination accuracy, the results revealed significant moderately strong correlations between learners' perception gains and auditory selective attention, reaching significance for the /æ/-/ʌ/ contrast ( $r=.522, p=.031$ ) and approaching significance for the /i:/-/ɪ/ contrast ( $r=.441, p=.076$ ). This suggests that auditory selective attention predicts a considerable amount of variance (about 27%) in how much learners could benefit from the training. No significant associations were found between ABX discrimination gains and attention switching (*AudAttSw*) or inhibition (*AudInh*) scores. However, pre-test ABX discrimination accuracy scores were significantly related to the auditory attention switching measure ( $r=-.510, p=.037$ ) and the L1 inhibition measure ( $r=.520, p=.032$ ) for the /æ/-/ʌ/ contrast, suggesting that at pre-test both attention switching skill and L1 inhibition skills predicted a significant amount of variance (about 25%) in the learners' ability to discriminate the /æ/-/ʌ/ contrast. In addition, the RT L1 inhibition measure was strongly related to pre-test RT scores for both the /æ/-/ʌ/ ( $r=.619, p=.008$ ) and /i:/-/ɪ/ ( $r=.701, p=.002$ ) contrasts, explaining more than 40% of the variance in the discrimination response speed at pre-test. Finally, the relationship between the attention switching cost measure and the  $d'$  gain scores (gains in perceptual sensitivity) for the /æ/-/ʌ/ contrast approached significance ( $r=-.476, p=.063$ ), and reached significance in the case of the adjusted  $\delta$  accuracy gain measure in the lexical decision task ( $r=-.594, p=.015$ ), suggesting that attention switching skill may be implicated in effecting changes in the lexical encoding of phonological contrasts. It should be noted, however, that perception gains between pre-test and post-test measured through the lexical decision task did not reach significance.

#### **4. Discussion and conclusion**

The main aim of the present study was to explore the contribution of cognitive attention control to L2 phonological development. We tested a group of L1-Catalan learners of L2 English on their attention control skills and trained them on the perception and production of two difficult L2 vowel contrasts (/æ/-/ʌ/ and /i:/-/ɪ/) through minimal-pair nonwords. We then assessed their gains and related them to the attention control measures.

The results revealed robust improvement from pre-test to post-test for both contrasts in response accuracy and speed, as well as consistent generalization effects to untrained nonwords and words. A major finding regarding phonetic training gains is that training based exclusively on minimal-pair nonwords, and therefore void of lexical content, led to

improvement in the perception of minimal-pair words participants had not been trained on, suggesting that improvement in perceptual sensitivity to phonetic contrasts at the phonetic perceptual level may effect changes and lead to improvement in corresponding phono-lexical representations exploiting the same contrasts. However, the lexical decision task, which provided a measure of sensitivity to the /æ/-/ʌ/ and /i:/-/ɪ/ contrasts encoded lexically, only revealed little (and non-significant) improvement in sensitivity to the contrasts in nonwords. These apparently contradictory findings may result from the nature of the lexical decision task and the stage of development of the learners' L2 phonology. In a lexical decision task improvement in performance is based on the participants' ability to identify nonwords based on the phonological distinction between two members of a contrast in a lexical context, a task that required our learners to have accurately encoded the /æ/-/ʌ/ and /i:/-/ɪ/ contrasts lexically in their phonologies. Further research is needed to explore the efficiency of phonetic training in developing or changing phono-lexical representations. In particular, it would be interesting to carry out a follow-up training study based on words (rather than nonwords) and assess generalization to new lexical items through a lexical decision task.

The relationship between the cognitive attention control measures revealed an association between learners' performance on the auditory selective attention and the auditory inhibition tasks. Although the former did not require test-takers to inhibit a language through attention to voice, both tasks were based on voice competition and apparently required the recruitment of similar attentional resources. Similarly, participants' attention switching skills were related to their ability to inhibit a voice in the L1 when attending to L2 speech, suggesting that attention switching is implicated in L2 speech processing.

As regards the relationship between the cognitive attention control and the L2 vowel perception measures, a moderately strong correlation between L2 gains in the perception of the /æ/-/ʌ/ contrast and auditory selective attention suggests that learners' ability to focus their attention to specific speech dimensions is related to L2 phonological acquisition, confirming previous findings (Darcy et al., 2014; Safronova, 2016). Stronger associations between attention control and gains in L2 perception could have surfaced for tendencies identified in the current study with a slightly larger sample size.

The present study has contributed to research on individual differences in L2 speech learning suggesting that cognitive attention control plays an important role in L2 speech learning. The fact that attention control explains a substantial amount of variance in L2 vowel perception has important implications for L2 pronunciation instruction beyond phonetic training. In particular, cognitive attention control is likely to play an important role in the context of communicative language teaching where recent research (Gurzynski-Weiss, Long, & Solon, 2017) has shown that meaning-oriented tasks with a focus on phonetic form making L2 pronunciation essential for task resolution is effective in developing L2 speech perception and production.

### **Acknowledgments**

We would like to thank Isabelle Darcy for kindly sharing the lexical decision task with us, Natalia Wisniewska for helping recruit participants and collect data, and Zhifei Zhang for help in data collection. This study was supported by grant 2017SGR560 from the Catalan government.

### **References**

- Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, *47*, S53-S71.
- Astheimer, L. B., & Sanders, L. D. (2009). Listeners modulate temporally selective attention during natural speech processing. *Biological Psychology*, *80*(1), 23-34.
- Astheimer, L. B., Berkes, M., & Bialystok, E. (2016). Differential allocation of attention during speech perception in monolingual and bilingual listeners. *Language, Cognition and Neuroscience*, *31*(2), 196-205.
- Baese-Berk, M., Bent, T., Borrie, S., & McKee, M. (2015). Individual Differences in Perception of Unfamiliar Speech. In *Proceedings of the 18th International Congress of the Phonetic Sciences*.
- Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240-250.
- Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.40, retrieved from <http://www.praat.org/>



- Bohn, O.-S. (1995). Cross-language perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379-410). Timonium, MD: York Press.
- Darcy, I. (2018). Executive functions and L2 phonological development: a study in multiplicity. Paper presented at the *SimPhon.Net Workshop*, 4-6 June 2018, Stuttgart, Germany.
- Darcy, I., Mora, J. C., & Daidone, D. (2014). Attention control and inhibition influence phonological development in a second language. In *Proceedings of the International Symposium on the Acquisition of Second Language Speech*. Concordia Working Papers in Applied Linguistics, 5, 115-129.
- Filippi, R., Karaminis, T., & Thomas, M.S. (2014). Language switching in bilingual production: Empirical data and computational modelling. *Bilingualism: Language and Cognition*, 17(2), 294-315.
- Filippi, R., Leech, R., Thomas, M. S., Green, D. W., & Dick, F. (2012). A bilingual advantage in controlling language interference during sentence comprehension. *Bilingualism: Language and Cognition*, 15(4), 858-872.
- Flege, J. E. (1995). Second-language Speech Learning: Theory, Findings, and Problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229-273). Timonium, MD: York Press.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods*, 35(1), 116-124.
- Gurzynski-Weiss, L., Long A. Y., & Solon M. (2017). TBLT and L2 Pronunciation: Do the benefits of Tasks Extend beyond Grammar and Lexis? *Studies in Second Language Acquisition*, 39, 213-224.
- Hazan, V., & Kim, Y. H. (2010). Can we predict who will benefit from computer-based phonetic training? In *Online Proceedings of the INTERSPEECH 2010 Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology (L2WS 2010)*, Tokyo, Japan.
- Humes, L. E., Lee, J. H., & Coughlin, M.P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *The Journal of the Acoustical Society of America*, 120(5), 2926-2937.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7(3), 134-140.
- Mora, J. C. & Darcy, I. (2016). The relationship between cognitive control and pronunciation in a second language. In T. Isaacs & P. Trofimovich (Ed.), *Second Language Pronunciation Assessment: Interdisciplinary Perspectives* (pp. 95-120). Bristol, UK: Multilingual Matters.
- Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002). An investigation of elicited imitation tasks in crosslinguistic SLA research. In *Second Language Research Forum*, Toronto.
- Ou, J., Law, S. P., & Fung, R. (2015). Relationship between individual differences in speech processing and cognitive functions. *Psychonomic Bulletin & Review*, 22(6), 1725-1732.

- Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, 35, 73-89.
- Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). The test of everyday attention (TEA). *Bury St Edmunds: Thames Valley Test Company*.
- Safronova, E. & Mora, J. C. (2013). Attention control in L2 phonological acquisition. In A. Llanes Baró, L. Astrid Ciro, L. Gallego Balsà and R. M. Mateus Serra (Ed.), *Applied Linguistics in the Age of Globalization* (pp. 384-390). Lleida: Edicions de la Universitat de Lleida.
- Safronova, E. (2016) *The Role of Cognitive Ability in the Acquisition of Second Language Perceptual Competence*. Unpublished PhD Thesis. Universitat de Barcelona.
- Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. London, UK: Routledge.
- Segalowitz, N., & Frenkiel-Fishman, S. (2005) Attention control and ability level in a complex cognitive skill: attention-shifting and second language proficiency. *Memory and Cognition*, 33, 644-653.
- Wager, T. D., Jonides, J., & Smith, E. E. (2006). Individual differences in multiple types of shifting attention. *Memory & Cognition*, 34(8), 1730-1743.



## A Non-critical Period for Second-language Learning

James Emil Flege

University of Alabama at Birmingham

### Abstract

Early learners usually enjoy greater success in second-language (L2) learning than Late learners do. This is often interpreted to mean that the capacity for L2 learning diminishes after the close of a critical period. However the seeming limits on Late learners' success in learning an L2 following immigration, even after years of regular L2 use in the host country, may not be the unwanted consequence of normal neurocognitive maturation. It may instead arise from differences in the quantity and quality of input that Early and Late learners typically receive. This hypothesis was supported by the research reviewed in this chapter for both L2 speech learning and some aspects of L2 morphosyntax learning, leading to the proposal that long-term success in L2 learning is determined probabilistically by a *non-critical period* defined by age-related variation in L2 input.

### 1. Introduction

Eric Lenneberg (1967) laid out a nativist account of second-language (L2) acquisition that continues to influence research. He provided convincing evidence that native-language (L1) acquisition has a strong biological component and that, to be completely successful, the L1 must be learned before the close of a *critical period*. Lenneberg then extended his critical period (CP) hypothesis for L1 acquisition to the learning of L2 speech based on a simple observation, namely that a foreign accent (FA) is usually evident in the speech of those who began learning their L2 after puberty (1967, p. 176).

As Lenneberg (1967) showed, normal neurological development tends to follow a fixed schedule across individuals. If the capacity for

---

Anne Mette Nyvad, Michaela Hejrná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 501-541). Dept. of English, School of Communication & Culture, Aarhus University.

learning L2 speech diminishes at a certain point, the effect of a reduced capacity for learning should be evident at roughly the same developmental state – and by extension, roughly the same chronological age of first exposure – for everyone who learns an L2. Although Lenneberg (1967) did not say so explicitly, he probably assumed that immigrants to a predominantly L2 speaking country receive abundant input from native speakers who provide a correct model of how the L2 should be pronounced. If that were true, then the observation of detectable foreign accents in immigrants who arrived in the host country after puberty might reasonably be interpreted as evidence for a diminished capacity for L2 speech learning.

This chapter will examine FA research in the light of the CP hypothesis. As a starting point, I question the tacit assumption that all immigrants receive abundant and adequate native-speaker input (see also Moyers, 2009). In fact, we know relatively little about the input that immigrants receive. This is because, as noted by Piske and Young-Scholten (2008, pp. 12-13), “only recently [has] input begun to receive consideration”. Indeed, these authors acknowledge that we simply “do not know how much input second language learners actually get [nor] how much exposure a learner requires”.

Researchers have tended to overlook input as a potential cause of differences between individual L2 learners, potentially leading to the misattribution of inter-subject differences. For example, a difference between two learners who immigrated to a predominantly L2 speaking country at the same age and lived in the host country for the same length of time might be ascribed to *individual differences* of unknown etiology or to differences in language learning aptitude. However if more were known about the quantity and quality of L2 input the two hypothetical individuals had received the difference between them might simply be a manifestation of differences in the input that they had received.

Few researchers would agree that an understanding of how input varies across individuals and groups is crucial to an understanding of age-related differences in L2 learning. This is due, at least in part, to the mistaken view that L2 input can be adequately assessed by length of residence (LOR) in a predominantly L2-speaking country. LOR is often used in L2 research because it can be readily obtained from language background questionnaires- It is, unfortunately, an imprecise measure and sometimes misleading index of the quantity of L2 input immigrants have received. This is because not all immigrants begin using their L2 immediately (e.g., Flege, Munro & MacKay, 1995a, Table I) nor use their L2 on a regular basis (Moyer, 2009, p. 162). The results of Flege and Liu (2001) suggested

that LOR provides a valid index of quantity of L2 input only for immigrants who have had both the opportunity and the need to use their L2 on a regular basis. These two crucial conditions for success in L2 learning are more likely to exist for Early learners than for Late learners (Moyers, 2009, pp. 360-363).

Another problem is that the LOR variable provides no information regarding quality of L2 input. Immigrants can hardly avoid trying to use their L2 when speaking to monolingual speakers of the target L2 after arriving in the host country. It may be just as difficult for them to avoid using the L2 in *linguistically mixed company*, that is, in conversations involving a monolingual L2 native speaker and one or more fellow immigrants from the same L1 background. In such situations the foreign-accented L2 speech to which immigrants are exposed is likely to provide an incorrect model of the target L2, one that reinforces the learners' natural tendency to adapt L2 speech to their existing L1 phonetic/phonological system.

This chapter is organized as follows. Section 2 considers speech learning, first through an examination of FA and then segmental production and perception accuracy. The focus of Section 3 is the learning of L2 morphosyntax. Section 4 directly compares the learning of L2 speech and morphosyntax. Finally, based on the preceding synthesis, Section 5 proposes that a non-critical period exists for L2 speech and morphosyntax learning and that Early vs Late differences derive primarily from age-related differences in the quantity and quality of input received.

## 2. L2 speech learning

### 2.1 AOA conditions input

Flege, Munro and MacKay (1995a) examined English sentences spoken by 240 Italian adults who immigrated to Canada between the ages of 2-23 years and had lived in Ottawa, ON for decades. The Italians recruited for this study had not received formal classroom instruction in English before im-

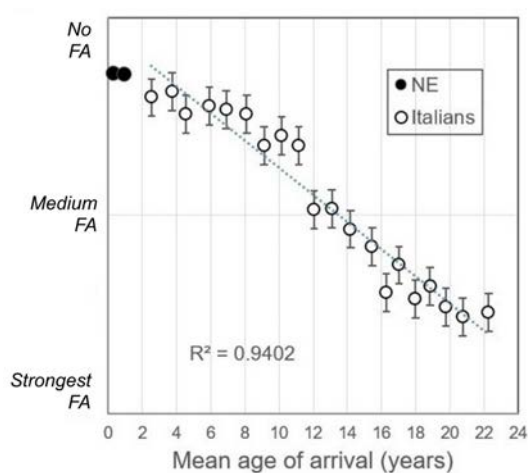


Figure 1. Mean FA ratings obtained for 20 groups of Italians and two groups of native English speakers (+/- 1 Sem)

migrating to Canada, and all continued to speak Italian, especially at home, with close friends, and at church-related events. A delayed repetition task was used to elicit the sentences, which were later presented along with sentences spoken by native English (NE) speakers to listeners who were native speakers of Canadian English. The listeners rated the randomly presented sentences for overall degree of perceived FA using a continuous scale ranging from 1 (*strongest foreign accent*) to 256 (*no foreign accent*).<sup>1</sup>

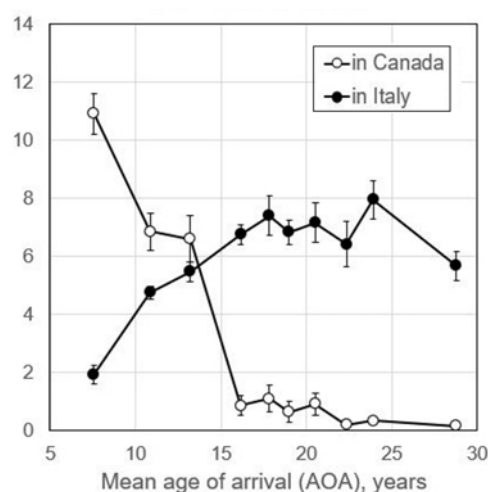


Figure 2. Mean years of formal education for 10 groups of 19 each.

Fig. 1 shows the mean FA ratings obtained for 20 groups of 12 Italians each differing in age of arrival (AOA) in Canada. The groups' mean AOA values accounted for 94% of the variance in the mean FA ratings, and a small increase in strength of FA is evident at an AOA of 11.7 years. Many would consider these findings to be convincing proof that a CP exists for L2 speech learning and that the capacity for L2 learning diminishes after about the age of 12 years as the result of normal neurological maturation.

Both conclusions may be unwarranted. First, as far as I know, age of exposure to an L2 has never been directly linked to state of neurological development (see Flege, 1987; Hartsorne, Tenenbaum & Pinker et al., 2018, p. 274). Second, there is another, potentially better way to interpret these

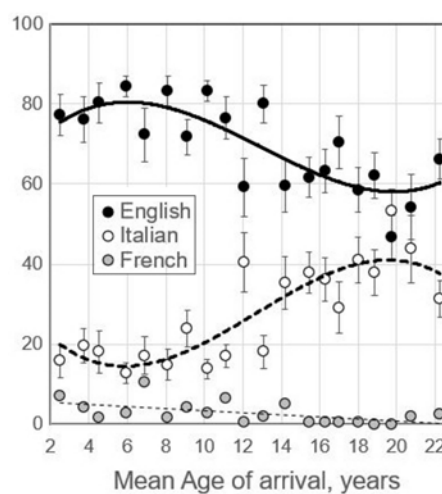


Figure 3. Mean self-estimated language use for 20 groups of 12 Italians each differing in AOA.

<sup>1</sup> The five sentences examined were randomly presented four times each. The first set of ratings were treated as practice and so discarded. The mean rating computed for each participant was thus based on 150 ratings. The 240 Italians were originally assigned to 10 AOA-defined groups of 24 each rather than the 20 AOA-defined groups shown here.

data. The potency of AOA as a predictor of strength of FA may be due to its association with variation in input rather than with state of neurological and/or cognitive maturation at the time of first exposure to an L2. Immigrants' AOA largely conditions their later experience with the L2 (Stevens, 1999, p. 556) and so immigrants' success in learning the target L2 may be the result of differences in the input they receive.

Fig. 2 shows the mean number of years of formal education that AOA-defined groups of Italian immigrants had received both in Italy and later, after arriving in Canada.<sup>2</sup> Those who arrived in Canada before the age of 15 years obtained a substantial amount of formal education in English-speaking Canadian schools whereas most who arrived after that age received little if any formal education in Canada.

A long period of education in Canadian schools was unlikely to have directly affected the phonetic variables of interest here. However it was likely to have impacted the quantity and quality of English language input that the Italian immigrants to Canada later received over the course of their lives (Stevens 1999, p. 563). Specifically, the Italians who were enrolled at a local school soon after arriving in Canada learned English from their NE teachers and their NE classmates, with whom they often developed lifelong friendships and sometimes married. However most of those who arrived in Canada after the age 15, and so did not begin attending school on a full-time basis, lacked an important opportunity for establishing a strong social network in the English-speaking community.

LOR continues to be used in L2 research, but self-estimated L2 use provides a better index of quantity of input, especially if used in combination with LOR. Fig. 3 shows the Italians' mean self-estimates of percentage use

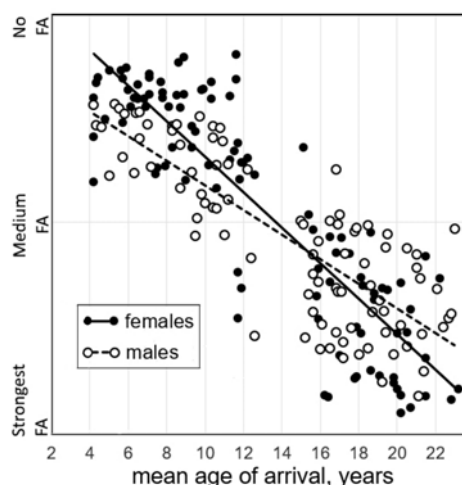


Figure 4. Foreign accent ratings obtained for Italian males and females differing in age of arrival in Canada.

<sup>2</sup> These data were drawn from an unpublished study of aging that examined participants drawn from the same population in Ottawa ON as those tested by Flege et al. (1995a).



of English and Italian. The earlier the Italians arrived in Canada, the more they tended to use English and the less they used Italian.<sup>3</sup> The fact that most (182 or 76%) of the Italians tested by Flege et al. (1995a) reported using English more than Italian is hardly surprising because the Italian-speaking community in Ottawa is relatively small (< 20,000).

Percentage use estimates like these are limited in two ways. They are based on self-report and, perhaps more importantly, they tell us nothing about the quality of L2 input. It is plausible to think that the more frequently the Italians spoke their native language, the more often they were involved in linguistically mixed conversations in which they were exposed to Italian-accented English. The influence of foreign-accented input can be inferred from the gender effects reported by Flege et al. (1995a).

Fig. 4 shows the mean FA ratings obtained for 96 Early learners (AOA = 4.2-12.6) and 96 Late learners (AOA = 15.0-23.2) differing in gender. As expected (e.g., Geary, 1998, p. 263 ff), Early females had a significantly better pronunciation of English than Early males did. The expected female advantage was not evident for Late learners, however. Late females' pronunciation of English was in fact significantly worse than that of Late males ( $p < .05$  by Mann-Whitney  $U$  tests).

This gender effect can be attributed to quality of input. As already mentioned, Italians who arrived early in life learned English at school from NE teachers and classmates. Males who arrived after the age of 15 years typically worked outside the home and learned English from both NE speakers and fellow Italian immigrants (the proportion is unknown). Female Late learners, on the other hand, usually stayed at home in their first few years in Canada. Their first model of English was likely to have been the foreign-accented English spoken by male relatives.

This reconstruction of the Italians' earliest phase of L2 learning was supported by an analysis of rate of learning. A variable called "Time needed to learn English" was derived by subtracting the Italians' age of arrival from their estimates of the age at which they were first able to speak English "comfortably" (see Flege et al., 1995a, Table 1). The times needed by Early males and females to reach this important milestone in L2 acquisition ( $M=11.8$  vs  $10.2$  months) did not differ significantly (Mann-Whitney  $U(1)=.966$ ,  $p > .05$ ). However Late females needed a year longer to speak English comfortably according to self-report than Late males did

<sup>3</sup> The low frequency of French use for some participants who arrived in Canada before the age of 15 reflects the fact that French is taught as a foreign language in English-speaking Canadian schools.

( $M=28.6$  vs  $16.4$  months; Mann-Whitney  $U(1)=-2.94$ ,  $p<.05$ ). The Late males' relatively rapid learning of English was likely the result of an opportunity and need to use English at work, motivations that Late females may have lacked.

This inference was supported by a factor analysis of language background questionnaire data. Flege et al. (1995a) carried out separate Principal Components Analyses of responses obtained for males and females. The factors identified for the two genders were then used as predictors of FA in step-wise multiple regression analyses. Some factors identified for both males and females accounted for a significant amount of variance in the FA ratings. Other factors, however, were unique to one gender. For males but not females, variables designated *Languages used at work*, *Strength of concern for pronunciation*, and *Instrumental motivation* accounted for a significant amount of variance in the FA ratings. For females but not males, factors named *Overall language use* and *Language loyalty* were significant predictors of FA.

## 2.2 Problems for the CP hypothesis

For some researchers, the observation that AOA accounts for more variance than any other variable that can be derived from a language background questionnaire provides strong support for the existence of a CP. However, this form of evidence does not in itself prove that a CP exists nor that the basis for a CP – should one exist – is L2 learners' state of neurological maturation at the time of first exposure to the L2.

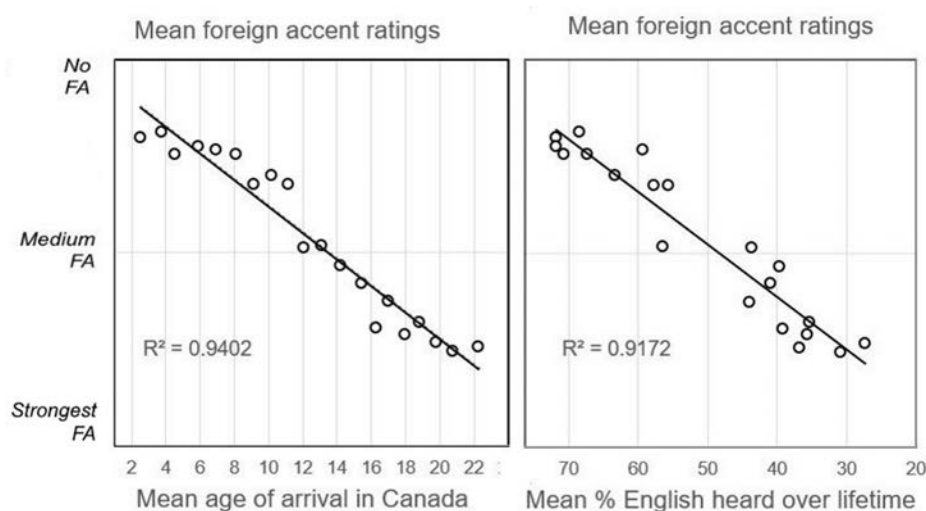


Figure 5. The relation between strength of foreign accent and age of arrival (left) and a measure of input (right) for 20 groups of 12 Italians each.

A substantial amount of variance in FA can also be accounted for without reference to AOA. I derived an input variable called “Percent English Heard in Life” from the Italians’ age at the time of testing, length of residence in Canada, and self-estimated percentage use of English. As seen in Fig. 5, this derived input variable captured nearly as much variance in the FA ratings as AOA did (92% vs 94%).

The CP hypothesis formulated by Lenneberg (1967) predicts the presence of FA for the Italians who arrived after puberty but not the presence of FA in individuals who arrived in Canada as young children. Asher and Garcia (1969) was one of the first studies to show that even children may speak their L2 with a FA. The authors recorded 71 Cuban children living in the San Francisco area. None were judged by NE-speaking listeners to speak English without a FA. However more children who had lived in the US for 5-8 years were judged to have a “near native” pronunciation of English than those who had lived in the US for just 1-4 years (51% vs 15%). This suggested that the Cuban children were making progress over time in the pronunciation of English but left open the question of whether the children would eventually manage to speak English without a detectable FA.

The Italian adults tested by Flege et al. (1995a) had lived in Ottawa far longer than the Cuban children had lived in San Francisco. Some of the Italian Early learners – adults who had begun learning English as young children – obtained FA ratings that were more than 2 SDs below the mean rating obtained for NE speakers. The aim of Flege, Frieda and Nozawa (1997), therefore, was to determine if these Early learners spoke English with a detectable FA.

Flege et al. (1997) re-examined sentences spoken by 40 Early learners drawn from the Flege et al. (1995a) study. These Early learners arrived in Canada at a mean age of 5.8 years and had lived there for an average of 34 years. Most Early learners tested by Flege et al. (1995a) reported using English more than Italian, but seven of the Italian Early learners examined by Flege et al. (1997) reported using Italian more than

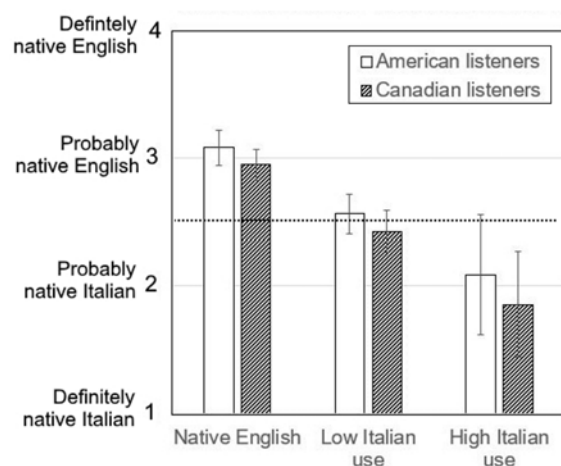


Figure 6. Mean ratings of English sentences spoken by native speakers of English and groups of Early learners differing in language use.

English. The Early learners examined Flege et al. (1997) were assigned to subgroups of 20 each according to percentage use of Italian ( $M=3\%$  vs  $36\%$ ). The two groups did not differ significantly in AOA ( $M=5.6$  vs  $5.9$  years) or the time that elapsed between arrival in Canada and first exposure to English (4 vs 3 months).

NE-speaking listeners from both Canada and the US classified sentences spoken by the Early learners and 20 NE speakers as having been (1) *definitely* spoken by a native Italian (NI) speaker, (2) *probably* spoken by a NI speaker, (3) *probably* spoken by a NE speaker, or (4) *definitely* spoken by a NE speaker. These classifications have been treated in Fig. 6 as a 4-point rating scale. An ANOVA examining the ratings in Fig. 6 yielded a significant main effect of Group, but not a significant main effect of Listener group (Canadians from Ontario vs Americans from Alabama) nor a significant interaction.

A Tukey test revealed that all between-group differences were significant ( $p<.05$ ). This means that both groups of Early learners spoke English with a FA, and that the strength of the Early learners' foreign accents depended on language use. The foreign accents were stronger for the Early learners who used English seldom/Italian often than for the Early learners who used English often/Italian seldom. Neither finding is compatible with the CP hypothesis.

Flege et al. (1997) carried out additional analyses to verify these theoretically important findings. The ratings just considered were based on 216 judgments obtained for each of the 60 participants (12 listeners x 2 listener groups x 3 sentences x 3 replicate judgments). The ratings obtained from the Canadian and American listener groups can be considered replicate experiments. To determine if the results would generalize to other listeners, *Listener-based* analyses were also carried out. In these, a mean rating was obtained for each of the 12 Canadian and 12 American listeners by averaging over the responses obtained for the 20 talkers in each group. The same findings were obtained again.

Significant differences between the two Early learner groups were also obtained using a non-parametric, bias-free measure of sensitivity to the presence of foreign accent (A-prime). The A-prime scores were based on the number of *hits* (i.e., "Definitely" and "Probably native Italian" responses for sentences spoken by the Early learners) and *false alarms* (the same responses given to sentences spoken by NE speakers). Both the Canadian and American listeners were significantly more sensitive to the presence of FA in sentences spoken by Early learners who used English

seldom/Italian often than for the Early learners who used English often/Italian seldom ( $p < .05$ ).

I carried out additional analyses for this chapter to better understand the presence of FA in the Italian Early learners. This involved selecting the 13 (of 40) Early learners who had the best pronunciation of English (the relatively *Good pronouncers*) and the 12 who had the worst pronunciation of English (the *Poor pronouncers*). English sentences spoken by the Poor pronouncers were usually judged to have been “definitely” spoken by a native speaker of Italian whereas the Good pronouncers’ sentences were usually classified as “probably” having been spoken by a NE speaker.

The ratings obtained for the 25 Early learners just described are shown in Fig. 7 in correspondence with the ratings obtained for the same sentences by Flege et al. (1995a). The two sets of ratings were strongly correlated,  $r = .94$ , despite a difference in the scaling procedures used in the two studies (a continuous vs 4-point scale) and even though the ratings obtained by Flege et al. (1997) occupied just a small portion of the range of perceptibly different strengths of FA. The crucial point to note, however, is that the two groups of Early learners were judged independently to differ in strength of FA by three groups of NE-speaking listeners.

I next examined the 25 Early learners’ responses to 7-point rating scales on a language background questionnaire. Participants were asked, in separate items, to self-rate their accuracy in pronouncing English and Italian. The two groups’ self-rated pronunciation of English did not differ significantly, but the Poor pronouncers of English judged their Italian pronunciation to be significantly better than the Poor pronouncers did ( $p < .05$  by Kruskal-Wallis test). As well, the Good pronouncers of English reported using English significantly more often, and Italian significantly less often than the Poor pronouncers did ( $p < .05$  by Kruskal-Wallis tests). These results support the conclusion that differences in language use measurably affected the Early learners’ pronunciation of English after decades of predominantly English use.

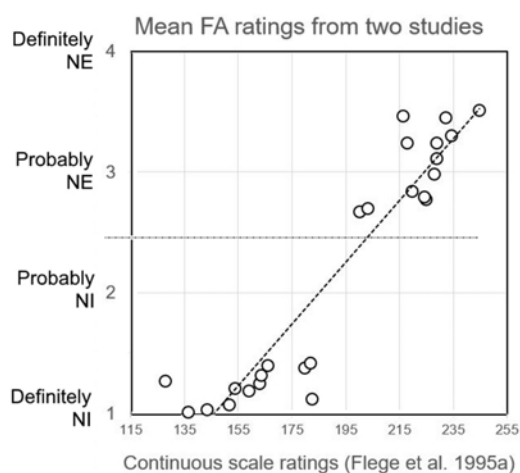


Figure 7. The mean foreign accent ratings obtained for 25 native Italian Early learners of English using two procedures for evaluating strength of foreign accent.

Another problem for the CP hypothesis is that foreign accents grow increasingly strong after the supposed closure of the CP. It is widely assumed that the ill effects of beginning to learn an L2 after the closure of a biologically based CP will be much the same for all *post-critical period* learners because the biological changes that trigger the close of the CP occur at roughly the same chronological age for all normally developing individuals (see, e.g., Lenneberg, 1967, Johnson & Newport, 1987; Birdsong, 2013).

Flege & MacKay (2011) tested this prediction by extending upward the AOA range examined in previous research. The participants in this study were three groups of 20 Italians each who had arrived in Canada at mean ages of 10, 18 and 26 years as well as an age-matched group of NE speakers. The ratings obtained for the NE speakers and the three AOA-defined groups of Italian immigrants all differed significantly from one another ( $p < .05$ ).

The CP hypothesis proposed by Lenneberg (1967) correctly predicted a stronger FA for members of the AOA-18 than the AOA-10 group. However it did not predict the significantly stronger foreign accents evident for members of the AOA-26 group than the AOA-18 group. Johnson and Newport (1989), in apparent agreement with Lenneberg, observed that “there are not many important maturational differences between ... the brain of a 17-year old and the brain of a 27-year old” (p. 79). Nor did the CP hypothesis predict the significantly lower ratings obtained for the pre-critical period members of the AOA-10 group than for the NE speakers.

### 2.3. Experienced L2 learners are bilinguals

The CP hypothesis presented by Lenneberg (1967), and later variants of this hypothesis, failed to consider a basic aspect of L2 learning. Immigrants who learn an L2 through immersion become bilinguals possessing two partially over-lapping phonetic subsystems which they usually cannot turn on or off either instantly or completely. Learning an L2 influences the native language (see Hopp & Schmid, 2013). According to the Speech

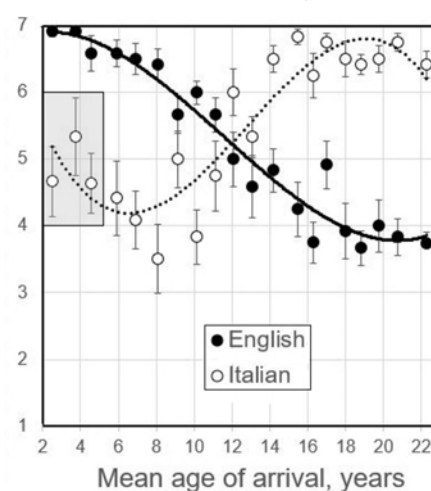


Figure 8. Mean self ratings of English and Italian pronunciation by Italians differing in AOA.

Learning Model (Flege, 1995) the phonetic elements comprising the L1 and L2 phonetic subsystems of a bilingual necessarily interact because they occupy the same phonetic/phonological space.

The 240 Italians tested by Flege et al. (1995a) self-rated their ability to pronounce English and Italian using 7-point rating scales. Fig. 8 shows the mean ratings obtained for 20 AOA-defined subgroups of 12 each. Most Italians who arrived in Canada before the age of 12 years reported having a better pronunciation of English than Italian whereas the reverse held true for those who arrived later in life. The self-ratings obtained for the 36 Italians with AOA values ranging from 1.7-5.3 years (grey box) may seem anomalous but are probably due to the fact that these participants were kept at home when they first arrived in Canada and so heard mostly or only Italian until school age.

The data in Fig. 8 supported the “common space” hypothesis but at the time of publication were judged to be of limited value because they were based on self-report. To further test the common space hypothesis, therefore, Yeni-Komshian, Flege and Liu (2000) recruited six male and six female Koreans at each age of arrival in the US ranging from 2 to 22 years. The 240 Koreans will be described in greater detail later when we consider the learning of English morphosyntax. For now, I simply note that the Koreans’ AOA values correlated as expected with their chronological ages,  $r(238)=.51$ , self-estimates of English use,  $r=-.50$ , self-estimates of Korean use,  $r=.51$ , years of US residence,  $r=-.42$ , and years of US education,  $r=-.92$ . All these variables, in turn, were inter-correlated ( $p<.05$ ).

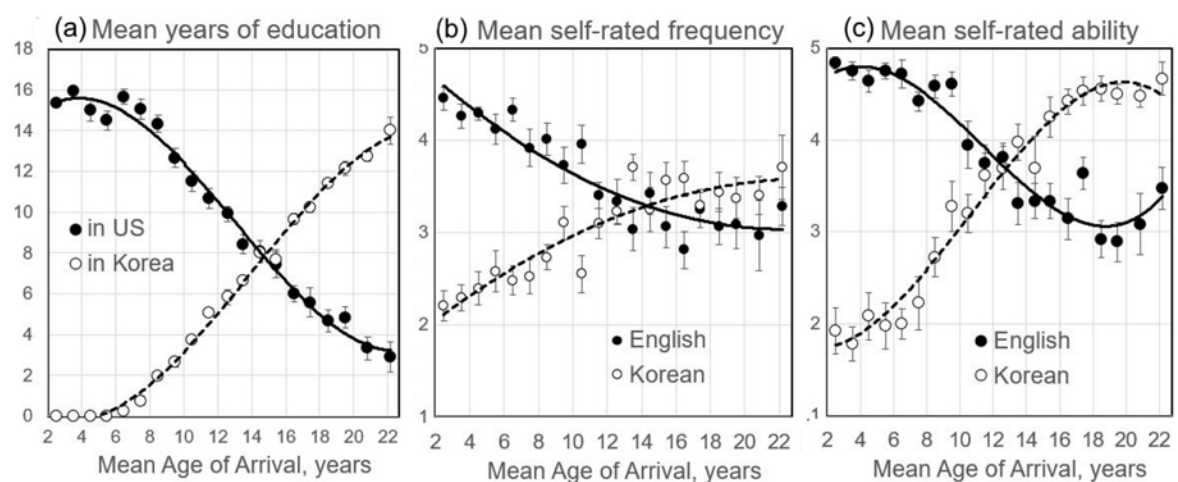


Figure 9. Relation between AOA and (a) years of formal education, (b) self-estimated frequency of use, and (c) self-estimated proficiency in Korean and English for groups of 12 immigrants each.

Fig. 9 shows the relation between the Koreans' mean AOA values and variables thought likely to influence their learning of English. Third-order polynomials have been fit to the mean values in all three panels to visually organize the data. Fig. 9(a) shows that an inverse relation existed between years of formal education obtained in Korean and American schools. Fig. 9(b) shows that Koreans who arrived in the US before the age of 12 years generally used English more than Korean whereas the reverse usually held true for Koreans who arrived after the age of 12 years. Finally, Fig. 9(c) shows that most Early learners judged themselves to be more proficient in English (5-point rating scales regarding pronunciation, reading/writing ability, and grammatical knowledge) than in Korean whereas the reverse usually held true for Late learners.

As observed earlier for Italian immigrants to Canada, the Koreans' use of their two languages depended importantly on context. Fig. 10 shows the mean ratings of frequency of Korean use obtained using scales that ranged from 1 (*very little*) to 5 (*very much*). The Koreans indicated using Korean far more frequently with their parents and when at home than while at work. We might think of these contexts as representing *obligatory contexts* in which the use of Korean or English was the norm. The remaining four contexts shown in Fig. 10, however, were intermediate in value to those observed in the home and work contexts. The ratings obtained for 120 Late and 120 Early learners in these *optional contexts* differed substantially ( $M=2.34$  vs  $3.51$ ,

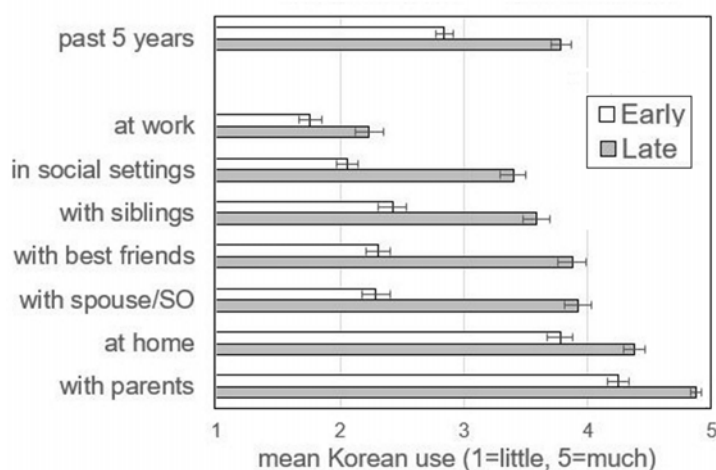


Figure 10. Mean ratings of Korean use in seven contexts by Early and Late learners.

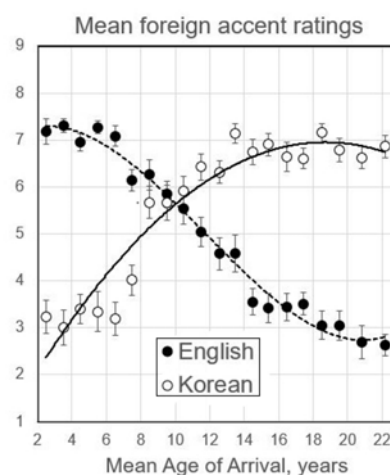


Figure 11. Mean foreign accent ratings obtained for Korean and English sentences.



$F(1,238)=134.5, p<.05$ ). Variation in language use in optional contexts like these was probably due to social factors rather than to the Koreans' state of neurocognitive maturation when they immigrated to the US.

Yeni-Komshian et al. (2000) had English and Korean listeners rate the 240 Koreans' pronunciation of Korean and English sentences for degree of FA using a 9-point scale. Fig. 11 reveals a crisscross pattern much like the one seen earlier for the Italians' self-ratings of L1 and L2 pronunciation ability. Higher ratings, indicating the presence of milder foreign accents, were obtained for Early learners' productions of English than Korean sentences whereas the opposite held true for the Late learners. This finding supported the hypothesis (Flege, 1995) that bilinguals' L1 and L2 phonetic systems interact in a common space.

## 2.4 Segmental production and perception

Lenneberg's (1967) CP hypothesis focused on global pronunciation of the L2, that is, the presence or absence of foreign accents. It was soon extended to phonetic research examining other aspects of L2 speech, namely the production and perception of L2 vowels and consonants. Some of that work will be presented here.

Later work in Ottawa with Italian immigrants employed an orthogonal design in which participants were selected on the basis of both AOA and language use. In one project 36 Early and 36 Late learners differing in AOA ( $M=7.5$  vs  $20.0$  years) were subdivided into subgroups of 18 each based on self-reported Italian use (Low Italian use  $M=7-10\%$ , High Italian use  $M=43-53\%$ ). Given that the Italians' use of English was inversely related to their use of Italian, I will refer to the groups differing in language use as the *English often/Italian seldom* and the *English seldom/Italian often* groups.

The 72 Italians took part in experiments examining overall degree of perceived FA (Piske, MacKay, & Flege, 2001), the identification of word-initial and word-final consonants (MacKay, Meador, & Flege, 2001), the production of English vowels (Piske, Flege, MacKay, & Meador, 2002), and the perception of English vowels (Flege & MacKay, 2004). In all four studies the Italians who used English often/Italian seldom obtained significantly higher scores than those who used English seldom/Italian often. This held true for both Early and Late learners. When taken together, the results of this research demonstrated the importance of language use but did not, of course, rule out a possible role of neurological maturation at the time the Italians immigrated to Canada.

One of the perception experiments carried out by Flege and MacKay (2004) focused on the vowels in English words like *beat* and *bit* (/i/, /ɪ/). These English vowels were of special interest because Italian has an /i/-quality vowel but not an /ɪ/-quality vowel. When Italians are first exposed to English they tend to hear both English vowels as being instances of their Italian /i/ category. However one experiment by Flege & MacKay (2004) suggested that although it is initially difficult for Italians to perceptually distinguish English /i/ from /ɪ/, doing so is eventually possible because of a perceived difference in the relation of the two English vowels to Italian /i/.<sup>4</sup>

Flege and MacKay (2004) examined the Italian immigrants' ability to detect errors in the production of English /i/ and /ɪ/. The stimuli were short phrases edited from the spontaneous conversations of Italian immigrants from an earlier study. The stimulus set included (a) correct productions of the vowel /i/ in phrases like "*speak the*", (b) correct productions of /ɪ/ in phrases like "*my kids they*", (c) incorrect productions of the target vowel /i/ as [ɪ], as in the phrase "*and reading*", and (d) incorrect productions of the target vowel /ɪ/ as [i], as in the phrase "*very difficult*".

The test phrases were presented auditorily and visually on a computer screen. An asterisk replaced the target vowel in the written phrases (e.g., sp\*k the; very d\*fficult) to localize the target vowel of interest in each phrase. The task on each trial was to decide if the target vowel had been produced correctly or incorrectly. An unbiased measure of sensitivity,  $A'$ , was computed based on hits (correct detections of errors) and false alarms (classifications of correct productions as incorrect).

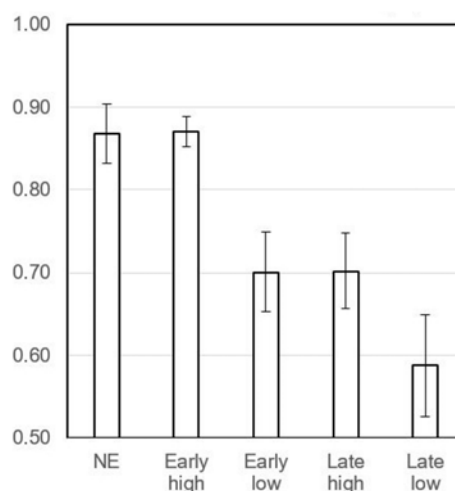


Figure 12. The mean perceptual sensitivity of NE speakers and four groups of Italians differing in AOA (Early vs Late) and language use (High vs Low) to vowel production errors.

<sup>4</sup> Young Italian adults who had little conversational experience in English classified English /ɪ/ tokens as Italian /i/ less often than they classified English /i/ as Italian /i/ ( $M=65\%$  vs  $87\%$ ). Moreover, the English /ɪ/ tokens were rated as being poorer instances of Italian /i/ than the English /i/ tokens were ( $M=2.9/5$  vs  $4.2/5$ ).

The mean A' scores obtained for the four Italian groups are shown in Fig. 12. Three Italian groups (Early-low use of English, Late-high use of English, Late-low use of English) showed significantly less perceptual sensitivity to vowel production errors than the NE speakers did ( $p < .01$ ). The only Italian group that did not differ from the NE group was Early-high, the group consisting of Early learners who used English often/Italian seldom. These findings are problematic for the CP hypothesis in two ways. First, a group of Early learners (Early-low) differed significantly from the NE speakers. Second, Early and Late learner groups (Early-low, Late-high) obtained virtually identical scores.

Language use turned out to be a slightly better predictor of the error detection scores than AOA did<sup>5</sup> and captured slightly more variance than AOA ( $\eta^2 = .162$  vs  $.158$ ). However the small advantage of Language use over AOA does not provide convincing evidence that input is more important than age of first exposure to an L2. Language use and AOA together accounted for only 32% of the variance in the error detection scores. This means that some other dimension(s) contributed importantly to the error detection scores.

The important dimension that was missing, I suspect, was the quality of English input that the Italian immigrants had received during their many years of Canadian residence. The more often the Italians used English, the more often they were likely to have taken part in linguistically mixed conversations involving a NE speaker and one or more other Italian immigrants. If so, they may have been more likely than Italians who used English seldom to hear English vowels spoken with an Italian accent, and so to have developed inaccurate perceptual representations for English vowels.

### 3. Learning L2 morphosyntax

In this section we turn our attention to the learning of English morphosyntax.

---

<sup>5</sup> AOA showed a stronger correlation to the error detection scores than either % English use or LOR examined individually,  $r(61) = -.402$  vs  $.340$ ,  $.230$ ). However the strength of correlation between the detection scores and *Years of English use* (LOR multiplied by % English) and AOA did not differ,  $r(61) = .412$  vs  $-.402$ . The correlation between % English use and the error detection score correlation remained significant when AOA was partialled out,  $r(60) = .26$ ,  $p < .05$ , but the AOA-detection correlation became non-significant when % English was partialled out,  $r(60) = -.24$ ,  $p > .05$ .

### **3.1. Maturational limits**

Johnson and Newport (1989) extended Lenneberg's (1967) CP hypothesis from L2 speech learning to the learning of L2 morphosyntax. These authors (henceforth "J&N") recruited a total of 46 Korean and Chinese adults who had arrived in the US between the ages of 3-39 years. To ensure that all participants had attained their ultimate level of performance in English, all were required to have lived in the US for at least five years, at least three years of which had to be continuous years of US residence prior to testing at the University of Illinois.

J&N developed a 276-item grammaticality judgement test (GJT) to evaluate knowledge of the "most basic aspects of English sentence structure" (1989, p. 72). Their test consisted of grammatical and ungrammatical versions of sentences such as *Last night the old lady died/\*die in her sleep*. The task of the adult native and non-native participants was to judge the randomly presented sentences as grammatical or ungrammatical. The authors described their test as "minimally demanding" because 6 to 7-year old NE-speaking children obtained "virtually perfect" scores on it (1989, p.70).

The percent correct scores obtained for the Korean and Chinese participants generally decreased as AOA increased. The scores obtained for seven immigrants having AOAs in the 3 to 7-year range did not differ significantly from the scores obtained for 23 NE speakers ( $M=97.6\%$  vs  $97.4\%$ ,  $p>.05$  by two-sample  $t$ -test). However the scores obtained for participants having AOA values in the 8 to 10-year range ( $n=7$ ,  $M=92.8\%$ ), in the 11 to 15-year range ( $n=8$ ,  $M=85.5\%$ ) and in the 17 to 39-year range ( $n=23$ ,  $M=76.2\%$ ) were all significantly lower than the NE speakers' scores ( $p<.05$ ). J&N also noted the existence of a significant correlation between AOA and test scores for participants having AOA values in the 3 to 15-year range but not in the 17 to 39-year range ( $r=-.87$  vs  $-.16$ ).

These findings led J&N to conclude that a critical (or sensitive) period exists for the learning of L2 morphosyntax. The period in question, which ranged from AOA values of 7 to 15 years, represented a period of rapid decreases in the GJT scores flanked by AOA ranges in which the GJT scores varied little or not at all. The authors hypothesized that the mechanism underlying the CP was either the reduction of an unidentified "language [learning] faculty" or a general change in "cognitive abilities involved in language learning", itself a consequence of normal neurological and/or cognitive maturation (p. 61).

The conclusions drawn by J&N regarding both the mechanism underlying the hypothesized CP as well as its offset have been contested. Hakuta and Bialystok (1994) noted, for example, that when the CP offset was shifted from an AOA of 15 to 20 years, a significant correlation was found to exist between AOA and the scores obtained for both the newly defined pre- and post-critical period learners. Hartshorne et al. (2018, p. 271) observed that J&N would almost certainly have observed a significant correlation between test scores and the AOA values of Later arrivals had they simply recruited more participants. This observation applies equally to the apparent existence of an *optimal age*, that is, the absence of a significant difference between NE speakers and the seven nonnatives having AOA values in the 3 to 7-year range.

## **3.2. Korean immigrants in the US**

### **3.2.1 Participants**

The aim of Flege et al. (1999) was to replicate and extend the classic Johnson and Newport (1987) study. These authors tested 295 Korean adults in 1992-1996 at the University of Maryland (UMD). All were tested individually by a bilingual Korean-English research assistant in a single 1.5-hour session. Most participants were current or former UMD students or faculty members. All were able to speak English as shown by their ability to repeat English sentences fluently following a filled delay. Twenty-six Koreans had to be excluded from the study because they could not fluently repeat Korean sentences following a delay and so might not have been bilinguals. Another 29 Koreans were excluded for one or more of the following reasons: not speaking the Seoul dialect of Korean, speaking a language other than English and Korean; having lived in a country other than Korea and the US.

The experimental protocol specified a minimum of 10 years of residence in the US. When it became evident after three years of recruitment that the 10-year minimum would make it impossible to fill all 20 one-year AOA bins with 6 males and 6 females, the minimum was reduced to 8 consecutive years of US residence. Two Koreans retained for the study were married to a NE speaker but even they reported using Korean in the home. The ages of the 240 Koreans retained for the study differed little ( $M=26$  years,  $range=17-46$ ) from the ages of the 24 NE speakers who formed the comparison group ( $M=27$  years,  $range=20-45$ ). Statistical analyses presented by Flege et al. (1999) focused on groups formed by combining adjacent 1-year

AOA groups, thereby creating 10 AOA-defined groups of 24 each. The mean AOA values of these groups of 24 each ranged from 3-22 years.<sup>6</sup>

### **3.2.2 Grammaticality Judgment Test**

The 18-min test used by Flege et al. (1999) was derived from the GJT developed earlier by Johnson and Newport (1987). Sentences from the original test that probed knowledge of Auxiliaries, Word order and the Present progressive were eliminated because they had served little to distinguish Early from Late learners. By adding to the length of the test, these sentences may have contributed to errors due to inattention or fatigue. Sixteen new items testing lexically specified subject/object raising were added, however, yielding a GJT test that consisted of 144 sentences, half grammatical and half ungrammatical.

The sentences comprising the new GJT were recorded at a constant moderate rate by a single native speaker of English (JEF) who took care to articulate word-final stops (e.g., the /d/ in *died*) and fricatives (e.g., the /s/ in *paints*). A short break occurred between the two halves of the GJT. The grammatical and ungrammatical versions of each sentence pair were presented in separate halves of the test. The written presentation of each sentence was accompanied by a single aural presentation of the same sentence, with a fixed 4.0 sec interval between sentences. Participants were required to respond “Yes” (grammatical) or “No” (ungrammatical) to each sentence before moving on to the next sentence.

### **3.2.3 Results**

As in the study by Johnson and Newport (1987), the GJT scores obtained for Korean immigrants by Flege et al. (1999) generally decreased as AOA increased. Korean groups having mean AOA values in the 7 to 22-year range, but not those having mean AOA values in the 3 to 5-year AOA range, obtained significantly lower GJT scores than the NE speakers did

---

<sup>6</sup> Flege et al. (1999) largely avoided a confound between AOA and years of formal education. The Koreans' highest educational attainment averaged 15.7 years ( $SE=.11$ ,  $range=12-21$  years). The effect of Group (10 levels) on education was significant,  $F(9,230)=2.1$   $p<.05$ . However just one pair-wise difference between the ten groups reached significance, that between Koreans having AOA values of 5 and 19 years ( $M=14.8$  vs  $16.5$ ,  $p<.05$  by Tukey test). This difference arose because ten members of the AOA-5 group had completed less than 3 years of college because of their young age (17-20 years) but this held true for just three members of the somewhat older AOA-19 group.

(Bonferroni-corrected  $p < .05$ ). A Gumpertz-Makeham growth function was fit to the percent correct GJT scores obtained for the 240 Koreans to visually organize the data. Inspection of this function suggested that most Koreans having AOA values in the 2 to 6-year AOA range obtained scores resembling those of the NE speakers. Scores for Koreans in the 7 to 15-year AOA range decreased in a near linear fashion as AOA increased whereas the scores for Koreans having AOA values in the 16 to 22-year range continued to decrease as AOA increased, but at a slower rate.

Flege et al. (1999) submitted the Koreans' responses to 39 questionnaire items to a Principal Components Analysis (PCA) to identify factors underlying the AOA effects just described. Factor scores derived from the PCA were then used as predictors of the morphosyntax scores in a step-wise multiple regression analysis. A factor named *Age of L2 learning* accounted for far more variance in the GJT scores than a factor named *Length of residence* did (67.7% vs 3.6%). This might be taken as support for the traditional view of AOA effects, namely that Early vs Late differences are due to the loss of capacity to learn an L2 following closure of a CP and that variation in input contributes little if at all to Early vs Late differences.

There are two reasons to question this interpretation. First, consider the regression analysis used to identify factors underlying age-related variation in the GJT scores. The factor accounting for most (67.7%) of the variance was named Age of L2 learning because the questionnaire variable having the highest loading on it (.912) was AOA. However, five other variables also had high loadings on the Age of learning factor: Years of education in the US (-.856), Use of Korean with spouse (.786), Use of Korean with close friends (.729), Use of Korean at social gatherings (.737) and Use of English at social gatherings (-.712). The multi-collinearity lurking beneath the surface of the Age of L2 learning factor made it impossible to directly evaluate the role of AOA in differentiating Early from Late learners of English.

Second, the influence of AOA was not the same for all sentences examined. AOA correlated significantly with all nine sentence types in the GJT used by Flege et al. (1999). However the strength of correlation ranged from a low of  $r(238) = -.44$ , for sentence pairs examining Third-person singular (e.g., *Every Friday our neighbor washes/\*wash her car*), to a high of  $r(238) = -.74$ , for pairs examining Determiners (e.g., *The boy is helping the man build a/\*Ø house*). Johnson and Newport (1989, Fig. 3) also noted substantial differences across sentence types for Korean and Chinese and immigrants.

The analysis of sentence types provided little general understanding of the learning of L2 morphosyntax. This is because the grammatical structures that proved difficult for the Koreans might not prove difficult, or else manifest a different degree of learning difficulty, for speakers of other languages. To provide a more general understanding of L2 morphosyntax learning, therefore, Flege et al. (1999) calculated two morphosyntax subscores that will be referred to here as the *Rule-based* and *Lexicon-based* scores. The two scores represented a functional rather than syntactically motivated grouping of sentences from the GJT since the items upon which the scores were based were drawn from multiple sentence types (see Appendices 1 & 2 in Flege et al., 1999). Both scores were based on responses to 44 sentences (half grammatical, half ungrammatical) identified through Principal Components Analyses.

The Rule-based scores probed knowledge of regular, productive and generalizable rules of the surface morphology of English. All involved case or number assignment on nouns, or person, or tense markers on verbs. For example: \*A/The boys are going to the zoo this Saturday; The man \*paints/painted his house yesterday, \*Them/They worked on the project all night. The Lexicon-based sentences, on the other hand, probed irregular and ungeneralizable aspects of English morphosyntax involving the proper assignment of particles or prepositions with verbs, or knowledge of idiosyncratic features of English verbs. For example: The farmers were \*hoping/hoping for rain; The little boys \*laughed/laughed at the clown. All ungrammatical Lexicon-based sentences could be made grammatical by replacing the verb (e.g., changing “lets” to “permits” in \*The man lets his son to watch TV). However the ungrammatical Rule-based sentences could not be corrected in this way.

The Lexicon-based and Rule-based scores were likely to draw on different forms of memory: declarative vs procedural. Also, the Lexicon-based scores were likely to be more susceptible to variation in input than the Rule-based scores (Pinker & Prince, 1991; Prasada & Pinker, 1993;

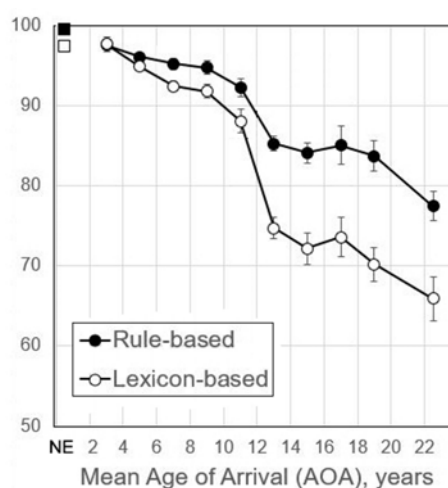


Figure 13. Mean Rule-based and Lexicon-based scores obtained for native English (NE) speakers and 10 groups of Korean differing in age of arrival (AOA).



Beck, 1997; see also Birdsong & Flege, 2001). That being the case, we expected to observe larger Early vs Late differences for Lexicon-based than Rule-based scores. This expectation was confirmed as can be seen in Fig. 13. The effect of AOA was greater for the Lexicon-based than Rule-based scores. AOA showed a significantly stronger correlation with the Lexicon-based than Rule-based scores ( $r=-.71$  vs  $-.58$ ,  $X(1)=12.3$   $p<.05$ ) and captured more variance ( $\eta^2=.513$  vs  $.318$ ).

Differences between the NE group and 10 AOA-defined groups of 24 Koreans each were evaluated using one-sample  $t$ -tests. All 10 Korean groups obtained significantly lower Rule-based scores than the NE speakers did ( $t$ - values ranging from  $-3.1$  to  $-10.1$ , Bonferroni-corrected  $p<.05$ ). However just eight Korean groups – those having mean AOA values in the 7 to 22-year range – obtained significantly lower Lexicon-based scores than the NE speakers ( $t$ - values ranging from  $-3.5$  to  $-13.3$ , Bonferroni-corrected  $p<.05$ ). Koreans having mean AOA values of 3 and 5 years did not obtain significantly lower Lexicon-based scores than the NE speakers ( $t=0.4$  and  $-2.9$ , respectively, Bonferroni corrected  $p>.05$ ).

The absence of significant differences in Lexicon-based scores between NE speakers and Koreans with AOAs of 3 and 5 years was probably not the result of a ceiling effect for the NE speakers. There was less variance, to be sure, in the NE speakers' Rule-based scores ( $M=99.5\%$  correct,  $SE=.19$ ,  $range=97.7$  to  $100.0$ ) than in their Lexicon-based scores ( $M=97.3$ ,  $SE=.90$ ,  $range=81.8$  to  $100.0$ ). Of the 24 NE speakers tested, 19 were at ceiling from the Rule-based scores as compared to just 12 for the Lexicon-based scores. Had a more difficult test been administered more NE speakers would have been off ceiling for the Rule-based scores; but even with 19 of 24 at ceiling, significant native vs nonnative differences were detected. A more likely explanation for the absence of a difference between the NE and two Korean groups was a small effect size. If so, native vs nonnative differences for the Lexicon-based scores would probably have been obtained for NE speakers and Koreans with mean AOAs of 3 and 5 years if a larger number of Koreans had been tested.

Two-sample  $t$ -tests were used to evaluate differences between the Rule-based and Lexicon-based scores. The Koreans having mean AOA values in the 13 to 22-year range obtained significantly lower Lexicon-based than Rule-based scores ( $t$ -values ranging from  $4.6$  to  $8.9$ ,  $df=23$ , Bonferroni-corrected  $p<.05$ ) whereas the Lexicon-based and Rule-based scores obtained for Koreans having mean AOA values in the 3 to 11-year range did not differ significantly ( $t$ -values ranging from  $-0.2$  to  $2.2$ ,  $df=23$ ,  $p>.05$ ).

Why did an AOA of 12 years mark the point of demarcation between the Rule-based and Lexicon-based scores? In the popular mind – and that of many researchers as well – the notion of a critical period is associated with the age of 12 years. As seen in Fig. 9(c) an AOA of 12 years was the point of demarcation between Koreans who usually judged themselves to more proficient in English than Korean from those who usually reported the opposite. The age of 12 years is also relevant in another way. It is the age at which mandatory instruction in English begins in Korean schools. Johnson and Newport (1987, p. 69) raised the issue of whether immigrants' *age of exposure* to an L2 should be defined as the age at which they began to study the L2 at school in their home country or age of emigration to a predominantly L2-speaking country.

To evaluate the influence of school study before immigration, I examined the results obtained for Koreans in 16 of the original 20 one-year AOA bins of 12 each. Participants who arrived in the US before the age of 4 years were excluded from this analysis because of uncertainty as to whether they only (or mostly) heard Korean until school age after arriving in the US. The Koreans who arrived in the US at the age of 12 years were also excluded because of uncertainty as to whether they had already begun to study English at school before departing for the US. Finally, to create balanced groups of Early and Late learners, Koreans who arrived in the US after the age of 20 years were also excluded.

This procedure yielded groups of 96 Early and 96 Late learners having AOA values that both spanned a seven-year range: 4 to 11 years for the Early learner and 13 to 20 years for the Late learners. As expected, significantly higher scores were obtained for Early than Late learners ( $M=93.1\%$  vs  $77.2\%$ ,  $F(1,190)=179.5$ ,  $p<.05$ ) and for the Rule-based than Lexicon-based scores ( $M=88.9\%$  vs  $81.5\%$ ,  $F(1,190)=158.3$ ,  $p<.05$ ). A significant interaction was obtained ( $F(1,190)=62.8$ ,  $p<.05$ ) because the Rule-based and Lexicon-based scores differed more for Late learners ( $83.2\%$  vs  $71.1\%$ ) than for Early learners ( $94.5\%$  vs  $91.8\%$ ).

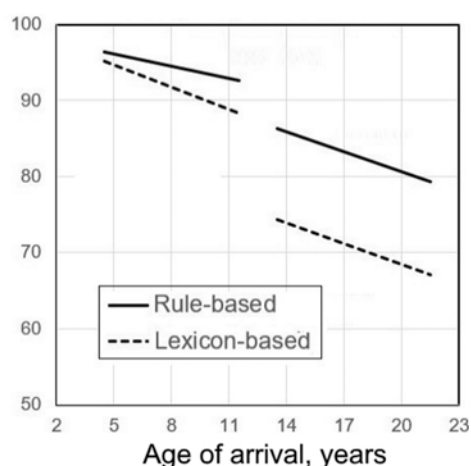


Figure 14. Best fitting functions relating AOA to scores obtained for Early and Late learners.

The best fitting linear functions for the two GJT scores are shown in Fig. 14 for the Early and Late learners. The slopes of all four functions differed significantly from zero ( $t(94)=-2.35$  to  $-4.29$ , Bonferroni-corrected  $p<.05$ ) indicating that the GJT scores decreased as AOA increased over a 7-year range for both groups. The Early and Late learners' slopes did not differ significantly from one another for either the Rule-based or Lexicon based scores ( $p>.05$ ). If exposure to English at school in Korea really mattered, we would have expected steeper slopes for the Early than for the Late learners, whose age of first exposure to English at school in Korea was a constant 12 years.

There is no guarantee, of course, that the downward trends seen in Fig. 14 were due to 7-year increases in AOA, nor that the abrupt decrements in performance at an AOA of 12 years were due to the closure of a CP at that age. As seen earlier in Fig. 9, AOA was correlated with variables that might have affected the Rule-based and Lexicon-based scores: Years of US education, Korean use, and English use. As well, Years of US education and frequency of English use correlated with one another,  $r(238)=.53$ ,  $p<.05$ . As years of education in American schools increased, the likelihood that the Korean participants had received explicit instruction on some of the grammatical structures probed by the GJT was also likely to have increased. Perhaps more importantly, Years of education in the US was probably related to the number of enduring relationships the Korean immigrants established with NE speakers. This may have resulted in greater exposure to correct models of English morphosyntax and more English input overall.

Bahrack et al., (1994) noted that AOA and Years of education is commonly confounded in studies of immigrant populations. This is important in that Years of education exerts a strong effect on immigrants' self-reported proficiency in the L2 (Hakuta et al., 2003). Flege et al. (1999) used a subgroup matching task to circumvent the AOA-

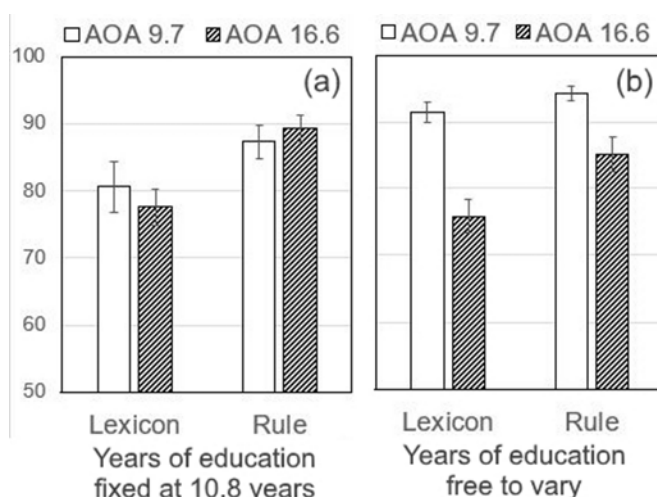


Figure 15. Mean Lexicon-based and Rule-based scores for subgroups of 20 Koreans each who did or did not differ in years of education in the U.S.

Years of education confound. Two subgroups of 20 Koreans each were identified from the larger sample of 240 participants. Members of the two groups had non-overlapping AOA values ( $M=9.7$  vs  $16.6$  years) but were matched for Years of education in the US ( $M=10.8$  years for both).

The GJT scores obtained by Flege et al. (1999) for the *matched* subgroups are shown in Fig. 15(a). As expected, the Rule-based scores were significantly higher than the Lexicon-based scores ( $M=89.4\%$  vs  $79.1\%$ ,  $F(1,38)=31.3$ ,  $p<.05$ ). Importantly, the difference between the Early and Late learners ( $M=84.0\%$  vs  $83.5\%$ ) was non-significant for both GJT scores ( $p>.05$ ). The absence of a significant difference between the education-matched groups of Koreans having mean AOA values of  $9.7$  vs  $16.5$  years undermines the view that Early vs Late differences are due to age-related differences in state of neurocognitive maturation at the time L2 learning begins.

The absence of an AOA effect for matched subgroups might be attributed to a lack of statistical power. Flege et al. (1999) therefore carried out a control analysis comparing two new groups of 20 Early and 20 Late learners. These *unmatched* subgroups had the same mean AOA values as the those in the matched subgroup analysis ( $9.7$  and  $16.6$  years). However participants in the two new groups of Early and Late learners differed significantly in terms of Years of education in the US ( $M=14.4$  vs  $8.0$ ,  $p<.05$ ) and the ratio of English/Korean use ( $M=1.6$  vs  $0.9$ ,  $p<.05$ ), because they were selected without regard to any variable other than AOA.<sup>7</sup>

As shown in Fig. 15(b), the unmatched Early and Late learners differed significantly as is usually the case ( $F(1,38)=24.4$ ,  $p<.05$ ). A two-way interaction was obtained,  $F(1,38)=7.91$ ,  $p<.05$ , because the difference between Rule-based and Lexicon-based scores was significant for the Late learners ( $p<.01$ ) but not the Early learners ( $p>.05$ ). This finding suggests that the lack of a significant difference between Early and Late learners in the matched analysis was not due to a lack of statistical power.

As seen earlier in Fig. 9(b), AOA was also confounded with language use in the sample of Korean immigrants tested by Flege et al. (1999). Accordingly, a similar matched subgroup analysis was undertaken to evaluate the effect of AOA when language use was controlled. This analysis compared subgroups of 20 Koreans each who were matched for AOA but differed in terms of language use. Participants selected for the “High-ratio” subgroup had English/Korean use ratios greater than

---

<sup>7</sup> The small LOR difference between members of the two unmatched groups ( $M = 14.0$  vs  $12.1$ ) was non-significant ( $p > .05$ ).

1.4 ( $M=2.07$ ,  $range=1.43-3.24$ ) whereas those selected for the “Low-ratio” subgroup all had ratios less than 1.0 ( $M=0.73$ ,  $range=0.50-0.99$ ).<sup>8</sup> Both subgroups had a mean AOA of 13.0 years ( $range=5-22$  years). The matched subgroups did not differ significantly in either age ( $M=25.4$  vs 27.1 years) nor Years of education in the US ( $M=11.0$  vs 11.6 years).

Selecting participants for the High-ratio and Low-ratio groups required the identification of Koreans who did not show the usual effect of AOA on language use. As seen in Fig. 9(b), most Late learners in the original sample of 240 used English slightly less often than Korean, yielding English/Korean ratios that were less than 1.0. On the other hand, most Early learners used English more than Korean, yielding ratios that were greater than 1.0. These statistical regularities did not hold true for all individuals, however. For example, a Late learner might have had an English/Korean use ratio greater than 1.0 if s/he used English exclusively at work and had mostly NE-speaking friends. An Early learner might have had an English/Korean ratio less than 1.0 if s/he needed to use Korean at work and was married to a Korean Late learner who wanted to speak only Korean at home.

Fig. 16 shows the scores obtained for the High-ratio and Low ratio subgroups. An ANOVA examining these scores yielded a significant interaction,  $F(1,38)=4.49$ ,  $p<.05$ , for two reasons. First, the Rule-based and Lexicon-based scores differed significantly for the Low-ratio group ( $M=87.5\%$  vs  $79.7\%$ ,  $p<.05$ ) but not for the High-ratio group ( $M=92.2\%$  vs  $87.5\%$ ,  $p>.05$ ). More importantly, the Koreans who used English more than Korean (i.e., High-ratio group members) obtained significantly higher Lexicon-based scores than those who used English less than Korean (i.e., Low-ratio group members,  $M=87.5\%$  vs  $79.7\%$ ,  $p<.05$ ). However the High-ratio and Low-ratio groups did not differ significantly for the Rule-based scores ( $M=92.2\%$  vs  $91.3\%$ ,  $p>.05$ ). This find-

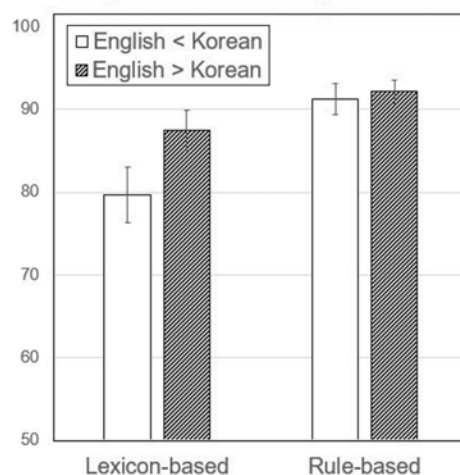


Figure 16. Mean percent correct scores obtained for Koreans who differed in language use but were matched for AOA.

<sup>8</sup> The ratio of self-reported use of Korean (9 questions) and English (7 questions) were used here to select participants. However Flege et al. (1999) used the mean frequency of Korean use.

ing suggests that a relatively large quantity of English input facilitates the learning of morphosyntactic properties of English, but only those properties that must be learned on a word-by-word basis (i.e., the Lexicon-based scores).

It is important to note, however, that both the High-ratio and Low-ratio groups obtained significantly lower Rule-based and Lexicon-based scores than the NE speakers did ( $p < .05$  by Bonferroni-corrected one-sample t-tests). This should probably not be interpreted to mean that complete learning is impossible, however. This is because we cannot be sure that the Koreans had obtained adequate English input.

First, we have no information regarding how much of the English input the Koreans received consisted of correct productions of English sentences by NE speakers. If the Koreans often heard ungrammatical English sentences produced by other Koreans they would not, of course, be expected to respond just like monolingual native speakers of English. This assumes, of course, that the NE speakers who formed the comparison group had heard only well-formed utterances produced by fellow NE speakers.

Second, the quantity of input the Korean immigrants had received may have been insufficient for complete learning to have occurred. I calculated percent English use values from the ratings presented earlier. Members of the High-ratio group used English more frequently than members of the Low-ratio group did ( $M = 66.7\%$  vs  $41.8\%$ ) and had also lived somewhat longer in the US than members of the Low-ratio group had ( $M = 14.5$  vs  $12.7$  years). Multiplying the percentage use estimates by LOR provides an estimate of Years of full-time English input. By this measure, members of the High-ratio group had received substantially more English input than members of the Low-ratio group had ( $M = 9.7$  vs  $5.4$  years,  $p < .05$ ).

However, 9.7 years of full-time English input distributed over 15 years of residence may not have been enough for the Korean immigrants to have learned some aspects of English morphosyntax completely. Even though members of the NE comparison group had an average age of 27 years, only half of them were at ceiling for the Lexicon-based scores. Learning some things takes time. Indeed, some of the 90 native English adults tested by Dąbrowska (2018) performed at a chance level for certain English grammatical constructions even though they had a mean age of 38 years ( $range = 17-65$ ) and had performed well on other constructions. Importantly, Dąbrowska (2018) found that some aspects of grammatical knowledge depended on amount of formal education and print exposure.

This finding points to the importance of input, even for monolingual adult native speakers of English.

Hartshorne et al. (2018) proposed that NE speakers may need as much as 30 years of full-time input to completely learn English grammar. One member of the High-ratio group had in fact lived in the US for 30 years. However this outlier had received only the equivalent of 20.3 Years of full-time English input because he used English only part of the time. His Lexicon-based score was 91% correct. One wonders if this participant's learning of English morphosyntax had stopped irreversibly, or if it would continue improving slowly over time until reaching a native-like level of attainment once 30 years of full-time English input had been received.

#### 4. Speech vs morphosyntax learning

The CP hypothesis has been applied to the learning of both L2 speech and morphosyntax. Investigators have tended to treat L2 pronunciation and morphosyntax as separate entities that require different kinds of learning and different explanations for age-related effects. Some think that a CP closes sooner for the learning of L2 speech than morphosyntax (e.g., Long, 1990) and others that a CP exists only for L2 speech learning (e.g., Scovel, 1988; Bahrick et al., 1994). The seemingly greater difficulty in pronouncing an L2 without a FA than in obtaining morphosyntax scores equaling those of L2 native speakers has been attributed to differences in the extent to which learners relate structures found in the L1 and L2 (e.g., MacWhinney, 1992) or the neuromotor component involved in speech production (Zatorre, 1989) and possibly speech perception (Best, 1995).

Oyama (1973) and Patkowski (1980, 1990) reported somewhat stronger correlations between AOA and L2 pronunciation than between AOA and L2 morphosyntax scores. Not surprisingly, Flege et al. (1999) observed a stronger correlation between AOA and overall degree of FA than between AOA and GJT scores,  $r(238) = -.85$  vs  $-.75$ ,  $X(1) = 19.9$ ,  $p < .05$ .

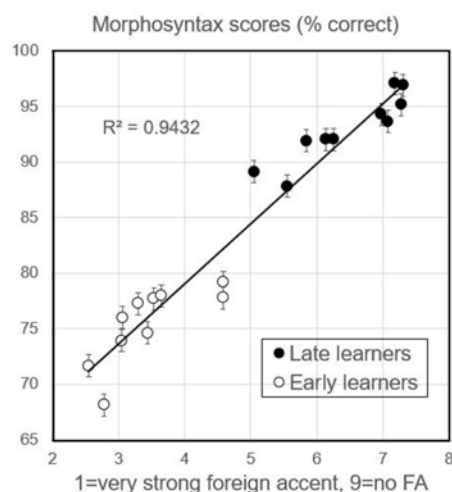


Figure 17. Mean foreign accent ratings and morphosyntax scores for 20 groups of 12 Korean each who differed in AOA.

I carried out an *F*-test comparing the Early and Late learners tested by Flege et al. (1999). AOA accounted for slightly more variance in FA ratings than GJT scores did ( $\eta^2=.623$  vs  $.548$ ). Fig. 17 shows that the FA ratings and GJT scores obtained for 20 AOA-defined Korean groups correlated strongly with one another ( $r(18)=.97$ ,  $p<.05$ ). The FA ratings and GJT scores also correlated significantly when the scores obtained for 120 individual Early learners and 120 individual Late learners were examined ( $r(118)=.62$  vs  $.58$ ,  $p<.05$ ).

The similar correlations obtained here for learning in the two linguistic domains, both for Early and for Late learners, suggest that an important commonality exists for the learning of L2 speech and morphosyntax. That underlying commonality may be input. Differences in the quantity and quality of input that Early and Late learners receive exert a strong effect on the learning of L2 speech and at least some aspects of L2 morphosyntax learning.

## **5. A non-critical period for L2 learning**

### **5.1 LOR and ultimate attainment**

The critical period (CP) hypothesis proposed by Lenneberg (1967) for L2 speech learning derived from the observation that people who learn an L2 after puberty usually speak it with a FA. Lenneberg probably assumed that a detectable FA remains evident in the speech of post-pubescent learners even after they have received abundant native-speaker input for many years. Had this not been so, Lenneberg's observation regarding foreign accent would not have evoked widespread interest and launched a new research subdiscipline.

The unspoken assumption I just attributed to Lenneberg (1967) later surfaced in the L2 acquisition literature under names such as *ultimate attainment*, *endstate* and *asymptotic learning* (Birdsong, 2013). Researchers soon became attentive to the potential role of L2 experience. For example, to ensure ultimate attainment in English morphosyntax learning, DeKeyser (2000) required that all 57 of the Hungarian immigrants he tested in the Pittsburgh area have lived in the US for at least 10 years. The GJT scores obtained for Hungarians who arrived in the US before vs after the age of 15 years showed little overlap. This led DeKeyser to conclude that Early vs Late differences derive from changes in cognitive processing. More specifically, DeKeyser concluded (2000, p. 518) that "somewhere between the ages of 6-7 and 16-17, everybody loses the mental equipment required for [learning] the abstract patterns underlying a human language".



DeKeyser also concluded (2000, p. 518) that input differences “are not a good explanation for age effects” on L2 morphosyntax learning. This conclusion was based on the complete absence of correlation between the Hungarians’ lengths of residence in the US and their GJT scores. It seemed that once the Hungarians had reached what DeKeyser thought was their ultimate attainment in English after 10 years of US residence further increases in LOR no longer resulted in much if any further improvement.

The conclusion drawn by DeKeyser (2000) regarding the inefficacy of LOR beyond 10 years of residence in a predominantly L2-speaking country was supported by two findings. For the 240 Koreans tested by Flege et al. (1999), GJT scores showed a modest correlation with LOR,  $r(238)=.39$ ,  $p<.05$ . However the strength of correlation between test scores and LOR decreased to  $r(181)=.23$  ( $p<.05$ ) for the subset of Koreans who had lived in the US for more than 10 years, and to a non-significant  $r(151)=0.12$  for Koreans who had lived in the US for more than 12 years. In an analysis of US census data, Stevens (1999) found that immigrants’ self-reported proficiency in English increased rapidly as a function of LOR until about 10 years of residence in the US, but that additional increments in LOR beyond 10 years were associated with little further improvement in self-reported English-language proficiency.

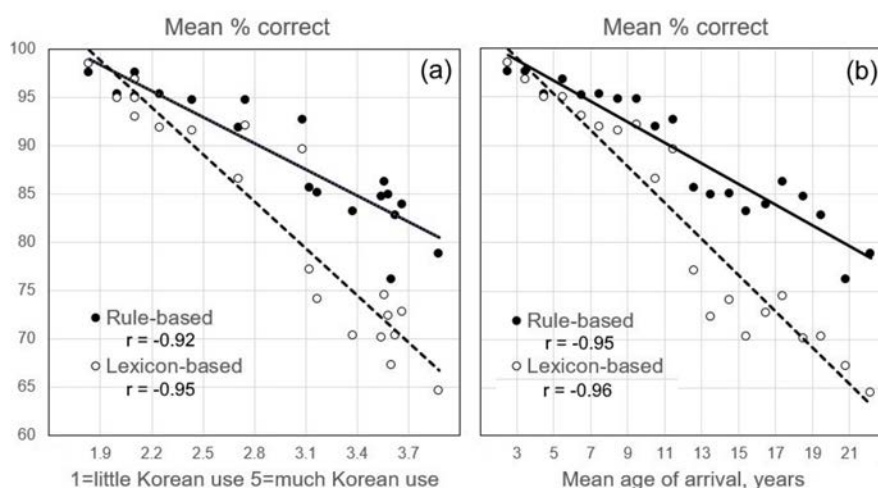


Figure 18. Mean morphosyntax scores obtained for 20 groups of 12 Koreans each as a function of self-estimated Korean use (left) and AOA (right).

The conclusion drawn by DeKeyser (2000) regarding the scant role of input in L2 morphosyntax learning, however, is questionable. This is because LOR does not provide an adequate index of quantity of L2 input (e.g., Flege & Liu, 2001) and offers no insight at all regarding the quality

of L2 input that immigrants receive. DeKeyser (2000) did not attempt to determine how much, with whom, in what social contexts, and for what purpose(s) his native Hungarian participants used English. Had this been done he might well have drawn a different conclusion regarding the role of input.

Fig. 18(a) shows the Rule-based and Lexicon-based scores obtained by Flege et al. (1999) for 20 AOA-defined groups of Koreans as a function of the groups' average estimated frequencies of Korean use in four language-optional contexts (see Fig. 10).<sup>9</sup> The same GJT scores are shown in Fig. 18(b) as a function of AOA.

The correlations between the GJT scores and the two predictor variables (frequency of Korean use, AOA) were both very strong. Somewhat weaker but still highly significant correlations between the GJT scores and the two predictor variables were obtained when the scores for all 240 Koreans were examined rather than mean values obtained for groups of 12 Koreans each. In these analyses the correlations between Rule-based and Lexicon-based scores and AOA were  $r(238) = -.60$  and  $-.74$ ,  $p < .05$ , while the correlations between the test scores and self-reported frequency of Korean use were  $r(238) = -.48$  and  $-.60$ ,  $p < .05$ .

The correlations with AOA obtained for individual participants were significantly stronger than the correlations with the Korean use estimates. This held true for both GJT scores ( $p < .05$ ). The difference in strengths of correlation was probably due to measurement precision. Language use estimates obtained using rating scales on a language background questionnaire are inherently noisier than AOA, an objective measure that is correlated with other variables likely to influence GJT scores. In fact, when the effect of Years of education in US schools was partialled out, the only correlation that remained significant was that between the Lexicon-based morphosyntax scores and Korean use,  $r(237) = -.21$ , Bonferroni corrected  $p < .05$ .

The results of this analysis suggest that there is no real justification, other than tradition, for concluding that AOA offers a better explanation of Early vs Late differences than variation in language use does. Indeed, it is plausible to hypothesize that variation in input is the single most important

---

<sup>9</sup> The bilinguals tested by Flege et al. (1999) spoke only English and Korean, which means that English use was the inverse of Korean use. In the analysis presented here I used Korean rather than English use estimates because the Koreans were not asked to estimate English use in the four context-optional contexts which, in retrospect, appear to have been the most indicative.

predictor of Early vs Late differences. AOA is regarded as a stand-in for other potentially causative variables such as state of neurocognitive maturation at the time of first exposure to an L2. Frequency of language use, on the other hand, relates directly to the input that is essential for the learning of L2 speech and at least some aspects of L2 morphosyntax.

## 5.2 Large sample studies

Most investigations of L2 learning by immigrants (e.g., Flege et al., 1999; DeKeyser, 2000) have examined relatively small numbers of participants. Here I consider two studies that examined far larger numbers of participants than is typical for L2 research.

Hartshorne et al. (2018) obtained responses to a 10-min grammar quiz from individuals who had learned English as a foreign or second language. These authors obtained quiz scores via the internet from 45,067 *immersion learners*, respondents who had spent at least 90% of their lives in an English-speaking country once having arrived there; from 266,701 *non-immersion learners*, who had spent at most 10% of their lives – but never more than one year – in an English-speaking country since first exposure to English; and from 246,497 native speakers of English from around the world. Some non-immersion learners reported having first been exposed to English at school and to have received subsequent input by watching TV programs and movies in English (2018, p. 266) although the sources of input and the context of first exposure to English were not systematically examined.

Hartshorne et al. (2018) obtained substantially higher grammar quiz scores for immersion than non-immersion learners of English (see Fig. 6). This supports the view that performance in an L2 depends on the amount of input received, assuming of course that the immersion learners had indeed received more English-language input than the non-immersion respondents had.

The aim of one analysis of special interest was to estimate the age of first exposure to an L2 beyond which “mastery ... [of L2 grammar] is no longer attainable” (Hartshorne et al., 2018, p. 270). This analysis examined scores obtained from 25% of the original immersion group respondents and 11% of the non-immersion respondents. To be included in this analysis respondents had to be less than 70 years of age and to have had at least 30 years of experience in English. The first criterion was meant to obviate the influence of cognitive losses due to normal aging. The second criterion was meant to ensure ultimate attainment in English proficiency. The

variable *Years of experience* indicated the difference, in years, between the respondents' age at test and the age of arrival in a predominantly English-speaking country (immersion learners) or the age of first exposure (non-immersion learners).

Hartshorne et al. (2018) found that the immersion learners showed "little decline in ultimate attainment [in English] until an age of first exposure of 12 years" whereas the non-immersion learners showed no decline until the age of 9 years and a "sharp decline" thereafter (p. 270). The authors concluded that to reach a "native-like level of proficiency in English [grammar]" the slow process of L2 learning must "start by 10-12 years of age" (p. 270). An early start is needed, according to the authors, because progress must begin well before a "sharp drop in learning rate [occurs] at about 17-18 years of age" (p. 270).

The conclusion that a CP for L2 learning closes at around 17-18 years of age might be questioned for several reasons. First, the test used by Hartshorne et al. (2018) to evaluate knowledge of English morphosyntax was an exclusively written instrument. It is safe to assume that all respondents to the internet grammar quiz could read English but not that all of them could speak English. Stevens (1999) found that the proportion of immigrants to the US who were unable to speak English increased steadily between the ages of 5 and 60 years.

Second, the Hartshorne et al. (2018) findings for immigrants (i.e., immersion learners) differed substantially from the findings obtained in another study examining a large number of immigrants. Hakuta et al. (2003) analyzed data obtained in the 1990 US Census for 2.02 million immigrants from a Spanish language background and 0.32 million immigrants from a Chinese language background. Respondents to the US census had been asked a series of questions regarding their proficiency in English. Their responses were used to construct a 4-point English proficiency scale for each respondent; these scores were then modelled. The functions obtained by Hakuta et al. (2003) showed no evidence of a discontinuity at the age of 12 or 15 years nor a discontinuity at any other age of immigration. Instead, self-rated proficiency in English was observed to decline gradually from ages of arrival that ranged from 5 to 60 years for both linguistic groups. The authors attributed this slow decline to normal cognitive aging over the lifespan.

Hartshorne et al. (2018, p. 269) questioned the validity of data obtained from the US census because these data were based on 4-point rating scales. However Hakuta et al. (2003, p. 32) cited validation research

which yielded moderate correlations of  $r=.52$  and  $.54$  with the English proficiency ratings they analyzed. Moreover, the findings of Flege et al. (1999) indicated that immigrants have a realistic understanding of their own L2 competence. The 240 Koreans rated their own proficiency in English using 5-point rating scales. Their self-ratings of English pronunciation correlated strongly with native English listeners' ratings of their pronunciation of English, and the Koreans' self-ratings of English grammatical knowledge correlated with the scores they obtained on the 144-item GJT described earlier ( $r=.64, p<.05$ ).

The two large-sample studies just cited led to very different patterns of data, and so to different interpretations of Early vs Late differences. The between-study differences may have arisen, at least in part, from sampling procedures. A general problem for L2 acquisition research is that participants are not randomly selected, leading to interpretive difficulties. For example, studies in which relatively well-educated participants are recruited on or near a university campus (e.g., Johnson & Newport, 1987; Flege et al., 1999) may not generalize to the population of persons around the world who learn English as a second or foreign language.

A potentially more serious problem arises for the analysis of samples that are systematically biased or skewed. For example, the census data analyzed by Hakuta et al. (2003) did not include respondents who reported no longer using their L1 while living in the US. This eliminated respondents who were likely to have lived longer than average in the US, and so to have achieved greater than average proficiency in English than respondents who continued to use their L1 and whose proficiency ratings were available for analysis (Stevens, 2004, p. 215). The possibility exists, therefore, that selection bias contributed to the absence of support for a CP in the Hakuta et al. (2003) study.

One wonders, in the same vein, if selection bias contributed to the strong support for a CP obtained by Hartshorne et al. (2018). The individuals who provided data for this innovative internet study were not, in fact, selected. They volunteered to participate after having learned about the grammar quiz, usually via social media. The authors selected a subset of respondents for the analysis mentioned earlier after having excluded respondents who provided obviously spurious data. The respondents retained for the analysis were nevertheless likely to have had a greater than average interest in language and language learning than the population of humans around the world who learn English as a foreign or second language. The influence of this selection bias, if indeed one existed, is unknown.

### **5.3 The crucial role of input**

The conclusion by Hartshorne et al. (2018) that a critical period for the learning of L2 morphosyntax closes at 17-18 years of age is difficult to evaluate in the absence of information regarding input. The authors did not ask respondents to indicate how often, with whom, in what contexts, or for what purposes(s) they used English. Use of the internet to obtain L2 acquisition data is an important step forward and is likely to become common. Such research will need to provide information regarding input, however, to yield interpretable results.

The need for input data is evident from inspection of individual data obtained by Hartshorne et al. (2018), which can be downloaded from <https://osf.io/pyb8s/wiki/home/>. Respondents who were speakers of the same L1 and reported having begun to learn English at the same age sometimes responded differently when asked to indicate their primary language(s). Many indicated having just one primary language, their native language. However others indicated that both their native language and English were primary languages, and others still indicated that English was their only primary language at the time of test.

Self-reported primary language(s) was not used in statistical modeling by Hartshorne et al. (2018), but one naturally wonders what accounted for such variation. The respondents who did/did not report English to be a primary language may have differed in one or more important ways, for example: living on a long-term basis with a native speaker(s) of English, being required to use English at work, or having the desire or need to use English for important social, recreational or religious activities.

The belief that a critical period (CP) exists for L2 learning has motivated a large amount of L2 research. Very little of this research, however, has aimed to determine if a CP actually exists. Most researchers simply assume the existence of a CP because of the ubiquitous presence of Early vs Late differences and simply seek to determine when and how rapidly the CP closes. Another aim of CP-inspired research has been to identify the underlying cause(s) of a loss or diminution of learning capacity that is hypothesized to occur after closure of the CP. This research, unfortunately, has met with scant success. To take an early example: Lenneberg (1967) thought that a CP for L2 speech learning closes when hemispheric specialization for language functions reaches completion. This proposed mechanism was soon discarded when evidence contradicting it emerged (e.g., Krashen, 1973).

Hartshorne et al. (2018, p. 263) defined the critical period for L2 learning as a “theory-neutral descriptor of diminished achievement, whatever its cause”. These authors culled seven possible causes for diminished achievement from the L2 acquisition literature. It has been hypothesized that, in comparison to Late learners, Early learners may: (1) have greater neural plasticity; (2) have the opportunity to learn the L2 over a longer period of time; (3) have less cognitive processing ability, which prevents them from being “distracted by irrelevant information”; (4) experience less “interference” from previously learned L1 structures; (5) show a “greater willingness to experiment and make errors”; (6) feel a “greater desire” to conform to peers; or (7) be more likely to immerse themselves in a “community of native speakers” (2018, p. 263).

It is noteworthy that four of the seven items on this list of potential causes of diminished achievement (viz., 2, 5, 6, 7) relate in some way to input. For example, someone who has a relatively strong desire to speak an L2 like his/her peers presumably spends time with those peers, which raises the question of “Who are they? Are they native or non-native speakers of the L2?” In a similar vein, one might ask “Who are the L2 learners who decide to immerse themselves in a community of L2 native speakers?” According to Stevens (1999, p. 574) learning an L2 at any age “requires exposure to the language, motivation, and opportunities to practice receptive and active skills. In short, language learning requires communicative and social interactions.” Moyers (2009, p. 161) argues that age of first exposure to an L2 “leaves much to be desired as an explanation for what is a very complex endeavor – one that is, by its nature, grounded in a social framework”.

## **6. Conclusions**

Evidence reviewed in this chapter indicated that the quantity and quality of L2 input that learners receive influence the long-term learning of L2 speech and at least some aspects of L2 morphosyntax learning. The age at which immigrants are first exposed to an L2 conditions the quantity and quality of L2 input they are likely to receive over the course of their lives in the host country. Early arrivals usually receive more and better L2 input than do immigrants who arrive later in life. This, in my opinion, is the primary basis for the Early vs Late differences that have been widely reported in the literature examining second-language acquisition.

My conclusion may surprise some readers given that input has tended to be ignored or downplayed in L2 research (Flege, 2009). The absence of attention to input seems to have derived, at least in part, from the mistaken belief that LOR provides an adequate measure of input. Another explanation is that many researchers believe that obtaining adequate measures of input is impossible. However the technology needed to obtain precise measures of the quantity and the quality of input now exists (Flege & Wayland, 2019). For those who are unwilling or unable to use this admittedly expensive and time-consuming approach, a simple paper and pencil instrument might be used to good advantage to assess L2 input.

As illustrated in Fig. 19, participants could be asked to indicate two percentages of language use, one for their L1 and the other for their L2, in a wide range of specific social contexts. When taken together, the responses would serve to define the participants' everyday language use patterns.

(The two percentages would necessarily sum to 100% in research with bilinguals.) In the final four items of such an instrument, participants would be asked to indicate their overall percentage use of the L1 and L2 in their first, third, and fifth years of residence in the host country, and also their overall language use in the year preceding the test. The purpose of these last four items is to evaluate possible changes over time in L2 use.

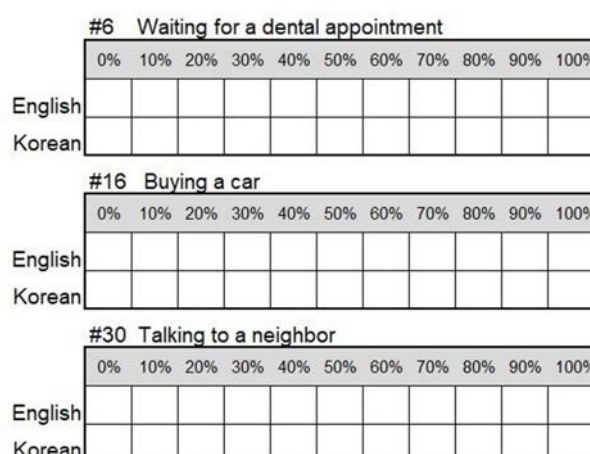


Figure 19. Illustration of an instrument for assessing language use in social contexts. Responses for the L1 and L2 must total 100%.

In summary, the pattern of data reviewed in this chapter leads me to conclude that long-term success in L2 learning is not limited by the closure of a critical period resulting in either the loss or the diminution of L2 learning capacity. Instead, degree of long-term achievement in L2 learning is determined probabilistically by a *non-critical period* that is defined primarily by age-related variation in the L2 input that learners normally receive owing to their differing motivations to learn the L2 and exposure to the L2 in differing social contexts.



Input is crucial for the learning of L2 speech and some aspects of L2 morphosyntax, just as it is for other socially defined human activities. Input is unlikely, of course, to be the only factor that influences long-term success in L2 learning. Other factors, for example, language learning aptitude (e.g., Abrahamsson & Hyltenstam, 2008), might also be found to influence long-term success in L2 learning. However, the role of these other factors cannot be understood via research designs that fail to control for variation in input.

The hypothesis that the closure of a critical period for L2 learning limits long-term success in L2 learning (Lenneberg, 1967) has inspired a large amount of research examining both L2 speech and morphosyntax. However the widespread appeal of the CP hypothesis has also impeded progress in L2 research focusing on age-related effects by encouraging investigators to ignore or downplay the crucial role of input. The time has come for the establishment of a new paradigm based on research designs favoring the evaluation of input and other factors that influence long-term success in L2 learning. The participants in future research with immigrants, in particular, should be selected on the basis of the input they have received at the time of test rather than on the basis of their presumptive stage of neurocognitive maturation years earlier at the time of immigration.

### **Acknowledgments**

This research was supported by grants from the National Institute of Deafness and Other Communicative Disorders to the University of Alabama at Birmingham. I thank Ocke-Schwen Bohn, Anders Højen, Thorsten Piske, Anja Steinlen, and Tullia Trevisan for comments on an earlier version.

### **References**

- Abrahamsson, J., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, 30(4), 481-509.
- Asher, J., & Garcia, R. (1969). The optimal age to learn a foreign language. *The Modern Language Journal*, 55(5), 334-341.
- Bahrack, H., Hall, L., Goggin, J., Bahrack, L., & Berger, S. (1994). Fifty years of language maintenance and language dominance in bilingual Hispanic immigrants. *Journal of Experimental Psychology: General*, 123(3), 264-283.

- Beck, M.-L. (1997). Regular verbs, past tense and frequency. Tracking down a potential source of NS/NNS competence differences. *Second Language Research*, 13, 93-115.
- Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-206). Timonium, MD: York Press.
- Birdsong, D. (2013). Age and the end state of second language acquisition. In W. Ritchie & T. Bhatia (Eds.) *The new handbook of second language acquisition* (pp. 401-424). Bingley, England: Emerald Group Publishing.
- Birdsong, D., & Birdsong, D. (2001). Regular-irregular dissociations in L2 acquisition of English morphology. *BUCLD 25: Proceedings of the 25<sup>th</sup> Annual Boston University Conference on Language Development* (pp. 123-132). Boston, MA: Cascadilla Press.
- DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499-533.
- Flege, J. (1987). A critical period for L2 learning to pronounce foreign languages? *Applied Linguistics*, 8, 162-177.
- Flege, J. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229-273). Timonium, MD: York Press.
- Flege, J. (2009). Give input a chance! In Piske & Young-Scholten (Eds.) *Input Matters in SLA* (pp. 175-190). Bristol: Multilingual Matters.
- Flege, J. (2018). It's input that matters most, not age. *Bilingualism: Language and Cognition*. <https://doi.org/10.1017/S136672891800010X>, Published online 15 May 2018
- Flege, J., Frieda, E., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, 25, 169-186.
- Flege, J., Frieda, E., Walley, A., & Randazza, L. (1998). Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition*, 20, 155-187.
- Flege, J., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, 23, 527-552.
- Flege, J., & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26, 1-34.
- Flege, J., & MacKay, I. (2011). What accounts for "age" effects on overall degree of foreign accent? In M. Wrembel, M. Kul, & K. Dziubalska-Kořaczyk (Eds.), *Achievements and perspectives in the acquisition of second language speech. New Sounds 2010*, Vol. 2 (pp. 65-82). Bern: Peter Lang.
- Flege, J., & Munro, M. (1994). The word unit in second language speech production and perception. *Studies in Second Language Acquisition*, 16, 381-411.
- Flege, J., Munro, M., & MacKay, I. (1995a). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125-3134.

- Flege, J., Munro, M., & MacKay, I. (1995b). Effects of age of second-language learning on the production of English consonants. *Speech Communication, 16*, 1-26.
- Flege, J., & R. Wayland (2019). The role of input in native Spanish Late learners' production and perception of English phonetic segments. *Journal of Second Language Studies, 2*(1). In press.
- Geary, D. (1998). *Male, female. The evolution of human sex differences*. Washington, D.C.: American Psychological Association.
- Hakuta, K., Bialystok, E. & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science, 14*(1), 31-38.
- Hartshorne, J., Tenenbaum, J., & Pinker, S. (2018) A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition, 177*, 263-277.
- Hopp, H., & Schmid, M. (2013). Perceived foreign accent in first language attrition and second language acquisition. *Applied Psycholinguistics, 34*, 361-394.
- Krashen, S. (1973). Lateralization, language learning, and the critical period: Some new evidence. *Language Learning 32*(3), 63-74.
- Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.
- Long, M., (1990) Maturation constraints on language development. *Studies in Second Language Acquisition, 12*, 251-285.
- MacWhinney, B. (1992). Transfer and competition in second language learning. In R. Harris (Ed.) *Cognitive processing in bilinguals* (pp. 371-390). Amsterdam: North Holland.
- Moyer, A. (2009). Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment. In T. Piske & M. Young-Scholten, M. (Eds.), *Input matters in SLA* (pp. 159-174). Bristol: Multilingual Matters.
- Oyama, S. (1973). *A sensitive period for the acquisition of a second language*. Unpublished Harvard University Ph.D. thesis.
- MacKay, I., Meador, D., & Flege, J. (2001). The identification of English consonants by native speakers of Italian. *Phonetica, 58*, 103-125.
- Patkowski, M. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning, 30*, 449-472.
- Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics, 11*, 73-89,
- Pinker, S., & Prince, A. (1991). Regular and irregular morphology and the psychological status of rules of grammar. In Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on the Grammar of Event Structure, pp. 230-251. DOI: <http://dx.doi.org/10.3765/bls.v17i0.1624>.
- Prasada, S., & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes, 8*(1) , 1-56.

- Piske, T., Flege, J., MacKay, I., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica*, 59, 49-71.
- Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191-215.
- Piske, T., & Young-Scholten, M. (2009). *Input matters in SLA*. Bristol: Multilingual Matters.
- Scovel, T. (1988). *A Time to Speak: A Psycholinguistic inquiry into the critical period for human speech*. New York: Newbury House Publishers.
- Stevens, G. (1999). Age at immigration and second language proficiency among foreign-born adults. *Language in Society*, 28, 555-578.
- Stevens, G. (2004). Using census data to test the critical-period hypothesis for second-language acquisition. *Psychological Science*, 15(3), 215-216.
- Yeni-Komshian, G., Flege, J., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition*, 3(2), 131-149.
- Zatorre, R. (1989). On the representation of multiple languages in the brain: Old problems and new directions. *Brain & Language*, 36, 127-147.



## Improvement in Young Adults' Second-language Pronunciation after Short-term Immersion

Anders Højen  
Aarhus University

### Abstract

This study examined the effect of short-term immersion in English-language communities in England on young native Danish adults' English pronunciation. Pronunciation ratings by a group of native judges revealed significantly higher pronunciation ratings when compared before and after 3-10 months of English immersion. A native Danish control group received virtually identical ratings by the judges at two different time points. The pronunciation gain score for the immersion group was significantly correlated with length of residence (LOR) in England. However, a stronger correlation ( $r=0.81$ ) was found between pronunciation gain score and a weighted input measure, *viz.* LOR weighted by self-reported proportion of English vs. Danish use during the immersion period. The results suggest that second-language (L2) learners' phonetic system is highly malleable and responds readily to new L2 input.

### 1. Introduction

Bilinguals typically speak their second language (L2) with a foreign accent. The age of L2 learning (AOL) is almost always found to be the strongest predictor of the degree of foreign accent, although a range of other factors – some of them often confounded with AOL – may also influence degree for foreign accent in the L2 (for a review, see Piske, MacKay, & Flege, 2001). The purpose of the present small-scale study was to examine the effect of experiential factors during young adults' short-term immersion

---

Anne Mette Nyvad, Michaela Hejrná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 543-559). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.

in an L2 community. The specific interest was in the immediate effect of the duration of L2 immersion experience and the degree of L2 use on L2 pronunciation.

AOL effects on L2 acquisition have been attributed to the passing of a critical period, after which it is no longer possible to make use of language input to build linguistic representations (Lenneberg, 1967). On the critical period account, biological or maturational changes specific to language underlie these difficulties, and the changes are often proposed to happen around the end of childhood although a number of different cut-off ages have been proposed (for an overview, see e.g., Singleton, 2005). In addition different critical periods have been proposed for different language domains (e.g., Granena & Long, 2013).

On other accounts, there is no biologically or maturationally defined endpoint that marks a categorical difference in the way an L2 is acquired. Rather, AOL effects on acquisition are assumed to arise from the state of development of the L1 when the L2 is acquired (e.g., Ellis, 2002; Flege, 1995; Hernandez, Li, & MacWhinney, 2005; MacWhinney, 2016). Moreover, these accounts generally assume that linguistic representations for L2 perception and production are acquired and entrenched via a high frequency of use and practice, just like other complex cognitive skills, using general cognitive processing mechanisms. The role of frequency for language acquisition is summarized thus by Gries and Ellis (2015, p. 230): “The most fundamental factor that drives learning is the frequency of repetition in usage. This determines whether learners are likely to experience a construction and, if so, how strongly it is entrenched, accessible, and its processing automatized.” Consistent with this notion, input frequency effects have been observed in the acquisition of L1 skills (e.g., Hart & Risley, 1995; Hurtado, Marchman, & Fernald, 2008; Weisleder & Fernald, 2013) as well as L2 skills (Højen & Flege, 2006; MacKay & Flege, 2004; Piske et al., 2001; Suter, 1976). Also consistent with the importance of input, effects of length of L2 use have been shown to influence L2 proficiency (Flege & Liu, 2001; Flege, MacKay, & Piske, 2002). In addition, the effect of duration of L2 experience has been found to be moderated by intensity of L2 input (Flege & Liu, 2001).

However, in their large-scale study of foreign accent in Italian immigrants who learned English as an L2 in Canada, Flege, Munro, and MacKay (1995), found no effect of the immigrants’ length of residence (LOR), which was a measure of the duration of their L2 experience. Flege et al. (1995) suggested that this was due to ceiling effects; all the

immigrants had lived in Canada for decades. Indeed, Flege & Fletcher (1992) had already suggested that most improvements in L2 pronunciation occur during the first year of exposure to native-produced L2 input in an L2 environment. In spite of this, few studies have examined L2 pronunciation changes occurring in the early phase of L2 input in an L2 environment. Previous studies examining the effect of L2 experience on degree of foreign accent have typically compared groups who had many years of L2 experience with groups who had about 6-12 months of L2 experience (e.g., Flege, 1991; Flege et al., 1995; Piske et al., 2001; Thompson, 1991; Yamada, 1995). Therefore, those studies were not designed to assess any immediate effects of experience in the initial phase of L2 exposure.

One study that examined L2 pronunciation development in the very early phase was that of Snow and Hoefnagel-Höhle (1977). They examined pronunciation of Dutch words by native English children and adults who had moved to the Netherlands. The participants were tested three times during their first 10-11 months of learning, and a significant improvement in the pronunciation of Dutch words was found for both child and adult learners. However, it is unclear how closely their first time of test coincided with the onset of native Dutch input. The participants were tested within six weeks after they started to speak Dutch (Snow & Hoefnagel-Höhle, 1977, p. 361), but presumably six months after they moved to the Netherlands (Snow & Hoefnagel-Höhle, 1978, p. 1115).

When examining the effect of native-produced L2 input on foreign accent, participants may differ according to the extent of nonnative L2 input. In the study by Snow and Hoefnagel-Höhle (1977), the participants had not learned Dutch in school before immigrating, whereas individuals immigrating to an English speaking country have often learned English in school to various degrees. In Denmark, students begin learning English after a few years in school and often become relatively proficient English speakers. However, the typical student only receives sporadic authentic input in inter-personal communication, because most teachers are non-native speakers of English. At the same time, English is very much present on various media platforms. As noted, Flege and Fletcher (1992) suggested that pronunciation typically improves during the first year of *authentic* input. However it is not clear how authentic input, e.g., via immersion in an L2 community, affects pronunciation of L2 speakers who are already relatively skilled. This would be the scenario for Danish students, who learned English in school from non-native teachers and heard native English on various media platforms. Therefore, the present study examined



the effect of short-term immersion in England on degree of foreign accent in native Danish young adults, who had learned English in school to a relatively high degree of receptive and expressive proficiency.

A foreign accent is manifested as the realization of phones in the L2 in a different way than native speakers typically do. In addition, L2 speech may differ on prosodic dimensions. Such deviances from the native phonetic norms can be identified using acoustic analysis. However, the above-mentioned studies of foreign accent generally examined degree of foreign accent using pronunciation ratings by native listener judges, which has been shown to be a reliable metric. Piske et al. (2001) reported a strong correlation between the pronunciation scores assigned to a set of sentences by two different groups of native judges on two different rating scales. This led Piske et al. to conclude that foreign accents can be scaled reliably by native listener judges. In addition, judges give highly similar pronunciation ratings across different sentences (e.g., Flege, 1988; Flege & Fletcher, 1992; Flege et al., 1995), suggesting that native listeners can identify and reliably scale a foreign accent based on a short speech sample.

## **2. Methods**

### **1.1 Participants**

Thirty female talkers participated in the study (which was part of a PhD dissertation based on a series of L2 speech perception and production experiments, Højen, 2003). Only females were recruited because the original intention was to examine only au pairs (who are mostly female). However, exchange students were added to the sample because the number of au pairs that could be recruited was insufficient. To keep the sample relatively homogenous, only exchange students who were females were recruited. The participants were assigned to three different groups. The Experience Group consisted of native Danish au pairs or exchange students ( $N=14$ ) who spent 3-11 months in England (the LOR). The No-experience Group ( $N=11$ ) served as an age-matched native Danish control group. The Native English Group ( $N=5$ ) consisted of native English speakers and served as a native English reference group. All participants reported normal hearing.

The native Danish participants had grown up in Denmark with native Danish parents. They had learned English in school for 7-10 years, but none of them ever had a native English teacher. Ten of the native Danish participants had never lived in an English speaking environment, and could

thus be said to be phonetically inexperienced with English with respect to everyday communication. However, four members of the Experience Group had previously lived in an English speaking environment for up to 12 months; this happened at a minimum age of 18 years. Participant characteristics are shown in Table 1.

	<i>N</i>	Age	T1 Prof	T1 Exp	LOR	EngUse
Experience Group	14	21.3 (3.4)	4.1 (0.8)	3.0 (5.0)	7.1 (3.2)	3.9 (0.8)
No-Experience Group	11	20.1 (2.4)	3.9 (0.8)	0.0 (0.0)	–	–
Native English Group	5	20.0 (2.2)	–	–	–	–

Table 1. Participant characteristics (*SD*). *Age*: years of age at Time 1. *T1 Prof*: Self-reported ability to speak English at Time 1 (1 = a little, 5 = very well). *T1 Exp*: Months of English-language experience at Time 1. *LOR*: Length of Residence during stay in England (in months). *EngUse*: Self-reported proportion of Danish and English use during their stay in England (1 = Danish only, 5 = English only).

The participants were informed about the purpose of the study, namely to examine effect of immersion on L2 speech perception and speech production. The participants in the Experience Group were tested in Denmark before and after their stay in England. The participants in the No-experience Group were tested two times with an interval of one week to 5 months. The No-experience Group stayed in Denmark during the interval between the two times of testing. The Native English Group was tested only once.

An additional three participants in the Experience group were tested before immersion but could not be tested after their stay; one participant returned to Denmark already after one month, and two participants did not return to Denmark at the time of retests. In addition, one more participant in the Native English Group was tested but recording failed.

## **1.2 Speech materials**

Previous research found that listener judges gave similar pronunciation ratings across different sentences (e.g., Flege, 1988; Flege & Fletcher, 1992; Flege et al., 1995). Therefore, to minimize the burden on the judges it was decided to base the listener judgments on just one sentence. The sentence 3. *Are “shock” and “hot” words?* was used to obtain listener ratings of foreign accent (the number 3 was read out and included in the sentence that was rated). The sentence was pragmatically odd because

it originally served to elicit specific speech sounds for acoustic analysis along with 11 other similar sentences. The specific sentence chosen for the present purpose was chosen because it contains several speech sounds with no direct Danish counterpart, namely [θ ɹ ʃ ɒ z] as well as syllable final [d] and syllable initial [w]. For the participants in the Experience Group and the No-experience Group, a production of the sentence was recorded at Time 1 and once again at Time 2. The talkers in the Native English Group were only tested once; therefore, two physically different repetitions of the sentence were recorded in one session.

### **1.3 Procedure**

Recordings of the target sentences by different talkers were compiled in a block of 60 sentences containing the Time 1 and Time 2 recordings from each native Danish speaker and the two single-session recordings from the native English speakers. The order of sentences was quasi-randomized. However, no two repetitions of the target sentence by the same participant were allowed to follow one another. The block was presented twice to ten 20-43-year-old listener judges, who were native speakers of British English. Because the block of sentences was presented twice, rating consistency within each judge could be assessed. Each of the 10 judges rated 120 sentences (30 participants × 2 sentence tokens × 2 blocks) for a total of 1200 judgments.

The judges heard the sentences over headphones and rated each sentence on a 100 point scale which was labeled "Strong foreign accent" at one end (corresponding to a rating of 1), and "No foreign accent" at the other end (a rating of 100). The judges were instructed to use the whole scale and to give the maximum score if they were sure they were listening to a native English speaker, and to give the minimum score to the most accented talkers. The participants were given 10 practice sentences to familiarize themselves with the task and the range of foreign accents. These sentences were produced by the participants of each group but differed from the test sentences. The judges indicated their rating of each sentence using a scale on a computer screen. The software UAB-soft was used to present the sentences and store the ratings.

## **3. Results**

### **1.4 Rating consistency in judges**

Before addressing the effect of immersion on foreign accent ratings, the judges' rating consistency was examined. The two times the block of

sentences was rated by the judges will be called the *first round* and the *second round*. Figure 1, left panel, shows the mean ratings assigned to all sentences by each judge in the first vs. second round. As shown, some judges were stricter than others, but their general rating level was similar in the two rating rounds and did not differ significantly ( $t(9)=0.903$ ,  $p<0.39$ ,  $d=0.57$ ). Mean ratings given in the first vs. second round to each group were also similar (Experience group, 39 vs. 40; No-experience Group, 29 vs. 32; Native English Group, 93 vs. 93).

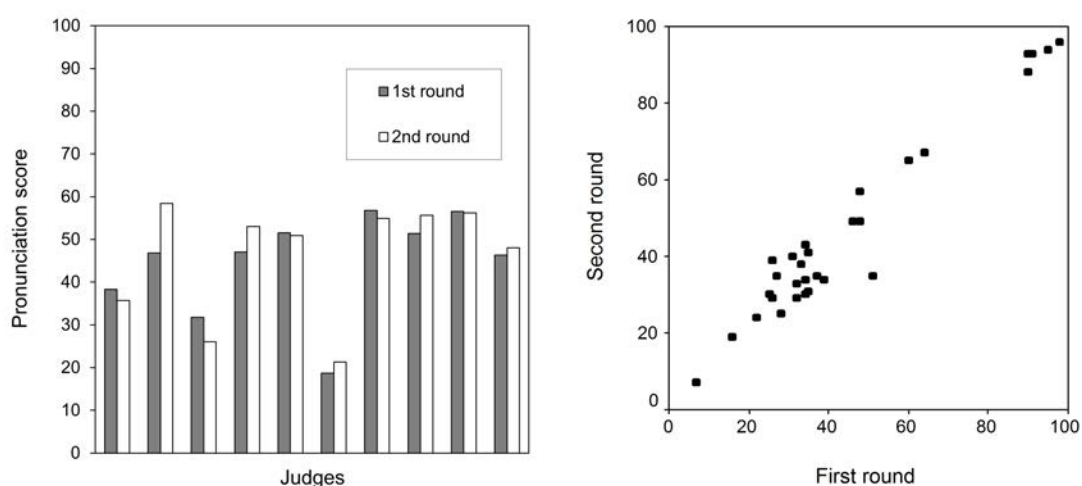


Figure 1. Left panel: The mean pronunciation score given on a 100-point scale by each judge to sentences in the first vs. the second round (not equivalent to Time 1 vs. Time 2). Right panel: The correlation between the pronunciation scores of each participant (averaged across test-time) at the first vs. the second rating round.

In order to test intra-judge consistency of rating of each participant, each judge's mean ratings given in the first vs. the second round were submitted to a Pearson correlation coefficient analysis. The first and second round judgments were strongly correlated ( $r(8)=0.93$ ,  $p<0.001$ ). This suggested that each judge was consistent in assigning scores. In addition, in order to test whether the judges as a group assigned the same pronunciation score to each participant in the two rating rounds, the talker-based mean scores were submitted to a Pearson correlation analysis. The correlation was highly significant ( $r(28)=0.94$ ,  $p<0.001$ ). A scatter plot of the talker-based round 1 and 2 correlation is shown in Figure 1, right panel. Note that, as expected, the five native English speakers, in the top right corner of the graph, consistently received very high scores, and that the variation among the native English speakers was similar in round 1 and round 2.

### 1.5 Pronunciation scores before and after immersion

The mean pronunciation scores given to each group at Time 1 and Time 2 are shown in Figure 2. The Native English Group received mean scores which approximated the maximum score of 100. This suggested that the judges successfully identified the native English speakers as speaking without a foreign accent, although the native speakers did not receive a perfect rating of 100. This suggests that on a few occasions, the judges were not completely sure that the speaker was a native speaker. This result aligns with previous studies of foreign accent rating (e.g., Flege et al., 1995; Yeni-Komshian, Flege, & Liu, 2000).

The Experience Group received higher scores at Time 2 than at Time 1. Before testing the significance of the difference, Figure 3 shows the distribution of scores in the three groups of participants. The figure shows that while the native English speakers (top panel) most often received near-maximum scores, the scores given to the No-experience Group (bottom panels) were skewed towards low scores at both Time 1 and Time 2.

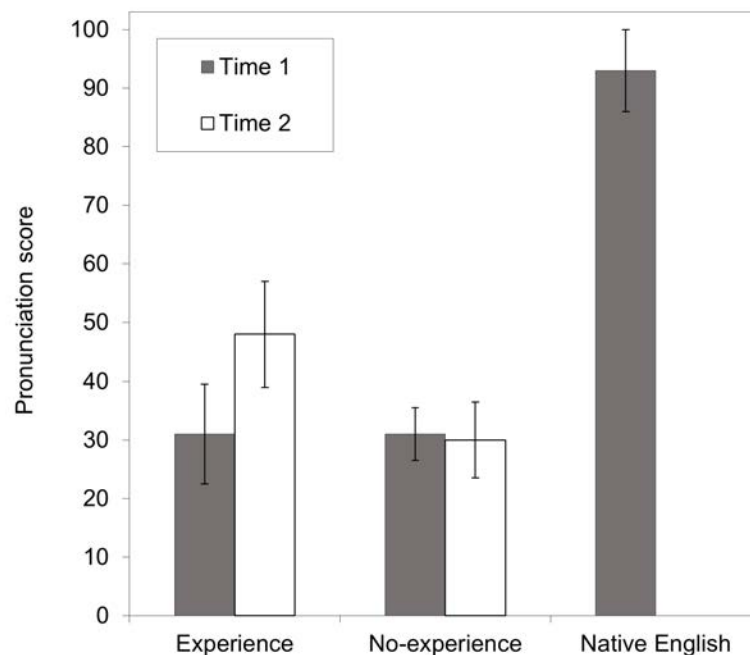


Figure 2. The mean pronunciation scores given to each group. A score of 100 indicates "no foreign accent; a score of 1 indicates "strong foreign accent". Error bars denote  $\pm 1$  SD.

However, the scores given to the Experience Group (mid panels) were less skewed towards low pronunciation scores at Time 2. Figure 3 provides important information about the pronunciation scores which are not apparent in Figure 2, namely that there is an appreciable spread of the scores. Notably, in each group of Danish speakers some sentences received high pronunciation scores, and with very few exceptions, all sentences produced by speakers in the Native English Group received very high scores.

The Native English Group was recruited mainly as a reference group, it was only tested once, and had an  $N$  of only 5. Therefore, only the scores of the Experience Group and the No-Experience Group were submitted to a two-way 2 (Group)  $\times$  2 (Test time) ANOVA with Test time as a repeated measure. The main effect of Group was nonsignificant ( $F(1, 23)=0.25, p=0.621, \eta^2=0.10$ ), the main effect of Test Time was significant ( $F(1, 23)=4.92, p=0.037, \eta^2=0.18$ ), and the Group  $\times$  Test time interaction was significant ( $F(1, 23)=6.30, p=0.020, \eta^2=0.22$ ). As expected, the source of the interaction was a significant simple effect of Test time for the Experience Group ( $t(13)=2.81, p=0.015, d=1.51$ ), but not the No-experience Group ( $t(10)=.42, p=0.684, d=0.27$ ), and a significant effect of Group at Time 2 ( $t(23)=2.81, p=0.010, d=1.13$ ) but not at Time 1 ( $t(23)<0.01, p=0.999, d<0.01$ ).

Whereas Figure 3 shows that the Experience Group (the middle panels) generally received more favorable pronunciation scores at Time 2 (right) than at Time 1, there were still many scores at the low end of the scale at Time 2. This indicates that some participants in the Experience Group did not improve their pronunciation during their stay in England. Individual scores for each of the 14 participants in the Experience Group at Time 1 and 2 are shown in Figure 4.

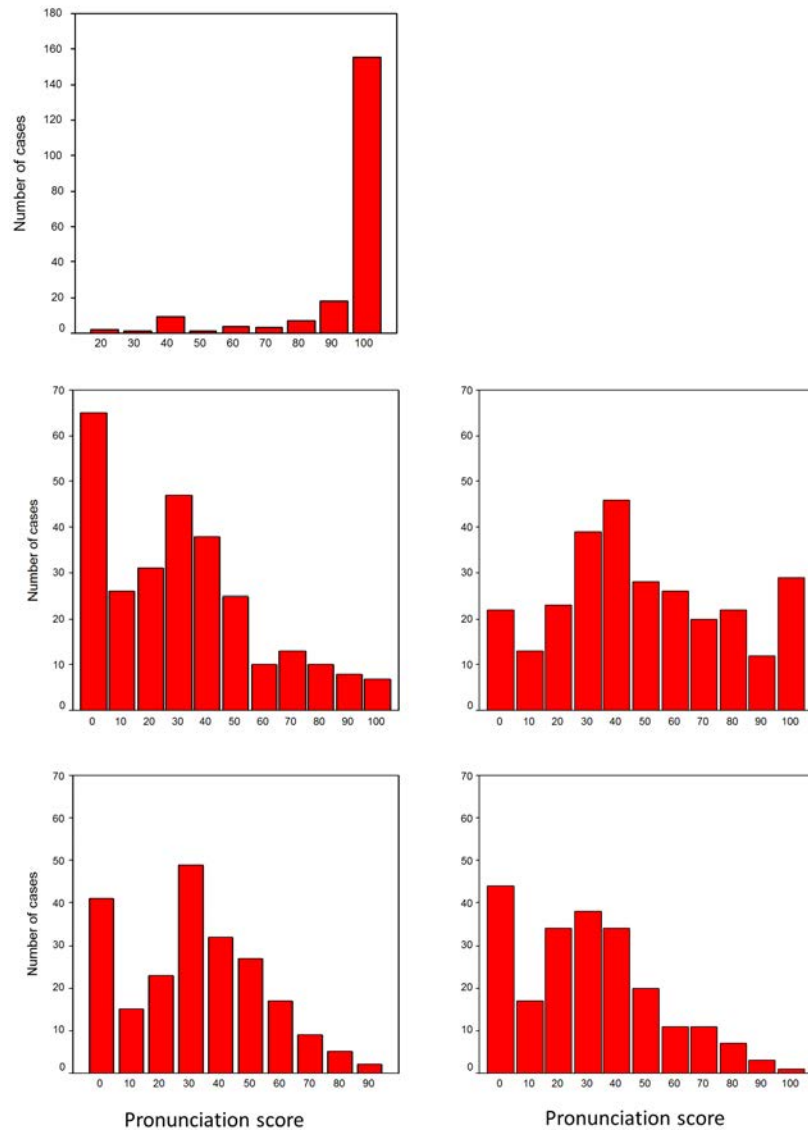


Figure 3. The frequency of pronunciation scores given to the Native English Group (top), the Experience Group (mid) and the No-experience Group (bottom). Time 1 results are shown in left panels, Time 2 results in right panels. The scores are bins of 10. Note that the frequency of scores is on different scales.

At Time 2, most (10 of 14) participants received higher scores than at Time 1; only 4 participants received lower scores. Recall that four of the participants had previous English immersion experience of between 9 and 12 months already at Time 1. These participants were number 1, 2, 5, and 14 in Figure 4. Participant 14 had a mediocre pronunciation score at Time 1, but received a high score at Time 2, which was comparable to the mean score of the Native English Group. However, the two participants who

showed the least progress – actually, they had nominally lower scores at Time 2 than Time 1 – were also two previously experienced participants. This suggests that amount of previous experience did not exert a uniform influence on pronunciation progress. However, participant 14 had an LOR in England of 11 months between Time 1 and Time 2, whereas participants 1 and 2, who also had previous immersion experience, had an LOR of only 3-4 months between Time 1 and Time 2. This suggests that LOR may also be an important factor in pronunciation score gain.

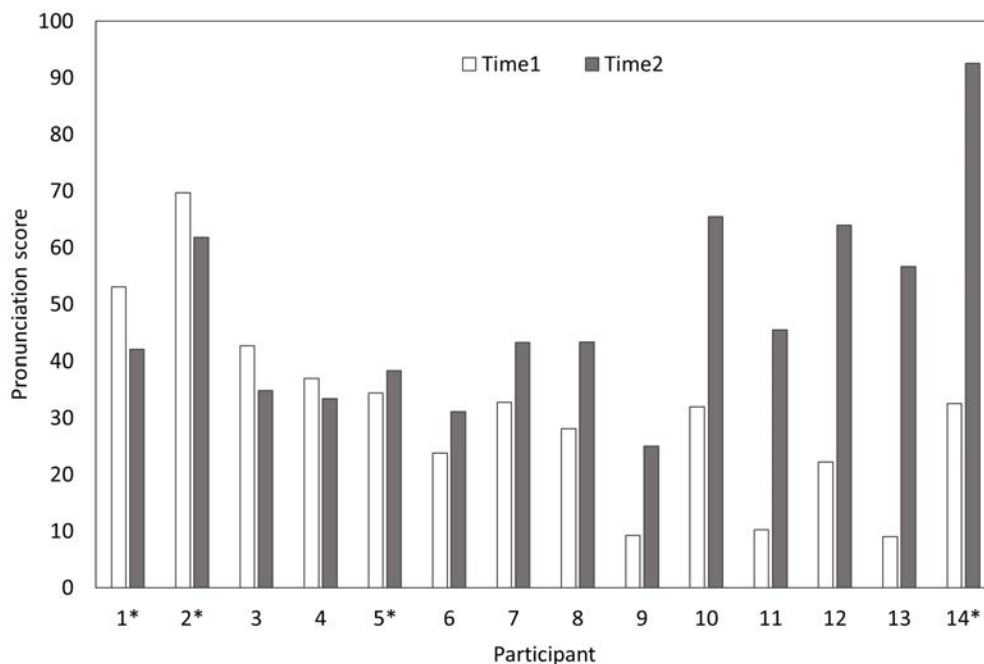


Figure 4. The mean pronunciation score assigned to each participant in the Experience Group at Time 1 and Time 2. Along the x-axis, the participants are ordered according to score increase from Time 1 to Time 2. Previously experienced participants are marked by a star.

To examine influences on participants' degree of improvement in pronunciation, a pronunciation gain score was derived by subtracting the Time 1 rating from the Time 2 rating. Figure 5, left panel, shows a scatter plot of the pronunciation gain score as a function of LOR for the 14 participants in Experience Group. In spite of some variation, the correlation between LOR and pronunciation gain score was significant ( $r(12)=0.61, p<0.022$ ). Note that no participant with an LOR of less than five months improved their pronunciation. Also note that three of the six participants with an LOR of 9-11 months showed little improvement. Why did they not?



As mentioned, Flege & Liu (2001) found that the LOR effect on L2 acquisition was modulated by the intensity of L2 input which the learners were likely to have had. For the present study, the participants in the Experience Group rated their use of Danish vs. English use in active interaction on a scale from 1 to 5 (1 = Danish only, 3 = equal use of Danish and English, 5 = English only). A weighted English-language input measure was derived by multiplying the LOR in months by self-reported proportion of English use. Although it is not known whether the relative importance of LOR and language use is reflected accurately in the weighted input score, it is likely to be a better measure of participants' total amount of L2 input during their stay in England than LOR or English use alone.

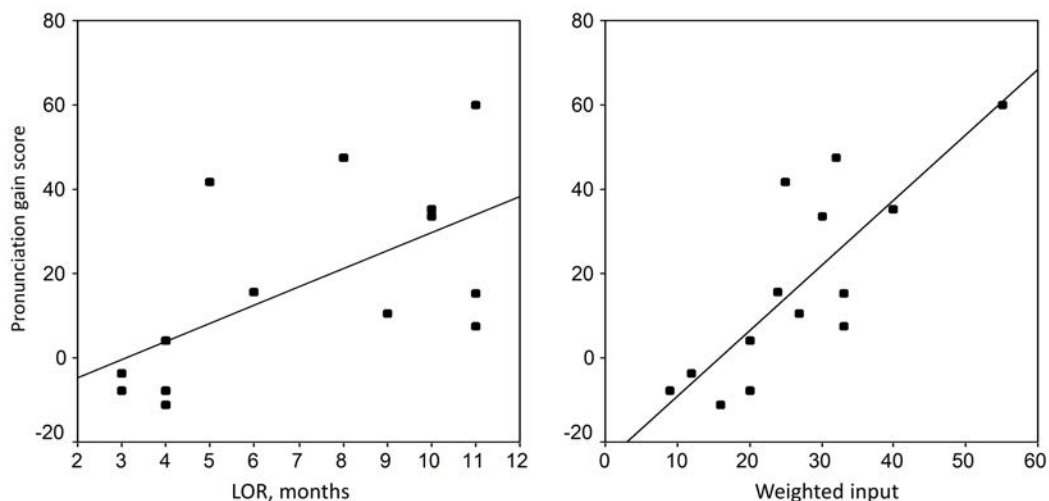


Figure 5. A scatter plot of pronunciation gain score as a function of LOR from Time 1 to Time 2 (left panel) and as a function of weighted input (LOR weighted by proportion English use, right panel).

Figure 5, right panel, shows a scatter plot of the pronunciation gain score as a function of weighted input. As shown, weighted input was quite successful at predicting pronunciation score gain, and the strong correlation between the two variables was significant ( $r(12)=0.81$ ,  $p<0.001$ ).

#### 4. Discussion

The purpose of this small-scale study was to examine the effect of short-term immersion in England on young native Danish adult females' pronunciation of English. The results showed a significant effect after an

average of just 7.1 months of immersion on native judges' pronunciation rating of a single sentence produced by the participants in the Experience Group before and after immersion. On the other hand, the No-experience Group received virtually identical mean ratings for their two productions of the same sentence at two different time points.

The improvement in pronunciation after immersion varied greatly in the Experience Group, such that some participants did not improve at all, whereas others went from the low end of the rating scale to the high end. Importantly, the pronunciation gain score was significantly correlated with LOR. The data suggest that an LOR of at least five months is needed for a detectable improvement in L2 pronunciation, but note that the low *N* means that this result should be interpreted with caution.

A stronger correlation was found between pronunciation gain score and a composite measure of input derived by a simple multiplication of LOR and self-rated degree of L2 use. These results indicate that even L2 learners who have learned an L2 as a foreign language in a school setting and spoken it for about 10 years, can improve their L2 pronunciation as a rather direct function of the amount of L2 input they receive during immersion in an L2 community. The strong correlation was likely to be due to selection of a highly motivated and relatively homogenous Experience Group, i.e. all females with similar ages who were self-selected for an interest in traveling abroad. The results support the suggestion by Flege & Liu (2001) that LOR may provide only a coarse measure of L2 input and that degree of L2 use moderates the effect of LOR (see also Flege, this volume).

The results suggest that the young adult L2 learners were able to perceive at least certain phonetic differences between their own pronunciation of English and the pronunciation of the English that they encountered during their immersion period. Moreover, the improved pronunciation after short-term immersion suggests that the organization of the phonetic system in L2 learners is malleable and responds readily to new input, allowing for an approximation to the native norm of the L2.

Some of the participants in the Experience Group did not receive higher accent scores at Time 2. Two of the participants who did not improve were previously experienced (9-12 months of English immersion in England and the United States, respectively). This might indicate that L2 learners do not improve L2 pronunciation much after the first year of immersion, as suggested by Flege & Fletcher (1992). However, the participant who improved the most also had 9 months of prior English immersion

(in British Columbia, Canada). This suggests that the drop in pronunciation scores of the two non-improvers was not simply explained as a slowing down of the rate of learning following their initial period of immersion experience.

The improvement in pronunciation scores in adult learners after short-time immersion in an L2 speaking environment and the improvements' close association with duration and intensity of L2 input suggest a quite malleable phonetic system underlying speech production. This finding runs counter to the critical period hypothesis, at least in its original formulation, which suggests that adult L2 learners cannot make automatic use of input and build L2 representations based merely on L2 exposure (Lenneberg, 1967). It is true that the general pattern of this study, and that of previous research (e.g., Flege et al., 1995; Yeni-Komshian et al., 2000), is that a foreign accent is extremely difficult to avoid for adult learners, and this is in accordance with more recent and less stringent formulations of the critical period hypothesis (e.g., DeKeyser & Larson-Hall, 2005; Long, 2005). The more recent formulations of the critical period hypothesis merely claim as evidence for the critical period hypothesis that the AOL function is not strictly linear across the lifespan (i.e., a *sensitive* rather than a strictly critical period).

However, as noted by Vanhove (2013), one problem with the multiple and watered down formulations of the critical period hypothesis is that it may in essence be impossible to falsify the hypothesis. But at the very least, it seems possible to state with certainty that a biologically or maturationally defined critical period does not suffice to explain bilinguals' deviances from (monolingual) native norms. This conclusion is supported by work showing deviances in the L2 of very early bilinguals, who should not have passed their critical period (e.g., Flege et al., 1995; Yeni-Komshian et al., 2000), and even in the L1 of early bilinguals (Ivanova & Costa, 2008; Yeni-Komshian et al., 2000) as well as late bilinguals (Ammerlaan, 1996; Pavlenko, 2000; Pavlenko & Jarvis, 2002; Pelc, 2001).

As mentioned in the introduction, other accounts of bilingual deviances stress the importance of L2 use and input as well as the interaction between the L1 and L2 systems, which may vary with age or state of entrenchment of the L1 system at the onset of L2 acquisition (e.g., Flege, 1995; MacWhinney, 2016). In addition, domain-general cognitive aging has been proposed to explain AOL effects on L2 acquisition (Hakuta, Bialystok, & Wiley, 2003). Probably all L2 researchers acknowledge the existence of use and interaction effects on L2 skills and perhaps also

cognitive aging effects. What seems to remain controversial is whether unexplained variance could or should be attributed to, as yet, unidentified language-specific biological/maturational changes during childhood.

In summary, the results of the present small-scale study suggest that L2 pronunciation improves in immersed adult L2 learners as a function of a measure of L2 input (LOR weighted by degree of L2 use). Even though most or all late bilinguals continued to speak with a foreign accent, the present findings also suggest that even late bilinguals possess a readily malleable phonetic system.

## References

- Ammerlaan, T. (1996). *'You get a bit wobbly...': exploring bilingual lexical retrieval processes in the context of first language attrition*. (PhD), PhD dissertation, Radboud University, Nijmegen. Retrieved from [http://repository.ubn.ru.nl/bitstream/handle/2066/146125/mmubn000001\\_215781937.pdf](http://repository.ubn.ru.nl/bitstream/handle/2066/146125/mmubn000001_215781937.pdf)
- DeKeyser, R. M., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism* (pp. 88-108). Oxford: Oxford University Press.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143-188.
- Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, 84(1), 70-79.
- Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, 89(1), 395-411.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.
- Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on the degree of perceived foreign accent. *Journal of the Acoustical Society of America*, 91(1), 370-389.
- Flege, J. E., & Liu, S. (2001). The effect of experience on adults acquisition of a second language. *Studies in Second Language Acquisition*, 23(4), 527-552.
- Flege, J. E., MacKay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, 23(4), 567-598.
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97(5), 3125-3134.
- Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language

- aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, 29(3), 311-343. doi:10.1177/0267658312461497
- Gries, S. T., & Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, 65(Suppl 1), 228-255. doi:10.1111/lang.12119
- Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical period hypothesis for second-language acquisition. *Psychological Science*, 14(1), 31-38.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes
- Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences*, 9(5), 220-225.
- Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, 11(6), F31-F39. doi:10.1111/j.1467-7687.2008.00768.x
- Højen, A. (2003). *Second-language speech perception and production in adult learners before and after short-term immersion*. (PhD dissertation), Aarhus University, Aarhus, Denmark.
- Højen, A., & Flege, J. E. (2006). Early learners' discrimination of second-language vowels. *Journal of the Acoustical Society of America*, 119(5), 3072-3084.
- Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, 127(2), 277-288. doi:10.1016/j.actpsy.2007.06.003
- Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.
- Long, M. H. (2005). Problems with supposed counter-evidence to the Critical Period Hypothesis. *International Review of Applied Linguistics*.
- MacKay, I. R. A., & Flege, J. E. (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. *Applied Psycholinguistics*, 25(3), 373-396.
- MacWhinney, B. (2016). Entrenchment in second language learning. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning* (pp. 343-366). New York: American Psychological Association.
- Pavlenko, A. (2000). L2 influence on L1 in late bilingualism. *Issues in Applied Linguistics*, 11(2).
- Pavlenko, A., & Jarvis, S. (2002). Bidirectional transfer. *Applied Linguistics*, 23(2), 190-214.
- Pelc, L. A. (2001). *L1 lexical, morphological and morphosyntactic attrition in Greek-English bilinguals*. PhD dissertation, City University of New York. Retrieved from [https://academicworks.cuny.edu/gc\\_etds/1782](https://academicworks.cuny.edu/gc_etds/1782)
- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, 29(2), 191-215.
- Singleton, D. (2005). The Critical Period Hypothesis: A coat of many colours *In-*

- ternational Review of Applied Linguistics in Language Teaching* (Vol. 43, pp. 269).
- Snow, C. E., & Hoefnagel-Höhle, M. (1977). Age differences in the pronunciation of foreign sounds. *Language and Speech*, 20(4), 357-365.
- Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, 49, 1114-1128.
- Suter, R. W. (1976). Predictors of pronunciation accuracy in a second language learning. *Language Learning*, 26, 233-253.
- Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, 41(2), 177-204.
- Vanhove, J. (2013). The critical period hypothesis in second language acquisition: A statistical critique and a reanalysis. *PloS One*, 8(7), e69172.
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143-2152.
- Yamada, R. A. (1995). Age and acquisition of second language speech sounds. Perception of American English /r/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 305-320). Timonium, MD: York Press.
- Yeni-Komshian, G. H., Flege, J. E., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition*, 3(2), 131-149.



## Understanding Vowel Perception Biases – It’s Time to Take a Meta-analytic Approach

Linda Polka & Yufang Ruan  
McGill University

Matthew Masapollo  
Boston University

### Abstract

This chapter reviews four recent studies designed to examine several theoretical accounts of directional asymmetries in vowel perception. The studies provide cross-language data on adults’ discrimination of vowels that fall within a given phonetic category. The results show that asymmetries emerge using unimodal acoustic and visual vowels, regardless of native language, and also using schematic non-speech visual analogs. We then integrate the data across these four studies in a mini-meta-analysis. Collectively, the findings provide strong support for the Natural Referent Vowel framework’s central claims that (1) asymmetries reflect a “language-universal” sensitivity to formant convergence (focalization) and 2) that this sensitivity is a speech-specific bias reflecting human sensitivity to the way that articulatory movements shape the acoustic and optical structures of speech. We advocate for further research adopting a meta-analytic approach.

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 561-582). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.



## 1. Introduction

In previous work we discovered that, in infants and adults, vowel discrimination is often asymmetric such that discriminating a vowel change in one direction is significantly easier compared to discriminating the same vowels in the reverse direction (Polka & Bohn, 2003; 2011). For example, infants were more accurate when discriminating a change from /ε/ to /ae/ compared to the reverse direction of change from /ae/ to /ε/. In infants, similar asymmetries are found across language groups showing that this pattern reveals a generic, universal bias rather than an effect of language-specific attunement or categorization. In adults, asymmetries have been observed for non-native and within-category vowel contrasts. Figure 1 (left panel) shows directional asymmetries that have been reported in the literature; the arrow connecting two vowels shows the direction in which discrimination of the vowel pair was significantly higher. Directional asymmetries follow a consistent pattern – the easier direction is the one in which the vowel to be detected (the B vowel in an AB sequence) is the more peripheral vowel within a standard articulatory/acoustic vowel space (F1/F2) vowel space. This suggests that perception favors vowels produced with more extreme vocal tract constrictions or configurations. The Natural Referent Vowel (NRV) framework was formulated to account for these findings and to guide research into the nature and significance of this perceptual bias (Polka & Bohn, 2011). According to NRV, directional asymmetries in vowel discrimination reveal a universal perceptual bias that is phonetically grounded in human capabilities for speech production and perception. This perceptual bias is posited to reflect our exquisite sensitivity to the way that articulatory movements shape the physical speech signal. Specifically, we propose that this bias is due to the increased salience of vowels produced with more extreme articulatory maneuvers, which give rise to well-defined spectral prominences in the acoustic speech signal due to formant frequency convergence, also known as *focalization*. The *focal vowel bias* is supported by cross-linguistic research on phonemic vowel contrasts (Polka & Bohn, 2011; Tsuji and Cristia, 2017).

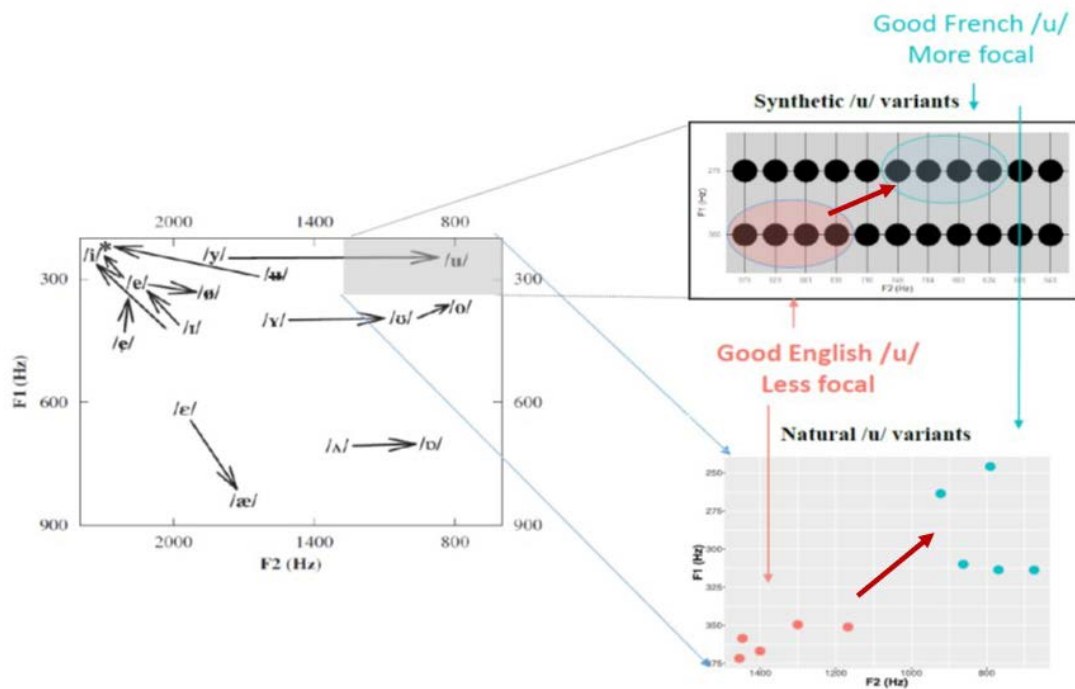


Figure 1.

**Left:** Directional asymmetries reported in the literature (from Polka & Bohn, 2011)

**Right top:** synthetic /u/ stimuli used in Masapollo et al JASA (2017): the red arrow shows the direction that was easier to discriminate for both French and English adults

**Right bottom:** natural /u/ stimuli used in Masapollo et al Cognition (2017); Masapollo et al JEP:HP& P (2018) and Masapollo et al JASA-EL (submitted); the red arrow shows the direction that was easier to discriminate for both French and English adults

The Native Language Magnet (NLM) model offers an alternative account of directional asymmetries in vowel discrimination (Kuhl et al, 2008). This model emerged from work investigating perception of within-category vowel variants. According to NLM, listening experience shapes perception to align with language-specific phonetic properties of native vowel categories. This leads to the formation of native language prototypes that act like perceptual magnets which attract less prototypic variants; NL magnets essentially warp the perceptual space around best or prototypic exemplars. One consequence of this magnet effect is asymmetric discrimination – detecting a change from a prototypic to a non-prototypic exemplar is harder compared to the reverse direction, i.e. a change from a non-prototypic to prototypic exemplar. Research focused on

vowel contrasts has aligned with the NRV predictions pointing to language universal biases, while research focused on within-category differences has aligned with NLM predictions pointing to language-specific processes. Thus, a more direct and systematic comparison of these predictions is needed.

Recently, we have made significant progress towards disentangling these alternative views and confirming several central claims of the NRV framework. In Masapollo, Polka, Molnar & Ménard (2017), we systematically examined the role of universal and language-specific factors in vowel discrimination asymmetries. To do so, we synthesized an array of vowels that fall within the /u/ category. This /u/ vowel array is shown in Figure 1 (top right panel). The variants systematically varied in the proximity between their F1 and F2 values, in equal psychophysical steps along the mel scale. Critically, these variants were all clearly categorized as /u/ by both English and French adults, but also varied such that the best /u/ exemplars in French (circled in blue) were more focal than the best /u/ exemplars in English (circled in pink). The difference in focalization was due in part to the greater lip-rounding and protrusion that occurs in production of French /u/ compared to English /u/, which also increases F1 and F2 convergence for French /u/ productions compared to English /u/ productions.

Adults performed a categorial AX discrimination task designed to assess whether they show an asymmetric pattern in their discrimination of more-focal/French /u/ and less-focal/English /u/ tokens. Both monolingual English and monolingual French adults showed asymmetric discrimination as predicted by the NRV framework – showing better discrimination for a change from a less focal/English /u/ to a more-focal/French /u/ compared to the reverse direction. It is important to note that the NLM predicts that discrimination would be asymmetric but in opposite directions for French and English perceivers. Specifically, within each language group discriminating the change from a poor to good /u/ exemplar was expected to be better compared to the reverse (good to poor) direction. However, both French and English adults showed an asymmetry in the same direction and magnitude; thus there was no evidence that this pattern was affected by language experience as proposed by NLM. These findings confirm that the NRV bias reflects a sensitivity to formant convergence and also firmly establishes the presence of universal vowel processing biases that are distinct from the effects of language-specific attunement or prototype categorization.

A follow-up study provided evidence that the focal vowel bias can be observed when perceiving natural speech. This was shown by using auditory-visual recordings of English /u/ and French /u/ produced by a simultaneous bilingual female talker (Masapollo, Polka & Ménard, 2017). The F1 and F2 measures for these natural auditory /u/ variants are shown in Figure 1 – bottom right panel. A static screen shot showing one French /u/ token and one English /u/ token (taken at vowel midpoint) is also presented in Figure 2. As these images show, French /u/ and English /u/ are visually distinct. Video analyses confirmed that the lip-rounding and protrusion differences between these /u/ variants are conveyed in the dynamic visemes of these vowels. The focal vowel bias predicted by NRV was replicated when adults discriminated the natural French /u/ and English /u/ tokens presented in an auditory vowel discrimination task. As with the synthetic stimuli, both French and English adults showed the same directional asymmetry, which did not interact with language experience. The same finding emerged when we tested French and English adults' discrimination of the French /u/ and English /u/ tokens in a visual-only condition. As well, the focal bias was observed when English adults were tested in a bimodal (audio-visual) condition in which the auditory and visual channels were phonetically congruent, but not in a bimodal condition in which the audio and visual channels were phonetically-incongruent (French auditory /u/ dubbed onto English visual /u/; English auditory /u/ dubbed onto French visual /u/). These findings supply further evidence that the NRV bias reflects a universal sensitivity to formant convergence, independent of native-language categorization processes. Importantly, the finding that the same pattern emerges in visual vowel processing provides strong support for the NRV claim that this bias is phonetically grounded, reflecting a sensitivity to articulatory information available across different perceptual modalities.

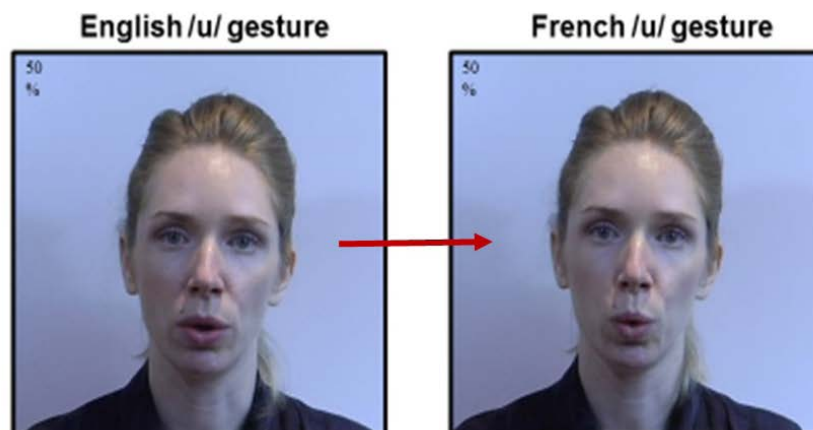


Figure 2. Model speaker's visual articulation at vowel midpoint. The red arrow shows the direction that was easier to discriminate for both French and English adults

In a subsequent study we probed the speech-specificity of the focal vowel bias in several ways (Masapollo, Polka, Ménard, Franklin, Tiede, & Morgan, 2018). First, we replicated the focal vowel bias in English adults using the same natural visual-only French /u/ and English /u/ stimuli while also tracking eye movements. Subjects attended selectively to the talker's mouth and also looked longer at the more focal /u/ tokens when discriminating these stimuli, confirming that articulatory features (increased lip rounding/protrusion) specifying French /u/ drew more attention to the talking mouth. In a second study, no asymmetry was observed when English adults were tested with still images of the model speaker's face at vowel midpoint (as in Figure 2) where significant differences in lip rounding are observed across the two vowel types. This finding lends further support to idea that the focal vowel bias is tied to dynamic articulatory information, which is absent in a static image.

We gained further insights by testing adult discrimination using non-speech visual analogs of the lip movements for each vowel type. One visual analog condition was a point-light movie of the lip movements for each vowel token created from the video recordings by tracking four dots, two placed at the corners of the mouth and two placed on the top lip and bottom lip at the mouth mid-line as illustrated in Figure 3 (right top panel). The moving dots provide information on lip shape and movements. Although the moving dots are not recognized as a mouth, the French point light movies track a larger and more dynamic change in lip aperture compared to English point light movies. The same directional asymmetry that we observed for natural auditory and visual vowel tokens was observed when adults discriminated these point light movies; this was the case when subjects were told that the dots track lip movements and when they were not provided this information. However, the asymmetry was much weaker and failed to reach significance when the point light movies were rotated counter-clockwise by 45 degrees; in this orientation the configuration of the dots convey the same lip movement patterns but no longer depict a mouth-like shape. The point light analog findings suggest that adults require both the lip shape and movement patterns of these vowels to elicit the focal vowel bias, but recognition of a moving mouth is not required. In a second visual analog condition (Figure 3 – right bottom panel) we replaced the dots with a sideways figure-8 ( $\infty$ ) shape (aka a Lissajou curve) that changed in width and height over time to track the lip movements of each vowel token. This visual analog conveyed the distinct kinematic

patterns present in lip movements over time for each vowel type but did not depict a mouth-like shape. Discrimination of the figure-8 analogs was not asymmetric providing further evidence that the focal vowel bias requires information specifying both lip shape and movement. Overall, the findings argue against an interpretation of the NRV bias as arising from simple, low level auditory or visual processes, and place the NRV bias squarely in the domain of speech perception.

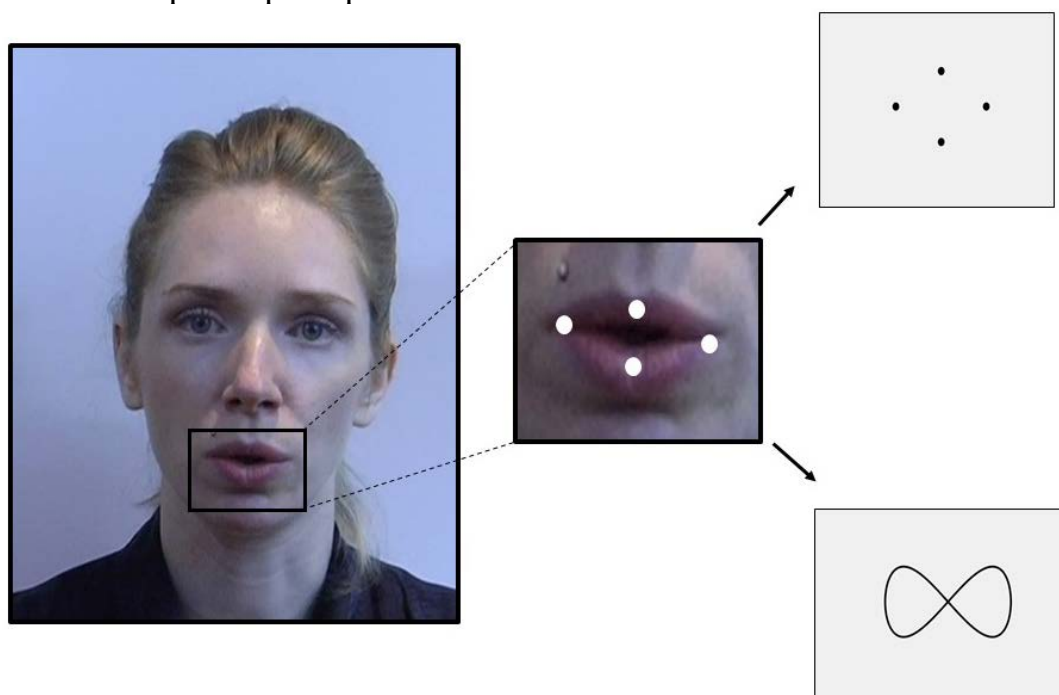


Figure 3. Dynamic non-speech visual analogs were created by tracking dots located at top/bottom and corners of the mouth. Point light movies (right top) conveyed lip shape and movement; Lissajou curves (right bottom) convey lip kinematics but not lip shape (Masapollo et al, 2018)

As a further test of the phonetic grounding of the NRV bias we examined task demands that impact phonetic processing (Masapollo, Franklin, Morgan, & Polka, submitted). In the work outlined above we used a categorical AX task with a 1500 ms inter-stimulus interval (ISI). This choice was based on prior work showing that phonetic processing is invoked by the memory demands imposed by a relatively long ISI. At a shorter ISI (e.g. 250 or 500 ms) perceivers can hold and compare acoustic details in auditory memory without engaging in phonetic encoding (e.g. Werker & Tees, 1983; Werker & Logan, 1985; Cowan & Morse, 1986). Auditory short memory fades quickly and thus when ISI is increased perceivers must rely on an encoded form of the stimulus to complete the task. Thus, prior work suggests

that auditory processing is engaged when the ISI is short and phonetic processing is invoked when the ISI is longer (e.g. 1000 ms or 1500 ms). Thus, if the NRV bias is a phonetic bias, it should be reduced or absent when the ISI is shortened creating memory demands that favor auditory processing. English adults tested with natural productions of English /u/ and French /u/ showed reliable directional asymmetries when the ISI was 1500 ms but not when the ISI was shortened to 1000 ms or 500 ms. This finding, which emerged for both visual-only and auditory-only stimuli, contributes further evidence that the directional asymmetries expose a bias that is phonetically grounded.

There is no doubt that speech perception is strongly influenced by experience with a specific language. Collectively, the work summarized above confirms that universal perceptual processes also play a role in shaping adult vowel perception. Two published meta-analyses support these same conclusions with respect to infant vowel perception. The first meta-analysis, which included 19 articles containing 119 experimental records obtained using different behavioral and physiological methods, established that attunement to the native language begins to emerge in the first year of life (Tsuji & Cristia, 2013). The second meta-analysis was conducted on an updated dataset that also includes acoustic measures of the stimuli used (Tsuji & Cristia, 2017). This meta-analysis showed that spectral acoustic distinctiveness and order effects predicted by the NRV framework are reliable predictors of effect size in infant vowel discrimination tasks.

The work of Tsuji and Cristia inspired us to take a meta-analytic approach to assess predictions from the NRV framework with respect to adult vowel perception. As a first step we conducted a mini-meta-analysis integrating data across the four adult studies summarized above. Our mini-meta-analysis addressed several questions. First, what is the effect size due to the focalization bias when data are combined across the four studies summarized above? Second, as predicted by NRV, is the focalization bias effect size similar across language groups and across stimulus modalities? Third, as predicted by NRV, is the focalization bias effect size reduced when memory demands are decreased (by ISI manipulations) to promote acoustic processing and disfavor phonetic processing? To address the latter two questions we analyzed the effect of several moderator variables on the focalization bias effect size.

## 2. Method

### Database

Data extracted from four studies were included in this mini meta-analysis (see Table 1). We included data from 16 test conditions that utilized dynamic speech or non-speech analogs. We excluded one record that utilized static visual images as our hypotheses pertain to perception of dynamic speech or speech-like events. The resulting data base included findings obtained with diverse stimulus types including synthetic speech, natural speech (in auditory, visual and AV modalities), as well as point-light and Lissajou ( $\infty$ ) analogs of vowel lip movements. Despite the variability in stimulus types, the data from these different studies are suited for a meta-analytic approach, since all conditions utilize the same AX discrimination task to test adults (for restrictions, see next sections). For this mini-meta-analysis we used A prime scores, which was the dependent variable reported in each study.

1. Masapollo et al (2018) JEP:HPP (experiment 5)
2. Masapollo et al (submitted) JASA - EL
3. Masapollo et al (submitted) JASA - EL
4. Masapollo et al (submitted) JASA - EL
5. Masapollo et al (submitted) JASA - EL
6. Masapollo et al (2018) JEP:HPP (experiment 4)
7. Masapollo et al (2017) JASA (experiment 2)
8. Masapollo et al (2017) Cognition (experiment 2)
9. Masapollo et al (2017) Cognition (experiment 2)
10. Masapollo et al (2018) JEP:HPP (experiment 1)
11. Masapollo et al (2018) JEP:HPP (experiment 3.2)
12. Masapollo et al (2017) Cognition (experiment 1)
13. Masapollo et al (2017) Cognition (experiment 1)
14. Masapollo et al (2017) Cognition (experiment 3)
15. Masapollo et al (2018) JEP:HPP (experiment 3.1)
16. Masapollo et al (2017) JASA (experiment 2)

Table 1. References for each condition entered in the meta-analysis

### **Moderator Variables – speech only conditions.**

The effect of three moderator variables – language, modality, and ISI – was examined for the 12 conditions conducted with speech stimuli. The four conditions using non-speech visual analogs were removed because there is no data on ISI or language with these stimulus types and our



hypotheses concerning these moderators pertain specifically to speech processing. For each condition, participants' native language, stimulus modality, and ISI (inter-stimulus interval) were coded as moderators. Only English- and French-speaking participants have been tested in included studies (EN=9, FR=3). The data base included three stimulus modalities: audio-only ( $n=6$ ), visual-only ( $n=5$ ), and audio-visual ( $n=1$ ) stimuli. Due to limited audio-visual data, the conditions were collapsed to form two modality types: audio-only and AV or visual-only. By collapsing AV with Visual-only conditions we can examine whether the focalization bias effect is affected by the presence vs absence of visual speech information. The data set included three levels of ISI: 1500ms ( $n=8$ ), 1000ms ( $n=2$ ), and 500ms ( $n=2$ ). Prior studies of vowel discrimination reveal a gradient decay in auditory memory (and decline in discrimination performance) as ISI is increased up to 2000 ms, especially for within category stimuli. This decay (and associated decline) is quite steep between 500 and 1000 ms and very gradual between 1000 ms and 1500 ms (Cowan & Morse, 1986; Experiment 2). For this reason (and given our limited data on ISI) the ISI conditions were collapsed to form to two ISI types: short ISI (500ms) and long ISI (1000ms and 1500ms).

### Meta-Analytic Procedures

The analyses were conducted with the open-source package “metafor” (Viechtbauer, 2010) in R (R Core Team, 2018). Our effect size of interest represented the difference in discrimination by direction of vowel contrast. We calculated effect sizes based on the unbiased accuracy score ( $A'$  prime score, see Masapollo, Polka, & Menard, 2017 footnote two for more details) for each direction. Since each participant was tested in both directions, there were two dependent outcome values per sample.

Based on these values, we calculated Hedges'  $g$  effect size to represent the difference between directions within a sample. Like Cohen's  $d$ , Hedges'  $g$  reports the effect size in standard deviation units of the dependent variable while including a correction factor for small sample sizes. We also used Pearson correlation coefficient  $r$  for a within-subject experimental design correction. The calculations are found in Borenstein, Hedges, Higgins, & Rothstein (2009a).

Effect sizes were weighted by their inverse variance and entered into a random-effects model (Borenstein, Hedges, Higgins, & Rothstein, 2009b). A random-effects model without any moderators was applied to estimate the overall effect of focalization: *model = effect size, effect size*

*variance, method = maximum likelihood estimator, weighted = TRUE.* was calculated to further estimate the proportion of heterogeneity over the total variability (Higgins, Thompson, Deeks, & Altman, 2003). Then, the *Q*-test of heterogeneity was performed to test whether the heterogeneity among the true effects was significant. We also conducted a sensitivity test on the overall effect of focalization (all conditions; no moderators), leaving one effect size out at one time, this was done to detect influential cases and check the stability of the overall focalization bias effect size.<sup>1</sup>

Next, we ran three analyses to examine the effects of each of our three moderators. Because only speech conditions were included in the moderator analysis, we initially fitted a random-effects model including only the 12 speech conditions as a base model, *model = effect size, effect size variance, method = maximum likelihood estimator, weighted = TRUE.* For each moderator analysis, a mixed-effects model with a moderator, *model = effect size, effect size variance, mods = ~ moderator, method = maximum likelihood estimator, weighted = TRUE,* was used. The effect of each moderator was estimated and a *z*-test was conducted to examine whether the coefficient was significantly different from zero. Then a likelihood ratio test was conducted to compare each full model (all speech conditions; including the moderator) with the base model (all speech conditions; no moderator).

### 3. Results

In total, four studies with 16 experimental conditions (with a total of 242 adults tested) were included in this mini meta-analysis. Each condition is assigned a number and Table 1 provides the specific reference for each condition. The forest plot shown in Table 2 includes all conditions listed in order of effect size (ES) magnitude, with smallest ES at the top running to the largest ES at the bottom. Each line in this Figure 2 provides details on one condition (“tree”) in the overall forest plot. The details provided include (from left to right) the assigned number, the language group tested, stimulus type, stimulus modality, ISI used in the AX task, the sample size, the mean and standard deviation for A prime scores for AX trials with a central (English /u/) to peripheral (French /u/) direction of change, mean and standard deviation for A prime scores for the reverse direction of

---

<sup>1</sup> We do not report a funnel plot analysis, which is typically conducted to assess publication bias, because this meta-analysis featured appropriate data from studies in our lab, all of which have been published or submitted for publication. Thus, there were no “file drawer” conditions excluded from this meta-analysis.

change, a plot showing mean difference in A prime scores across the two directions plotted as a square with 95% confidence intervals indicated. The dotted line is at zero (indicating no difference between directions); a mean value greater than zero corresponds to a directional asymmetry indicating a focalization bias. The squares that are plotted for each condition vary in size to show the weight of that condition in the overall meta-analysis, with larger squares denoting a greater weighting. The weighting is jointly determined by effect size, sample size, and margin of error (confidence interval). The last column on the far right, shows hedge's  $g$  for each condition and the confidence intervals (5%, 95%) around this effect size estimate. At the bottom of the forest plot the observed outcome is plotted as a diamond. The horizontal mid-point of the diamond corresponds to the overall effect size computed across all 16 conditions. The left and right points of the diamond correspond to the confidence limits (left = 5%, right = 95%) around the combined effect size. In this meta-analysis, the confidence interval is quite narrower for the combined effect size making it difficult to visualize. A smaller confidence interval is expected given that combining data across studies often yields a more precise estimate of the effect size. To the right of the diamond, Hedges'  $g$  and confidence limits (5%, 95%) around it are also reported. The statistics for the heterogeneity test ( $Q$  test of heterogeneity, residual heterogeneity proportion ) are indicated at the bottom left.

### **The Effect of Focalization**

The estimated overall effect size of the focalization bias was .34, with 95% CI [.24, .44],  $z=6.76$ ,  $p<.0001$  (See Table 2). This corresponds to a small to medium effect size using the classification offered by Cohen (1988). The amount of total heterogeneity (i.e. between studies variation),  $I^2$ , was .0069 and 18.11% of total variability was explained by the heterogeneity instead of sampling error,  $Q(15)=21.76$ ,  $p=.114$ . This indicates that inconsistency among studies was relatively low. The result of the sensitivity test showed that there was no influential case and the estimated overall effect size varied from .32 to .36. Thus the effect size is stable within the small to medium range.

Table 2. The forest plot of the overall model showing 16 effect sizes.

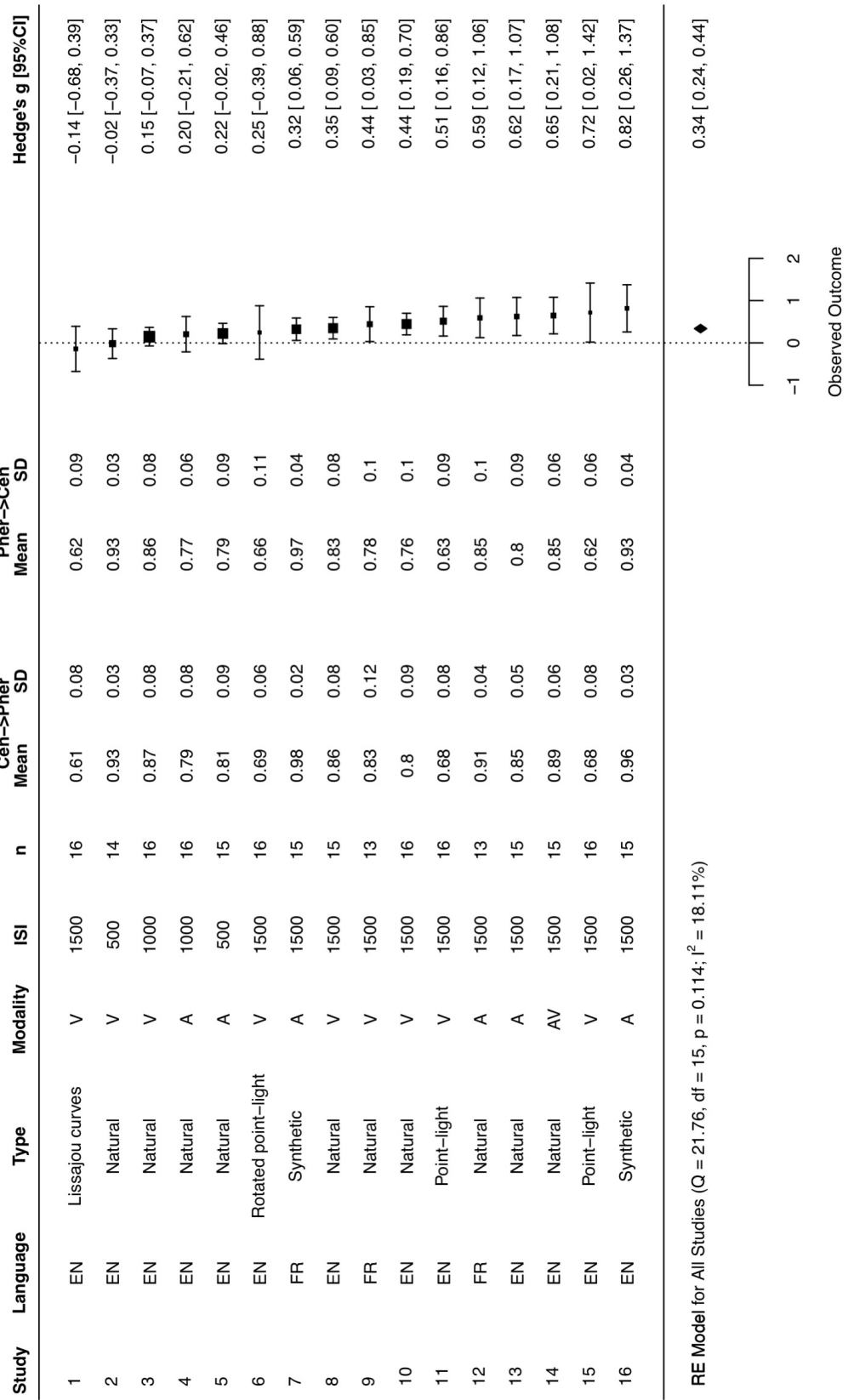


Table 2

## Moderator Analyses

### Effect of Focalization - Speech conditions only

We initially fit a base model without any moderators by using 12 speech conditions (2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 14, and 16), then each full model (including a moderator) was compared. The estimated effect size of the focalization bias for speech conditions was .33 with 95% CI [.23, .44],  $z=6.36$ ,  $p < .0001$ . This corresponds to a small to medium effect size. The inconsistency among conditions was relatively low ( $I^2 = .0057$ ,  $Q(11)=16.46$ ,  $p=.125$ ). The result is consistent with the overall effect of focalization reported above.

### Language

A moderator analysis was conducted with the 12 speech conditions to determine if the focalization bias effect size was modulated by language experience. The resulting forest plot is shown in Table 3. The upper portion of the Table 3 shows the conditions conducted with French adults listed in order of effect size magnitude; the lower portion show the conditions conducted with English adults listed in order of effect size magnitude. For each language group, the estimated effect size and CI for that language group is plotted right below the corresponding section and is superimposed on each condition as a light grey diamond. The magnitude of the focalization bias did not differ across the two language groups in our data set (beta=-0.10, 95% CI [-0.34, .15],  $z=-0.77$ ,  $p=.439$ ). As expected the full model with language as a moderator is not better than the base model (LRT=.57,  $df=1$ ,  $p=.449$ ). Both groups displayed a small to medium effect size for focalization bias. The weighted average effect size for French-speaking samples ( $n=3$ ) was .40 with 95% CI [.20, .60],  $z=3.90$ ,  $p < .0001$ . The weighted average effect size for the English-speaking samples ( $n=9$ ) was .32 with 95% CI [.19, .45],  $z=4.93$ ,  $p < .0001$ . French and English listeners completed identical perceptual tasks in the following conditions: 7 and 16, 8 and 9, 12 and 13. Within these matched conditions, the largest difference in focalization bias across language groups was observed for synthetic speech and within each direction discrimination performance was consistently higher for French adults than English adults.

**Table 3.** The forest plot for speech conditions: language as a moderator of focalization effect size.

Study	n	Cen->Pher		Pher->Cen		Hedge's g [95%CI]
		Mean	SD	Mean	SD	
<b>French</b>						
7	15	0.98	0.02	0.97	0.04	0.32 [ 0.06, 0.59]
9	13	0.83	0.12	0.78	0.1	0.44 [ 0.03, 0.85]
12	13	0.91	0.04	0.85	0.1	0.59 [ 0.12, 1.06]
RE Model						0.40 [0.20, 0.60]
<b>English</b>						
2	14	0.93	0.03	0.93	0.03	-0.02 [-0.37, 0.33]
3	16	0.87	0.08	0.86	0.08	0.15 [-0.07, 0.37]
4	16	0.79	0.08	0.77	0.06	0.20 [-0.21, 0.62]
5	15	0.81	0.09	0.79	0.09	0.22 [-0.02, 0.46]
8	15	0.86	0.08	0.83	0.08	0.35 [ 0.09, 0.60]
10	16	0.8	0.09	0.76	0.1	0.44 [ 0.19, 0.70]
13	15	0.85	0.05	0.8	0.09	0.62 [ 0.17, 1.07]
14	15	0.89	0.06	0.85	0.06	0.65 [ 0.21, 1.08]
16	15	0.96	0.03	0.93	0.04	0.82 [ 0.26, 1.37]
RE Model						0.32 [0.19, 0.45]

Moderator Analysis: beta = -0.10, 95%CI [-0.34, 0.15], p = 0.439.



Table 3

### Stimulus Modality

A moderator analysis was conducted with the 12 speech conditions to determine if the effect size for the focalization bias was modulated by stimulus modality. The forest plot is shown in Table 4. The upper portion shows the auditory-only conditions ordered by effect size magnitude. The lower portion shows the visual and AV conditions also ordered by effect size magnitude. For each modality, the estimated effect size and CI for that language group is plotted right below the corresponding section and is superimposed on each condition as a light grey diamond. Recall that two stimulus modality types were included in the analysis: audio-only and AV or visual-only. The effect size related to focalization did not differ across modalities types ( $\beta = -0.07$ , 95% CI [-0.28, .14],  $z = -0.66$ ,  $p = .511$ ). As expected, the full model is not better than the base model (LRT=.43,  $df=1$ ,  $p = .511$ ). A small to medium effect size was observed in each modality. The weighted average effect size for audio-only conditions ( $n=6$ ) was .36 with 95% CI [.22, .50],  $z=4.99$ ,  $p < .0001$ , while the estimated effect size for AV /visual-only conditions ( $n=6$ ) was .31 with 95% CI [.16, .45],  $z=4.09$ ,  $p < .0001$ .

### ISI

A moderator analysis was conducted with the 12 speech conditions to determine if the effect size for the focalization bias was modulated by ISI. The forest plot is shown in Table 5. Recall that two ISI types were included in the analysis: short ISI (500ms) and long ISI (1000 or 1500 ms). The upper portion shows the short ISI conditions ordered by effect size magnitude. The lower portion shows the long ISI conditions also ordered by effect size magnitude. For each ISI type, the estimated effect size and CI for that language group is plotted right below the corresponding section and is superimposed on each condition as a light grey diamond. The effect size related to focalization differ across ISI types ( $\beta = .23$ , 95% CI [ $< .01$ , .47],  $z = 1.96$ ,  $p = .050$ ). The full model is slightly better than the base model (LRT=3.77,  $df=1$ ,  $p = .052$ ). The estimated effect size for short ISI conditions ( $n=2$ ) was .15 with 95% CI [-0.05, .34],  $z = 1.44$ ,  $p = .149$ . The estimated effect size for long ISI conditions ( $n=10$ ) was .38 with 95% CI [.27, .50],  $z = 6.62$ ,  $p < .0001$ . The weighted averaged effect size was significant within the long ISI condition but not within the short ISI condition, which suggests that ISI is a potential moderator of the focalization effect size that may gain support if the data related to ISI was augmented.

Table 4. The forest plot for speech conditions: stimulus modality as a moderator of focalization effect size.

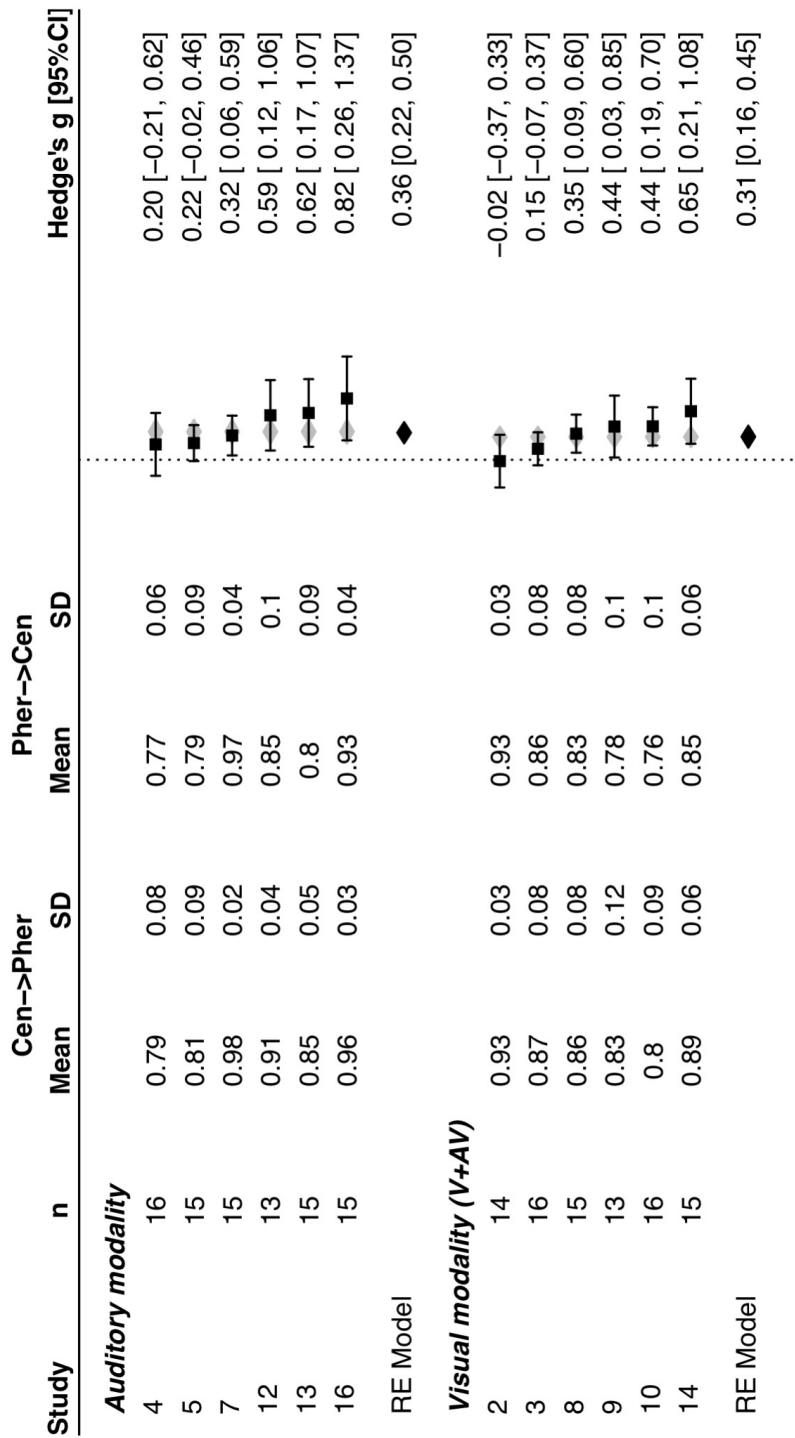


Table 4



Table 5. The forest plot for speech conditions: ISI as a moderator of focalization effect size .

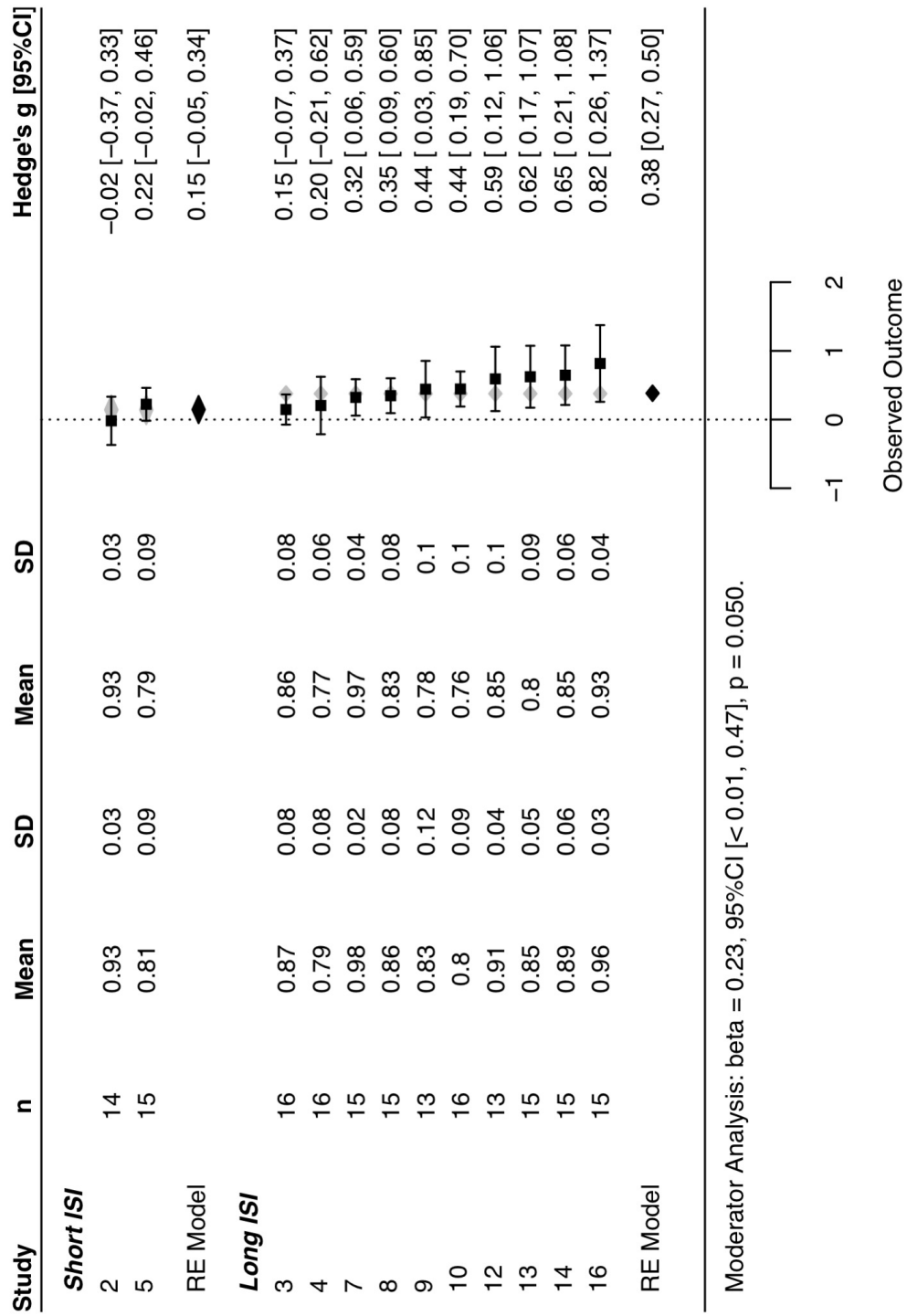


Table 5

#### 4. Discussion

In this chapter we present a qualitative review of recent adult cross-language research designed to examine the perceptual and cognitive processes underlying directional asymmetries in vowel perception. We also report the results of a mini-meta-analysis of this body of research which was undertaken to gain a more rigorous and comprehensive assessment of this work. This meta-analysis was limited to data from the series of studies reviewed above in which we assessed the focalization effect for a sub-phonemic contrast (more focal French /u/ vs less focal English /u/) across diverse stimulus types, language groups, and task demands.

Overall, the meta-analytic findings support the central tenets of the NRV framework. The collective evidence confirms that adults display a stable and reliable directional asymmetry in vowel discrimination that is tied to focalization differences. This bias has a small to medium effect size when measured in A prime units, which is a conservative (unbiased) index of discrimination.

Analyses of several variables that potentially moderate the focalization effect were also in line with NRV predictions. As predicted, the perceivers' native language did not modulate the focalization effect size. This further strengthens our claim that the NRV bias is distinct from the NLM effect. However, given the limited language diversity in this initial met-analysis, this should be re-assessed with an augmented data set. Importantly, although we claim that NLM and NRV describe distinct factors that shape vowel perception, they are not mutually exclusive. Interactions between these biases may emerge in other contexts or language groups.

Also as predicted, differences in stimulus modality did not modulate the focalization effect size. Thus, the focal vowel bias appears to be multi-modal and comparable in magnitude when assessed via vision or audition. This provides strong support for our claim that NRV is phonetically grounded and cannot be explained by general auditory processing biases alone.

The moderator analysis indicates a marginal trend for the focalization effect size to be modulated by the inter-stimulus interval used in the AX task. Thus, the NRV-based prediction regarding ISI was not firmly supported in this meta-analysis. However, the observed trends within the long and short ISI subgroups suggest that this factor may emerge in data set that is augmented with additional studies that include short ISI conditions. Thus, further research addressing this issue is needed to draw a firm conclusion regarding this task variable.

Overall most of the NRV frame-work predictions were supported in individual studies and also backed up in our integrative meta-analysis. As a next step, it will be informative to augment this meta-analysis to include data from other sub-phonemic and phonemic contrasts, other language groups, and other discrimination tasks. We invite researchers to contribute appropriate data to us as we begin to build a more comprehensive data set. Specifically vowel discrimination data (published or unpublished) that can be analyzed to assess effects of directional asymmetries in adult listeners, with native or non-native contrasts, will be informative.

Most of us are familiar with meta-analysis as a big undertaking that involves a thorough and comprehensive collection and integration of work within a specific field of research. However meta-analysis has much to offer and can be implemented on many different scales - with just a few experiments or with a large and multi-faceted data set. The main benefit of this approach is that it provides a way to look beyond an individual study and ground our interpretation in a more precise estimate of effect size gathered from a body of data rather than the dichotomous outcome of a single study. The focus on effect size (instead of null hypothesis tests) also pushes us to ask a deeper question – how big is an effect and is the magnitude of this effect modulated (or not) by specific factors as predicted by our hypothesis or conceptual framework. Thus integrating data in a meta-analytic framework provides both a more comprehensive and a more rigorous test of our hypotheses. By uncovering the strengths as well as the limitations of a body of research from a data analytic perspective, meta-analysis can also guide and motivate future research in productive directions. Following the example of Cummings (2012), we encourage our fellow speech scientists to add a meta-analytic perspective and tools to their research program.

### **Acknowledgements**

This chapter is supported by NSERC funding to L. Polka and CSC funding to Y. Ruan. Special thanks to Sho Tsuji and Christina Bergmann for their helpful input on the analyses and manuscript.

## References

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009a). Effect Sizes Based on Means *Introduction to Meta-Analysis*: John Wiley & Sons, Ltd
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009b). Fixed-Effect Versus Random-Effects Models *Introduction to Meta-Analysis*: John Wiley & Sons, Ltd.
- Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*, (2<sup>nd</sup> Edition). Hillsdale: NJ: Erlbaum.
- Cowan, N., & Morse, P. (1986). The use of auditory and phonetic memory in vowel discrimination. *Journal of the Acoustical Society of America*, 79, 500-507.
- Cummings, G. (2012). *Understanding the New Statistics – Effect sizes, Confidence Intervals and Meta-Analysis*, Routledge, Taylor and Francis Group, LLC.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e), *Philos. Trans. Royal Soc. B* 363, 979-1000.
- \*Masapollo, M., Franklin, L., Morgan, J., & Polka, L. (submitted). Asymmetries in vowel perception arise from phonetic encoding strategies. *Journal of the Acoustical Society of America – Express Letters*
- \*Masapollo, M., Polka, L., & Menard, L. (2017). A universal bias in adult vowel perception – By ear or by eye. *Cognition*, 166, 358-370. doi:10.1016/j.cognition.2017.06.001
- \*Masapollo, M., Polka, L., Menard, L., Franklin, L., Tiede, M., & Morgan, J. (2018). Asymmetries in unimodal visual vowel perception: The roles of oral-facial kinematics, orientation, and configuration. *J Exp Psychol Hum Percept Perform*. doi:10.1037/xhp0000518
- \*Masapollo, M., Polka, L., Molnar, M., & Menard, L. (2017). Directional asymmetries reveal a universal bias in adult vowel perception. *J Acoust Soc Am*, 141(4), 2857. doi:10.1121/1.4981006
- Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41(1), 221-231. doi:10.1016/s0167-6393(02)00105-x
- Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, 39(4), 467-478. doi:10.1016/j.wocn.2010.08.007
- R Core Team. (2018). R: A language and environment for statistical computing Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Tsuji, S., & Cristia, A. (2014). *Percetual attunement in vowels: A meta-analysis*. *Developmental Psychobiology*, 56(2), 179-191.
- Tsuji, S., & Cristia, A. (2017). *Which Acoustic and Phonological Factors Shape Infants' Vowel Discrimination? Exploiting Natural Variation in InPhonDB*. Pa-

- per presented at the Interspeech 2017.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J Stat Softw*, 36(3), 1-48.
- Werker, J. F., & Tees, R.C. (1983). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75(6), 1866-1878.
- Werker, J. F., & Logan, J.S. (1985), Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37, 35-44.
- Higgins, J. P. T., Tompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medicine Journal*, 327(7414), 557-560.

\*studies providing data for the meta-analysis.

## Second and Third Language Immersion Students' Pronunciation in Foreign Language English Oral Reading

Anja K. Steinlen & Thorsten Piske  
Friedrich-Alexander-University Erlangen-Nürnberg

Sophia Karmeli & Christine Mooshammer  
Humboldt University Berlin

### Abstract

The present study deals with the English pronunciation of majority and minority language children attending a German-English elementary school immersion program in Germany. In this program, 50% of the teaching time was conducted in English. By using a reading aloud task, we assessed phonemic accuracy as well as reading fluency in English and related both to (i) the English input the children received from their teachers and (ii) possible sources of transfer. So far, cross-linguistic influences in young learners' L1, L2 and L3 phonological acquisition have received only very limited attention.

Articulatory transcriptions of the immersion students' English reading data indicate transfer patterns from German to English, independent of the children's L1. These findings are discussed in the light of teacher input and various sources which may account for transfer in majority and minority language children's English pronunciation.

### 1. Introduction

In Germany (as in many other countries) the number of elementary schools offering bilingual programs is steadily increasing. Currently, there are over 300 private and public elementary schools, corresponding to 2% of all elementary schools (FMKS, 2014). Immersion (IM) programs represent

---

Anne Mette Nyvad, Michaela Hejrná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 583-605). Dept. of English, School of Communication & Culture, Aarhus University.

the most intensive type of bilingual education. In these programs, 50-100% of the teaching time is conducted in the target language. The effectiveness of these programs has been demonstrated in a large number of studies, which have mainly focused on majority language students' reading, writing, speaking, listening, and grammatical skills as well as on their attitudes and motivation (see reviews by e.g. Wesche, 2002, for North America, Pérez-Cañado, 2012, for Europe, and Piske, 2015, for Germany).

However, there are only a few studies that have assessed bilingual students' phonemic accuracy and fluency in the target language (see e.g., Harada, 2007; Rallo Fabra & Jakob, 2015; Wode, 2009). There are even fewer studies that have examined the phonological development of those students in bilingual programs for whom the target language is not the second but the third language (i.e. minority language students, e.g. Hart, Lapkin, Swain, 1987). In order to provide much needed additional data, the present study compares the phonological development of majority and minority language children enrolled in bilingual programs. All the children examined here attended a German-English IM elementary school program in Germany, in which 50% of the teaching time is conducted in English. We assessed accuracy (of selected English sounds) as well as fluency (i.e. speech rate) in English by relating it to (i) the English input which the children received from their English teachers and (ii) the transfer source (i.e., L1 German for the majority language children; and the minority language children's L1 and L2). The following review will be devoted to studies dealing with the L2 phonological development of majority language children in bilingual programs (section 1.1), studies examining phonological aspects in third language (L3<sup>1</sup>) acquisition (section 1.2) and studies examining the L2/L3 exposure students receive in the foreign language classroom, i.e. teacher input (section 1.3).

### 1.1 L2 phonology

Various models have been proposed to account for foreign accent in L2 speech, for example, Flege's *Speech Learning Model* (SLM, e.g. 1995), which posits that the processes and mechanisms used in the successful acquisition of the L1 sound system, including the ability to establish phonetic categories, remain intact across the lifespan and can also be

<sup>1</sup> The terms L2 and L3 will be used according to the chronological onset of acquisition, i.e. the term 'second language' (L2) refers to the first non-native language acquired by an individual, while 'third language' (L3) relates to the second non-native language being learned (see also Hahn & Angelovska, 2017).

exploited in the acquisition of L2 speech. However, the acquisition of L2 sounds depends on the perceived cross-language phonetic distance between sounds of the L2 and the L1 as well as on the state of development of the L2: An L2 sound that is not too similar to a native-language (L1) sound will be easier to acquire than an L2 sound that is relatively similar to an L1 sound (because it will be perceived as more obviously “different” by the learner). The SLM has also been applied to German learners of English and their production of English vowels (e.g. Bohn & Flege, 1992, see also Steinlen, 2005): Of particular interest for the present study is English /æ/ which is a phoneme not found in most dialects of German, including Standard German. Acoustic cross-language comparisons (Bohn & Flege, 1992) suggested that English /æ/ is a new vowel because there is hardly any spectral overlap between English /æ/ and the closest German vowels /e:, ε, a/; furthermore, English /æ/ is produced with a longer duration. Turning to the production of English /æ/ by German learners, the results of Bohn & Flege’s (1992) study were largely consistent with Flege’s hypothesis that extended L2 experience will enable adults to produce a new vowel in a natively-like fashion. The inexperienced learners, however, did not differentiate between English /æ/ and German /ε/, which suggests that they used only one vowel category where the native English speakers and experienced German speakers of English used two. We would, therefore, predict that German primary school children learning English would show similar production patterns as Bohn & Flege’s inexperienced adult learners, i.e. they would not be able to produce English /æ/ in a target-like manner. Their teachers, in contrast, would have established a separate phonetic category for English /æ/ and produce this sound in a target-like way, just like Bohn & Flege’s (1992) experienced German learners of English.

Rather problematic English consonants for German learners of English seem to be the dental fricatives /ð/ and /θ/, and the alveolar or retroflex approximant /r/ in prevocalic position. The last sound is often substituted with German /ʀ/, the dental fricatives are often realized either as labiodental or alveolar fricatives or alveolar stops (e.g. Eckert & Barry, 2002; König & Gast, 2012). Other transfer phenomena include syllable structure processes based on the learners’ L1 German, such as devoicing of final voiced obstruents<sup>2</sup> or devoice-ment of nonsyllabic-initial [ɫ]. These sounds (including English /æ/) have also been examined by Wode

---

<sup>2</sup> In English, voiced obstruents in word-final position are preceded by vowel lengthening, which additionally poses a problem for German learners of English (e.g. Smith, Hayes-Harb, Bruss, & Harker, 2009)



(2009) in a study of German-English immersion preschool and primary school students in Germany. In his paper, Wode (2009) stressed the large number of parallels between the errors produced by German learners of L2 English across different age groups (children and adults), who acquired the L2 in diverse learning situations (i.e. in naturalistic vs. IM vs. regular classroom contexts). Focusing on the IM context, Wode reported that preschoolers at age 3 already showed transfer patterns from their L1 German (apart from errors due to the development of children's L1 phonological system). These transfer-based substitutions included alveolar fricatives used instead of dental fricatives, clear /l/ instead of nonsyllabic-initial velarised [ɫ] and /ɛ/ for English /æ/. There were only a few cases of [ʁ] and [w] substituting for target /-r-/. Similar substitution patterns were noted for primary school IM children in Grade 4, who, according to Wode (2009), reflected the same segment substitutions, the same transfer patterns, the same range of individual variation, and the same kind of global German accent in their English as the IM-preschoolers. However, the frequency of the target-like productions increased from grade level to grade level, i.e. from 30% to 79% target-like productions for /ð/ and 69% to 86% for /θ/. However, [ɫ] did not show any more target-like production as a function of time (57% vs. 55%), and /æ/ was produced in a more target-like manner in only 9% and 13% of all cases, respectively. Similar substitution patterns are expected for the majority language students in the present study whose L1 is also German.

Only a few studies have examined L2 fluency in bilingual programs: For example, Rallo Fabra & Jacob (2015) focused on so-called CLIL (*Content and Language Integrating Learning*) programs, where only one subject, (i.e. History and Geography, respectively) was taught in English. They compared Spanish-English CLIL and non-CLIL students in Grade 8 with respect to fluency (operationalized as speech rate) in their L2 English, using reading-aloud data and extemporaneous speech. The results of their study indicated that both groups did not differ with respect to their speech rates, which the authors attributed to the teachers of either group who were not native English speakers, and who were, unfortunately, not tested for their speech rates in English. We would expect the primary school children in the present study to produce similar speech rates as their teachers because as IM students, they had received a very large quantity of English input from their teachers.

## **1.2 L3 phonology**

As regards third language (L3) acquisition, phonological aspects constitute a relatively unexplored research area. In contrast to learners acquiring a phonological system in the L2, L3 learners have already acquired an L2 and can thus make use of conscious linguistic knowledge as well as of language-learning experience and strategies (e.g., De Angelis 2007; Lloyd-Smith, Gyllstad & Kupisch, 2016). Not surprisingly, L3 acquisition is characterized by the simultaneous influence of more than one previously acquired language (i.e., the L1 and the L2, De Angelis, 2007).

### **1.2.1 Age**

Research on the relationship between age and cross-linguistic influence in L3 phonological acquisition has received very limited attention so far: Cenoz (2001) pointed out that cognitive and metalinguistic development may be related to cross-linguistic influence, and particularly, to psychotypology, because older children may have a more accurate perception of linguistic distance that could influence the source language they use when transferring terms from one of the languages they know.

Kopečková (2013) examined twenty 5th graders' productions of rhotic sounds in their L1 German, L2 English and L3 Spanish. Her results indicated that the intrinsic difficulty of the phonetic feature of the Spanish trill may have affected L3 pronunciation to a large degree as this sound requires a higher degree of articulatory and aerodynamic precision than the uvular fricative in German or the alveolar approximant in British English. Reyes, Arechabaleta-Regulez & Montrul (2017) examined Spanish rhotic sounds produced by Spanish native speakers, English native speakers acquiring Spanish as an L2 and Korean-English bilinguals acquiring Spanish as an L3. They reported that although all children rapidly developed a native-like pronunciation of the Spanish rhotic sounds, the Korean-English bilinguals outperformed the English-speaking children. According to Reyes et al. (2017), not only previous linguistic knowledge may thus play a role in L2 and L3 acquisition but children may overcome transfer errors because they are guided by universal developmental strategies from the initial stages of acquisition. If L3 learners have an advantage over L2 learners, this may be due to their complex linguistic knowledge and higher metalinguistic competence.

So far, minority and majority language students have only been compared in terms of their oral fluency in the new language (but not regarding their pronunciation accuracy) and even such studies are scarce and relate to IM programs only: Hart, Lapkin, & Swain (1987) compared the oral fluency of minority and majority language students in a middle IM program in Grade 8 and found that minority language students outperformed their majority language peers. In general, oral fluency ratings did not appear to be related to their parents' occupation, independent of the students' language background. Hart et al. (1987) reported similar results for early IM programs and also reported general effects of program, i.e. better oral fluency ratings for students in early IM programs than in middle IM programs.

Previous research in L3 acquisition – and most of the studies in L3 phonology – have examined more advanced adult L3 learners and possible transfer patterns. These have largely been discussed in the light of three models (see e.g. Lloyd-Smith, Gyllstad & Kupisch, 2016, for a review): The Cumulative Enhancement Model (CEM, e.g. Flynn, Foley & Vinnitskaya, 2004) maintains that any language available to the multilingual learner can be the source of transfer, irrespective of the order of acquisition. Transfer only occurs when such knowledge has a facilitative effect; otherwise it is neutralized or “blocked”. According to such a view, the learner does not transfer an entire system but only individual properties. According to the Typological Primacy Model (TPM, Rothman, 2011, 2015), multilingual transfer is determined by structural similarities between languages (Rothman, 2011, 2015), where transfer is assumed to occur completely from one previous system, much like in Schwartz and Sprouse's (1996) Full Transfer Model. Finally, the L2 Status Factor Model (L2SFM, Bardel & Falk, 2007) hinges on the distinction between L1 and L2 acquisition and predicts L2 transfer into L3 due to similarities in the learning procedures in L2/L3 acquisition as opposed to L1 acquisition. Lloyd et al. (2016) point out that although these models pertain to L3 transfer at the initial state, more advanced adults L3 learners have been used as subjects. In addition, studies conducted so far have not completely testified to the CEM, the TPM, or the L2SFM models.

Most studies to date point to the existence of the so-called “foreign language effect” in L3 phonological acquisition, which typically seems to exist in the early stages of L3 acquisition, suggesting that a foreign accent may be based on aspects such as age, L2 proficiency, L2 status, or psycho/typological distance (e.g., Ringbom, 1987).

### **1.2.2 Status**

Llisteri & Poch (1987) acoustically analyzed L3 French vowels and consonants produced by native speakers of Catalan and L2-Spanish and found that the learners' L1 affected their L3 oral production without any interference of their L2. Based on these results, they postulated a privileged status of the L1 system as the main source for L3 phonology. Similar results were reported by Wrembel (2012): Her participants were native speakers of Polish who were all proficient users of L2 English but differed in terms of their proficiency level in their L3 French. Their speech samples were evaluated online by expert raters who found that the prevailing source of transfer was the participants' L1 (although some L2 influence was also noticeable). Finally, a study with five Turkish-German heritage speakers learning L3 Spanish tentatively indicated that higher proficiency in the heritage language may also facilitate positive transfer from the L1 (Gabriel & Rusca-Ruths, 2014). The Turkish-German heritage speakers tended to produce the rhythm of L3 Spanish more monolingual-like than five German monolinguals, suggesting positive transfer from Turkish, which is syllable-timed like Spanish. This effect was stronger in individuals with a higher frequency of use in Turkish.

Studies in favor of L2 proficiency include Hammarberg's (2001) single-case study, in which an L3 Swedish learner with L1 English and L2 German was perceived to have a "prominent" German accent during her first year in Sweden, yet speech samples recorded one year later were perceived by the same raters as distinctly English. The activation of the L2 at the initial stage of acquisition was seen as an unconscious strategy employed by the speaker to cope with unfamiliar phonological forms. As proficiency in L3 increased, this strategy was overridden by the highly-automated articulatory patterns of the L1 (Hammarberg, 2001, p. 35). Similar results were reported by Wrembel (2010) who examined L1 Polish, L2 German, and low proficiency L3 English speakers who were mistaken as German speakers more frequently than those with a higher proficiency, suggesting that L2 transfer was more noticeable at the initial stage of L3 acquisition. However, this effect decreased with higher proficiency (see also Gut, 2010, for similar results). Finally, in their study of perceived foreign accent by German and German-Turkish adult learners of L3 English, Lloyd et al. (2016) found that the bilinguals with a high proficiency in German were predominantly perceived as German by English raters, while the others were perceived as non-German. In addition, the bilinguals' amount

of Turkish use seemed to be related to perceived accent in L3 English (although this relation did not yield any significance).

### **1.2.3 Typology**

Typological similarity between an L2 and an L3 are also believed to affect the process and the product of learning a third language in the sense that typological similarity may facilitate learning at the phonological level. For example, Bouchhioua (2016) found that her adult learners with L1 Tunisian Arabic and L2 French produced L3 English target words with French word stress patterns. Similar results were reported by e.g. Llama, Cardoso & Collins (2010) on L1-/L2 learners' pronunciation of L3 English, and Wrembel (2010, 2012) with L1 Polish, L2 French and L3 English. However, as Cabrelli Amaro (2012) critically pointed out, L3 phonological research has yet to agree on general aspects that constitute a typological relationship between languages (i.e., typological distance referring to the linguistic system as a whole, the phonological system as a whole, or the relationship of a single property across languages).

In a study that teased apart language status and distance in the production of VOT, Llama, Cardoso & Collins (2010) used adult groups with L1/L2 mirror images (L1 French/L2 English, L1 English/L2 French) acquiring L3 Spanish. The results showed that both groups transferred from L2, with the L2 French group producing target-like VOT values, and the L2 English group producing L3 stops with longer VOT than required in Spanish, a likely effect from English. Typological proximity was apparently not the motivating factor for transfer, although both French and Spanish are characterized by non-aspirated stops. In addition, psycho-affective factors may also account for transfer due to L2 status, as some participants of studies have been reported to express a desire to suppress their L1 in an effort to sound non-foreign (e.g. Lloyd-Smith et al. 2016).

### **1.3 Teacher input**

According to the Stifterverband (2013), 98% of the teachers in Germany have a German background. It is not clear, however, how many of the remaining 2% are native speakers of English. Medgyes (2013, p. 509) defines nonnative teachers as people "for whom the foreign language they teach is not their mother tongue; who usually work with monolingual groups of learners; whose mother tongue is usually the same as that of their students". Many studies have examined advantages and disadvantages of

being a nonnative or native teacher (e.g. Llurda, 2005) and foreign accent has been identified as one of the disadvantages of being a nonnative speaker. For the primary school context in Germany, in particular, it has often been criticized that the English teachers' pronunciation is far from being target-like (e.g. Süddeutsche Zeitung, 2012; FAZ, 2015). This was also shown in some studies examining English primary school teachers' pronunciation of English words in other countries, which found non-target like renderings on the segmental level as well as on the sub- and suprasegmental level (e.g. Kanoksilapatham, 2014; Yani, 2012). Thus, teachers' pronunciation errors may also be reflected in their students' speech, in particular because young learners like to imitate their teachers, who are, incidentally, the children's main source of foreign language input (e.g. Böttger, 2005; Piske 2008; Kanoksilapatham, 2014; Karakaş, 2012; Yani, 2012). However, studies relating teachers' pronunciation errors to those produced by their students have apparently not been conducted so far.

#### **1.4 Research questions**

In summary, previous research leaves open whether the same mechanisms that operate in majority language students also apply to minority language students, and thus, whether the existing models aiming to explain transfer in L2/L3 phonology can predict cross-linguistic influences for minority language children. Similarly, the role of teacher input has remained rather vague. The aim of the present study is, therefore, to address the following research questions:

- i. Do majority and minority language students attending an elementary immersion school program differ in their pronunciation of English, which is their L2 and L3, respectively?
- ii. Is there any relation between the English teachers' pronunciation and their students' English pronunciation regarding the general phonological error rate?

## **2. Method**

### **2.1 School**

The data presented in this paper were collected in a (non-private) district primary school in a city in the south of Germany. The school has offered a partial IM program since 2008, with one cohort per year. In this program, all subjects are taught in English from the first day of Year 1 onwards, except

for German language arts, religious education and math. The immersion students are thus exposed to both English and German for about 50% of the teaching time. Although technical terms are always introduced in both English and German, the subject lessons are taught entirely in English. The students usually receive their instruction from native speakers of German who studied English in order to become English teachers. The children are allowed to answer questions in German if they want to do so, but they are always encouraged to speak English (e.g. Steinlen & Piske, 2013).

## 2.2 Sample

For the present study, the data of 14 children (8 girls and 6 boys) in Year 4 were selected; they had all started the IM-program in Year 1 but attended different classes. On average they were 10.6 years old ( $SD=8,3$  months). Five of the children (i.e. 36%) had a minority language background, reflecting the overall demographics of the school, and nine had a majority language background. Such a background was attested when one or both parents were born abroad (see also OECD, 2016) and, most importantly, when a language other than the majority language German was spoken at home. The minority language children had all been born in Germany, and they all used their family language plus German at home. The parents' questionnaire, unfortunately, did not ask for information concerning the use of the family language and the use of German before the children had entered school. It is, therefore, not clear whether the minority language children had learned German as an L1 or an L2. In informal interviews, however, most parents stated that the family language was their children's L1, with German being acquired in preschool (at age 3) at the latest. The foreign language English is, therefore, the children's L3. The family languages included Turkish (2 children), Arabic (2) and Russian (1 child). The parents did not report any hearing problems of their children. The majority and minority language children were comparable in terms of their socioeconomic background as an informal look at the parents' questionnaires indicated.

In order to investigate how input contributes to the children's pronunciation of English sounds, data were also gathered from the students' four teachers in the IM program. All the teachers were female, between 27-34 years old at the time of testing and had a German background. They had studied English at a university in Germany (with a focus on bilingual teaching) and had spent at least a year in an English-speaking country

(Canada, Great Britain, New Zealand, South Africa, USA). Furthermore, they rated their English proficiency at level C2, following the levels proposed by the Common European Framework of Reference (Council of Europe, 2001).

### **2.3 Speech materials**

The *Gray Oral Reading Test* (GORT, Wiederholt & Bryant, 2001) was used to analyze the students' pronunciation after four years in the bilingual program. It was originally designed for L1 individuals aged 6 to 18 years and contains 14 separate stories. Each story is followed by five multiple-choice comprehension questions. Testing is discontinued if the student misses at least three of five comprehension questions for any one story. For the present study, however, the analysis of the data is restricted to the children and teachers reading aloud the first three stories (which were completed by all 14 children), disregarding the comprehension part.

### **2.4 Recordings**

At the school premises, the children were recorded in a quiet room by one of the members of the research group using an Olympus digital voice recorder (VN-3100/VN-3100PC). Two of the teachers, who were still working at the school at that time, were recorded with the same device. The other two teachers, who were not working at the school anymore because they had moved abroad with their families, sent their voice recordings via WhatsApp. All subjects were allowed a few minutes to silently read the text before they were recorded. Note that the recordings were originally not intended to be used for phonetic analyses.

### **2.5 Measurement procedures**

The three texts consisted of 113 words. All sound files were imported and annotated with the Praat program 6.0.05 (Boersma & Weenink, 2013) and transcribed orthographically as well as aurally. Because of the poor quality of the recordings (which were originally collected to assess oral reading skills and not pronunciation), the analysis of the number of syllables was conducted by hand, only pauses were detected automatically with Praat. The minimum silence interval duration was set at 0.2 seconds. Following Rallo Fabra & Jacobs (2015), the total number of syllables was divided by the total time required to produce the speech sample, including pauses, hesitations and fillers.



For the phonological error analysis, the words were marked in a separate annotation tier. After listening to the recordings, consonant and vowel identity was coded, using the symbols of the International Phonetic Association (1999). The focus of this pilot study is on the English targets /æ/ (9 targets, e.g. *at, can, have*), [ɫ] (20 targets, e.g. *little, play*), prevocalic /r/ (9 targets, e.g. *red, green*), the dental fricatives /ð/ and /θ/ (13 targets, e.g. *the, father, something*) and voiced obstruents in word-final position (20 targets, e.g. *rides, stars, good*) as these sounds are the most problematic ones for German learners of English (e.g., König & Gast, 2012). Altogether the corpus comprises of 1278 items (71 targets x 18 subjects). In a few cases, the children omitted a word while reading the text (14 omissions). Unfortunately, acoustic analyses of sounds were not possible due to the poor quality of the recordings.

### 3. Results

In order to examine differences between groups, mean speech rate measures as well as hit/miss scores for speech sounds obtained for each of the 18 subjects (fourteen children and four teachers) were submitted to one-way ANOVAs. The results of the descriptive analyses are presented in Table 1.

	Majority language students (N=5)	Minority language students (N=9)	Teachers (N=4)
<b>Fluency:</b>			
<b>speech rate</b>	2.28 syll/sec [SD=0.3]	2.34 syll/sec [SD=0.3]	2.97 syll/sec [SD=0.4]
<b>Accuracy:</b>			
/æ/	13,3%	38,9%	62,2%
[ɫ]	93.0%	96.2%	95.2%
prevocalic /r/	86.7%	100%	100%
dental fricatives /ð/ + /θ/	36.2%	61.5%	93.9%
w/f voiced obstruents	54.3%	54.8%	82.0%

Table 1. Descriptive analyses for mean speech rate (syllables per second) and mean hit rate in percent regarding the pronunciation of selected sounds (w/f = word final).

As Table 1 illustrates, teachers and students did not read the English texts at the same pace. This was confirmed by a one-way ANOVA, which yielded significant differences for group [ $F(2, 16)=6.262, p=.010, \eta_p^2=.963$ ].

Post-hoc tests indicated that the teachers' speech rates were considerably faster than those of the majority and minority language students ( $p < .05$ ). However, the two student groups did not differ significantly regarding their speech rate ( $p > .05$ ). Apparently, language background (majority vs. minority language students) did not exert any influence on speech rate but experience (teachers vs. students) did.

Some of the English sounds examined here are reported to be notoriously difficult for German learners of English to pronounce. However, the results listed in Table 1 suggest that this is not generally true: Indeed, prevocalic /r/ and [ʀ] were pronounced almost always in a target-like way by the three groups (the hit rate ranged between 87% and 100%). One-way ANOVAs did not yield any significant differences between the three groups, neither for [ʀ] [ $F(2, 16) = 1.847, p = .218, \eta_p^2 = .810$ ] nor for prevocalic /r/ [ $F(2, 16) = 2.381, p = .153, \eta_p^2 = .239$ ]. These two sounds apparently neither posed any difficulty for German learners of English (independent of their age/experience) nor for minority language students whose L1 was not German. In the few cases of incorrect pronunciation, [ʀ] was substituted for [l] and /r/ was replaced with /ʁ/, i.e. with the German sound that was most similar to the English sound.

The dental fricatives did not pose any problems for the English teachers; they were almost always pronounced in a target-like way, corresponding to a hit rate of 93.9%. They were, however, problematic for the students: Minority language children obtained a hit rate of 61.5%, whereas majority language children pronounced only a third of the dental fricatives correctly. A one-way ANOVA yielded significant differences of group [ $F(2, 16) = 14.806, p = .000, \eta_p^2 = .887$ ], and post hoc tests indicated significant differences between all three groups ( $p < .005$ ). Usually /d/ was used instead of /ð/ (only once did a child use /z/ instead of /ð/ for <the>), the same pattern applied to /ð/ in word-medial position (only one child produced a /t/ in <father>). Substitution patterns, however, varied for the word <with>: In two thirds of the cases, the children used /d/ instead of /ð/, followed by /f/ (4 instances), /t/ (2), and /s/ (1). The dental fricative in the word <something> was substituted by /f/ only.

Final devoicing posed a problem even for experienced learners of English: The teachers of the present sample obtained a hit rate of 82%. The students (independent of their language background) devoiced around half of all voiced obstruents in word-final position; language-specific patterns for devoicing were not detected. A one-way ANOVA revealed significant differences for group [ $F(2, 16) = 4.327, p = .033, \eta_p^2 = .857$ ], with teachers

performing considerably better than either group of students ( $p < .05$ ), who did not show any significant differences as a function of language background ( $p > .05$ ). Due to the poor quality of the recordings, acoustic measurements of vowel length could not be included in the analysis.

The teachers pronounced the vowel /æ/ in a target-like way in 67% of all cases, in the other instances they substituted English /æ/ with German /ɛ/. The majority language students showed stronger transfer effects with a hit rate of 13%. Minority language children, however, produced almost 40% of all /æ/ tokens in a target-like way. Therefore, there was a significant main effect of group in a one-way ANOVA [ $F(2, 16) = 11.985$ ,  $p = .001$ ,  $\eta_p^2 = .704$ ] and post-hoc tests revealed significant between-group differences ( $p < 0.05$ ) between teachers, minority and majority language children regarding their target-like use of /æ/.

#### 4. Discussion

The present study examined English reading-aloud data produced by majority and minority language children who all attended a German-English IM elementary school program in Germany, in which 50% of the teaching time was conducted in English. We assessed phonemic accuracy of selected English sounds as well as fluency (operationalized as speech rate) in English with regard to (i) the English input the children received from their English teachers and (ii) sources of transfer. So far, cross-linguistic influence in young learners' L2 and L3 phonological acquisition in educational contexts has received only very limited attention.

##### 4.1 Majority vs. minority language students

With regard to phonemic accuracy, the results for majority language children with a German background were very similar to those reported by Wode (2009) for preschool and primary school students in German-English IM programs in Germany: The 4<sup>th</sup> graders in our study indeed had problems with some English sounds (in particular /æ, ð, θ/). As expected, /æ/ was usually rendered as German /ɛ/ in almost all of the cases (87%), indicating that this vowel was still problematic for the learners. A similar result was obtained by Bohn & Flege (1992) for inexperienced adult L2 German learners of L2 English. Acoustic analyses would be a welcome addition to determine in more detail whether the children already show a slow phonetic shift from the native to the non-native vowel.

In contrast to Wode (2009), the dental fricatives were not only substituted by alveolar fricatives but also by labiodental fricatives and

by alveolar stops (see also König & Gast, 2012). For example, the word-initial sound in the very frequent word <the>, if not produced target-like, was pronounced with an alveolar stop, the same applies to word-medial /ð/ in <father>. The dental fricative in <something>, though, was regularly substituted by /f/, pointing to progressive assimilation processes at work. Prevocalic /r/ did, in line with Wode's study, not pose any difficulty for majority language children, the same applied to [ʔ]. However, the children based their pronunciation of English final voiced obstruents on their L1 German syllable structure and devoiced half of these items.

In terms of fluency, minority and majority language students in grade 4 did not significantly differ in their speech rate when they were reading the three texts aloud. This result differs from findings obtained by Hart, Lapkin & Swain (1987) who reported minority language students in grade 8 to outperform their majority language peers. It may be possible that four years are not sufficient for such effects to occur. However, as the sample size is only small, additional research with minority language students of different ages attending different IM programs is needed to examine such effects in more detail.

In general, the minority language children showed a more target-like production of the English sounds /æ/, [ʔ], /r/, /ð/, /θ/ and of voiced obstruents in word-final position, indicating that they were not disadvantaged compared to their majority language peers. However, the substitution patterns of both groups did not differ: English /æ/ was replaced by German /ɛ/, English dark [ʔ] by German clear [l], and the dental fricatives by either alveolar obstruents (e.g. <with>, <father>, or labiodental fricatives (<something>). Minority language students also devoiced obstruents in word-final position – just to a smaller extent as compared to their majority language peers.

The minority language children's data, therefore, suggest an influence of L2 German on their pronunciation of L3 English sounds: For example, the children's L1 Turkish and Arabic do not exhibit final devoicing, but they did not resort to their L1 when producing English voiced obstruents in word-final position but devoiced these sounds, in line with the phonological rules of the L2 German. Even though our learners have been exposed to English for four years, their foreign accent (at least with the English sounds being tested) is not based on their L1 Turkish or Arabic, rejecting their L1 as a possible source of transfer for L3 pronunciation. Aspects such as L2 proficiency and/or psycho/typological distance seem to play a greater role: For example, minority language children in the IM program of this particular school may generally be described as having a high command

of German, as shown in standardized tests of German reading and writing (e.g. Steinlen 2016, 2018). The same is true for the individuals of the present sample, as an informal look at their test values showed. Thus, they are highly proficient users of L2 German and use this language not only in the school context but also during the rest of the day with their friends and siblings, as an informal look at their questionnaires revealed. In line with many studies examining L3 phonology (e.g. Hammarberg, 2001; Lloyd et al., 2016; Wrembel, 2010), L2 proficiency is a likely candidate in order to account for L2 German transfer patterns in minority language students' L3 pronunciation of selected English sounds, in particular because the students mentioned in a questionnaire that L2 German was usually their dominant language.

However, it cannot be ruled out that the typological similarities between L2 German and L3 English may also have facilitated L3 learning at the phonological level. Such effects have been reported in other studies (e.g. Bouchhioua, 2016; Llana et al., 2010; Wrembel, 2010, 2012) but further studies are necessary to disentangle effects of L2 proficiency and typology by systematically comparing larger groups of speakers with various language backgrounds (e.g. Arabic, Turkish, including also other family languages such as Swahili or Urdu). It would also be interesting to examine L2s that are typologically closer to L3 English than L2 German (e.g. Frisian) or learners with L2s that are typologically closer to their L1 (e.g. different varieties of Arabic). In such studies, it could be determined whether a linguistic system as a whole, a phonological system as a whole, or single properties across languages are transferred from one language to the next (e.g. Cabrelli Amaro, 2012).

As regards the different models that have been proposed in order to account for transfer effects in L3 phonology, our results cannot be used to support any of the L3 phonology models, because we did not investigate initial state learners (or adult learners). In addition, we only considered selected L3 sounds of English and did not include the minority language learners' L1 to a sufficient extent in order to be able to prove or disprove any model of transfer in L3 phonology.

#### **4.2 Teachers vs. students**

In the present study, the teachers showed significantly faster (i.e. more native-like) speech rates and better phonemic accuracy in their English pronunciation than both groups of students: For example, [ʃ], /r/, /ð/ and /θ/ were produced almost always in a target-like way by the teachers. Even

voiced obstruents in word-final position did not pose a great problem for the English teachers who correctly produced these sounds in 80% of all instances.

Some inconsistencies, however, remain: Only two thirds of all /æ/ sounds were produced in a target-like way by the teachers, indicating that even these experienced learners are still in the process of forming a distinct phonetic category for English /æ/ (e.g. Bohn & Flege, 1992). These examples of mispronunciations may also have an impact on the students' pronunciation of English sounds because the teachers are their main source of foreign language input (e.g. Böttger, 2005; Piske, 2008; Yani, 2012, Kanoksilapatham, 2014). In other words, students' problems with voiced obstruents in word-final position or with /æ/ (see e.g. section 4.1. and 4.2) may not only be due to transfer patterns from their L1/L2 German (i.e. learner-inherent) but also to their teachers who provide them with input which is not native-like regarding these sounds. However, as the sample is very small, additional studies are warranted in order to examine the relationship between teachers' and students' pronunciation of English in the foreign language classroom in more detail.

In summary, the results of the present study suggest that the teachers are fairly adequate role models for their students in terms of their English accuracy and fluency. Furthermore, it has been reported that English learners seem to prefer a teacher who is easier to understand (i.e. one with the same language background), rather than one with a native accent (e.g. Braine, 2010 but see Butler, 2007 for different results). As previous research (e.g. Levis, 2005) indicates, the curricula for English as a foreign language nowadays rather emphasize intelligibility than nativeness in the foreign language classroom anyway, so that it is not regarded as problematic to let non-native qualified English teachers teach subject content as long as their competence in English is at least near-native like (Böttger, 2005; Piske, 2008; Kanoksilapatham, 2014; Karakaş, 2012; Yani, 2012).

### **4.3 Role of orthography**

In contrast to extemporaneous speech, pronunciation errors in reading-aloud data may also be "orthography induced" as a consequence of a mismatch between L1 and L2 grapheme-phoneme conversion rules. For example, Rallo Fabra & Jacobs (2015) reported their learners made fewer vowel errors when the target words had more transparent spellings and were closer to Spanish-Catalan phoneme-grapheme conversion rules, suggesting that in such cases, the learners had relied on orthography (see

also Piske, Flege, MacKay & Meador, 2002). In our sample, we only found a few instances of pronunciation errors that appeared to be due to grapheme-phoneme discrepancies in English, or German and English: These include instances of *come* (often realized as [kɔm] as in German *kommen*), *said* ([sɛɪd]), *ran* ([ɾʌn] pronounced as German /a/ as in <an>). Phonological coding (i.e. the recoding of written, orthographic information into a sound based code, e.g. Leininger 2014) as a source of error occurred only for the unknown words <pretty> and <laughed>, which were realized as [pɹɛti]) and [laʊɡed]. The last two words were apparently not familiar to the students who evidently resorted to the more familiar German phoneme-grapheme correspondences to read these words aloud. In line with Rallo Fabra & Jacobs (2015), it indeed seems to be easier for students in reading aloud tasks to pronounce English words in a target-like way if they are spelt transparently.

#### 4.4 Future studies

As this study only included a small sample of majority and minority language students and their teachers, there is a dire need of studies examining L1 and L2 effects in L3 acquisition with larger samples. It would be particularly interesting for the school context to also include students in mainstream programs in which English is taught as a subject for only 1-2 lessons per week with teachers who are not always qualified English teachers as it is still often the case, for example, in elementary schools in Germany. Moreover, many previous studies included foreign accent ratings obtained from native speakers of English (e.g. Lloyd et al., 2016), which would be a possibility to also evaluate global accuracy of our majority and minority language students' English pronunciation. Finally, an interesting question is whether Flege's SLM (e.g. 1995) could also be extended to L3 phonological acquisition, taking into account the acoustic properties of L1, L2 and L3 sounds as well as students' and teachers' perception of L2/L3 sounds in order to examine how L2 and L3 phonetic categories shift towards native-like categories in the course of acquisition. In times in which a steadily increasing number of people develop a multilingual competence inside and outside of the foreign language classroom it will become more and more important for language acquisition research to focus on the acquisition of more than two languages and to examine in detail the processes underlying and the factors affecting multilingual development.

## 5. Acknowledgements

The authors would like to express their gratitude to two reviewers whose insightful comments greatly improved this manuscript. All errors are, of course, entirely attributable to the authors. In addition, the support received by the colleagues of the elementary school in Germany is gratefully acknowledged. Most importantly, this study would not have been possible without the children's enthusiastic participation.

## References

- Bardel, C. & Falk, Y. (2007). The role of the second language in third language acquisition: the case of Germanic syntax. *Second Language Research*, 23(4), 459-484. doi: 10.1177/0267658307080557.
- Boersma, P. & Weenink, D. (2013). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.05. <http://www.praat.org/>
- Bohn, O.-S. & Flege, J.E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, 14, 131-158.
- Böttger, H. (2005). Englische Ausspracheschulung: Aspekte eines didaktisch-methodischen Designs für die Grundschule. *Neusprachliche Mitteilungen aus Wissenschaft und Praxis*, 58(3), 33-39.
- Braine, G. (2010). *Nonnative speaker English teachers. Research, pedagogy and professional growth*. New York: Taylor & Francis.
- Bouchhioua, N. (2016). Cross-linguistic influence on the acquisition of English pronunciation by Tunisian EFL learners. *European Scientific Journal* 12(5), 260-278. doi: 10.19044/esj.2016.v12n5p260.
- Butler, G. Y. (2007). How are non-native English-speaking teachers perceived by young learners? *TESOL Quarterly*, 41(4), 731-755.
- Cabrelli Amaro, J. (2012). L3 phonology: An understudied domain. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 33-60). Amsterdam: John Benjamins.
- Cenoz, J. (2001). The effect of linguistic distance, L2 status and age on cross-linguistic influence in third language acquisition. In J. Cenoz, B. Hufeisen & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 8-20). Clevedon: Multilingual Matters
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching and assessment*. Language Policy Unit: Strasbourg
- De Angelis, G. (2007). *Third or additional language acquisition*. Clevedon, UK: Multilingual Matters.



- Eckert, H. & Barry, W. (2002). *The phonetics and phonology of English pronunciation*. Trier: Wissenschaftlicher Verlag Trier
- FAZ (Frankfurter Allgemeine Zeitung, 2015). Nachhilfe für den Englischlehrer (Marvin Milatz). Retrieved from <http://www.faz.net/aktuell/beruf-chance/campus/ostdeutsche-altlasten-nachhilfe-fuer-den-englischlehrer-13605781.html>
- Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233-277). Baltimore: York Press
- Flynn, S., Foley, C. & Vinnitskaya, I. (2004). The cumulative-enhancement model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *International Journal of Multilingualism* 1(1), 3-16. doi: 10.1080/14790710408668175
- FMKS, Verein für frühe Mehrsprachigkeit an Kindertagesstätten und Schulen (2014). Ranking: bilinguale Kitas und Grundschulen im Bundesvergleich. Retrieved from <http://www.fmks-online.de/download.html>.
- Gabriel, C. & Rusca-Ruths, E. (2014). Der Sprachrhythmus bei deutsch-türkischen L3- Spanischlernern: Positiver Transfer aus der Herkunftssprache? In S. Witzigmann & J. Rymarczyk (Eds.), *Mehrsprachigkeit als Chance. Herausforderungen und Potentiale individueller und gesellschaftlicher Mehrsprachigkeit* (pp. 185-204). Frankfurt a.M.: Lang.
- Gut, U. (2010) Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, 13(4), 19-38. doi:10.1080/14790710902972248
- Hahn, A. & Angelovska, T. (2017). Input-practice-output: A model for teaching L3 English after L2 German with a focus on syntactic transfer. In T. Angelovska & A. Hahn (Eds.), *L3 syntactic transfer: Models, new developments and implications* (pp. 299-319). Amsterdam: Benjamins.
- Hammarberg, B. (2001). The roles of L1 and L2 in L3 production and acquisition. In J. Cenoz, B. Hufeisen & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 21-41). Clevedon, UK: Multilingual Matters.
- Harada, T. (2007). The production of voice onset time (VOT) by English-speaking children in a Japanese immersion program. *International Review of Applied Linguistics*, 45, 353-378.
- Hart, D., Lapkin, S. & Swain, M. (1987). *Early and middle French immersion programs: Linguistic outcomes and social character*. Toronto: Modern Language Centre, Ontario Institute for Studies in Education.
- International Phonetic Association (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
- Kanoksilapatham, B. (2014). Thai elementary school teachers' English pronunciation and effects of teacher variables: Professional development. *TESL-EJ* 18(1), 1-13. Retrieved from <http://www.tesl-ej.org/pdf/ej69/a2.pdf>.

- Karakaş, A. (2012). Foreign accent problem of non-native teachers of English. *Humanising Language Teaching*, 14(5), Retrieved from <http://www.hltmag.co.uk/oct12/mart06.htm>.
- König, E. & Gast, V. (2012). *Understanding English-German contrasts* (3rd edition). Berlin: Schmidt.
- Kopečková R. (2013). Crosslinguistic influence in instructed L3 child phonological acquisition. In M. Pawlak & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism. Second language learning and teaching* (pp. 205-224). Cham; Springer.
- Leininger, M. (2014). Phonological coding during reading. *Psychological Bulletin*, 140(6), 1534-1555. doi.org/10.1037/a0037830.
- Levis, J.M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369-377. doi.org/10.2307/358848
- Llama, R., Cardoso, W. & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism*, 7(1), 39-57. doi: 10.1080/14790710902972255
- Llisteri, J. & Poch, D. (1987). Phonetic interference in bilingual's learning of a third language. In T. Gamkrelidze (Ed.), *Proceedings of the Eleventh International Congress of Phonetic Sciences. Tallinn, Estonia, 1-7 August 1987*, Vol. 5 (pp. 134-137). Estonia: Academy of Sciences of the Estonian SSR.
- Lloyd-Smith, A., Gyllstad, H. & Kupisch, T. (2017). Transfer into L3 English global accent in German-dominant heritage speakers of Turkish. *Linguistic Approaches to Bilingualism*, 7(2), 131-162.
- Llurda, E. (Ed.). (2005). *Non-native language teachers: Perceptions, challenges, and contributions to the profession*. New York: Springer Science and Business Media.
- Megdyes, P. (2013). Non-native teachers. In: M. Byram & A. Hu (Eds.), *Routledge encyclopedia of language teaching and learning. Second edition* (pp. 508-511). London & New York: Routledge.
- OECD (2016). *PISA 2015 results: Excellence and equity in education*. PISA: OECD Publishing
- Pérez-Cañado, M. L. (2012). CLIL research in Europe: Past, present and future. *International Journal of Bilingual Education and Bilingualism*, 15(3), 315-341.
- Piske, T. (2008). Phonetic awareness, phonetic sensitivity and the second language learner. In J. Cenoz & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2<sup>nd</sup> edition), Vol. 6: Knowledge about language (pp. 155-166). New York: Springer
- Piske, T. (2015). Zum Erwerb der CLIL-Fremdsprache. In B. Rüschoff, J. Sudhoff, & D. Wolff (Eds.), *CLIL Revisited: Eine kritische Analyse zum gegenwärtigen Stand des bilingualen Sachfachunterrichts* (pp. 101-125). Frankfurt a. M.: Lang.
- Piske, T., Flege, J. E., MacKay, I.R.A. & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica*, 59, 49-71.

- Rallo Fabra, L. & Jakob, K. (2015). Does CLIL enhance oral skills? Fluency and pronunciation errors by Spanish-Catalan learners of English. *Educational Linguistics*, 23, 162-177.
- Reyes, A. M., Arechabaleta-Regulez, B. & Montrul, S. (2017). The acquisition of rhotics by child L2 and L3 learners. *Journal of Second Language Pronunciation*, 3(2), 242-266.
- Ringbom, H. (1987). *The role of the first language in foreign language learning*. Clevedon, UK: Multilingual Matters.
- Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research*, 27(1), 107-127.
- Rothman, J. (2015). Linguistic and cognitive motivations of the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism Language and Cognition*, 18(2), 179-190. doi: 10.1017/S136672891300059X.
- Schwartz, B. & Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12, 40-72. doi: 10.1177/026765839601200103.
- Smith, B. L., Hayes-Harb, R., Bruss, M. & Harker, A. (2009). Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics*, 37(3), 257-275.
- Steinlen, A. K. (2005). *The influence of consonants on native and non-native vowel production*. Tübingen: Narr.
- Steinlen, A. K. (2016). Primary school minority and majority language children in a partial immersion program: The development of German and English reading skills. *Journal of Immersion and Content-Based Language Education*, 4(2), 198-224.
- Steinlen, A. K. (2018). The development of English and German writing skills in a bilingual primary school in Germany. *Journal of Second Language Writing*, 39(1), 42-52.
- Steinlen, A. K. & Piske, T. (2013). Academic achievement of children with and without migration backgrounds in an immersion primary school: A pilot study. *ZAA Zeitschrift für Anglistik und Amerikanistik. A Quarterly of Language, Literature and Culture*, 61(3), 215-244.
- Stifterverband für die Deutsche Wissenschaft (2013). *Hochschulbildungsreport 2020*. Retrieved from <http://www.stifterverband.de/bildungsinitiative/hochschulbildungsreport.pdf>.
- Süddeutsche Zeitung (2012). Grundsüler – Lost in Translation (T. Baier). Retrieved from <http://www.sueddeutsche.de/karriere/englischunterricht-fuer-sechsjaehrige-grundschueler-lost-in-translation-1.1044176>.
- Wesche, M. B. (2002). Early French immersion: How has the original Canadian model stood the test of time? In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development* (pp. 357-379). Trier: Wissenschaftlicher Verlag Trier.

- Wiederholt, J. L. & Bryant, B. R. (2001). *Gray Oral Reading Tests*. Austin, TX: pro-ed. 4th edition.
- Williams, S. & Hammarberg, B. (1998). Language switches in L3 production: Implications for a polyglot speaking model. *Applied Linguistics*, 19(3), 295-333.
- Wode, H. (2009). Developing non-native pronunciation in immersion settings. In T. Piske & M. Young-Scholten (Eds.), *Input Matters in SLA* (pp. 238-256). Bristol: Multilingual Matters.
- Wrembel, M. (2010). L2-accented speech in L3 production. *International Journal of Multilingualism*, 7(1), 75-90.
- Wrembel, M. (2012). Foreign accentedness in third language acquisition. In J. Cabrelli Amaro, S. Flynn & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 281-310). Amsterdam: John Benjamins. doi: 10.1075/sibil.46.16wre
- Yani, A. (2012). Teachers' incorrect pronunciation and its impact on young learners: (A review of linguistic aspects of EFL classroom practices). In M. Syafe & H. Madjdi (Eds.). *Teaching English for young learners in Indonesia (TEYLIN): From Policy to Classroom* (pp. 179-189). Retrieved from [http://eprints.umk.ac.id/340/25/PROCEEDING\\_Teylin\\_2.185-195.pdf](http://eprints.umk.ac.id/340/25/PROCEEDING_Teylin_2.185-195.pdf).



## PAM-L2 and Phonological Category Acquisition in the Foreign Language Classroom

Michael D. Tyler  
Western Sydney University

### Abstract

Models of second language (L2) speech learning are designed to account for phonological acquisition in the L2. Both the Perceptual Assimilation Model of L2 speech learning (PAM-L2; Best & Tyler, 2007) and the Speech Learning Model (SLM; Flege, 1995, 2003) have based their predictions on L2 acquisition by immersion in a predominantly L2 environment. However, many second languages are learned via formal instruction in a foreign language (FL) classroom, often in the learner's native language environment. Piske (2007) outlined how the principles of SLM might apply to the FL classroom, concluding that formal instruction should begin at an early age, there should be intensive foreign language use over an extended period of time, learners should have exposure to high quality input, and there should be training focused specifically on perception and production. The aim of this paper is to explore how the principles of PAM-L2 might complement those suggestions. The paper provides a thorough overview of PAM-L2, before outlining key characteristics of FL learning in the classroom that are likely to impact on L2 category acquisition, either positively or negatively. It also discusses methodological factors to be taken into consideration for any study investigating L2 category acquisition from a PAM-L2 perspective. Applying PAM-L2 to the FL classroom, the paper concludes that FL students need learning experiences that provide opportunities for them to discover the phonetic differences that signal phonological contrast in the L2. These experiences need to be provided at the earliest possible stages of learning, prior to the establishment of a large L2 vocabulary. A range of suggestions is provided for how PAM-L2 principles might be incorporated into FL learning curricula to maximise the opportunity for acquiring sensitivity to L2 phonological distinctions.

---

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 607-630). Dept. of English, School of Communication & Culture, Aarhus University.

© The author(s), 2019.

## 1. Introduction

The outcome of second-language (L2) acquisition in childhood is markedly different from adult L2 acquisition, particularly in the domains of phonetics and phonology. Older learners are much more likely than younger learners to speak with a detectable foreign accent. Less obvious to casual observation is the fact that late L2 learners are also likely to *hear* with an accent (Jenkins, Strange, & Polka, 1995). While some have attempted to explain these differences between early and late L2 acquisition as biological in nature, the most plausible explanation seems to be that the effects are due to the fact that the learner already has a first language (L1). Indeed, results from studies on cross-language speech perception, where listeners are presented with stimuli from a never-before-heard non-native language, consistently show a profound influence of the L1 on the perception of non-native phones. It comes as no surprise, then, that the L2 learner's perception and production is heavily influenced by prior learning of the L1.

The two most influential models of how the L1 influences L2 speech learning are the *Speech Learning Model* (SLM; Flege, 1995, 2003) and the *Perceptual Assimilation Model of Second Language Speech Learning* (PAM-L2; Best & Tyler, 2007). SLM was designed to provide a framework for predicting the likelihood of acquiring new phonetic categories in the L2, and it applies to both L2 production and perception. PAM-L2, on the other hand, is concerned with perception only. Both models assume a learner with no prior knowledge of the L2, who is acquiring the L2 by immersion in an L2-dominant environment. However, many people successfully acquire communicative competencies in a formal instruction setting in a predominantly L1 environment. Finding optimal conditions for L2 category acquisition is important for designing foreign-language (FL) curricula, but neither model was designed with a formal instruction setting in mind. Here, the term L2 acquisition is reserved for L2 learning in an immersion setting, and *foreign language acquisition* (FLA) is used for a classroom setting. Both models can be applied to FLA, in principle, but it may be more difficult to make clear predictions about category acquisition because students come to the learning situation with a varying degree of prior experience with L2 and classrooms differ in the degree of native-speaker input received by students. Nevertheless, with carefully designed and well-controlled studies it may be possible to test general predictions in a classroom FL context. Piske (2007) has already outlined how the principles of SLM might apply to classroom FLA. He concluded

that FLA should begin in school at a young age, provided that there is high-quality L2 input, opportunities for intensive language use over a period of years, and that curricula should include specific training in perception and production of L2 phonemes. This chapter will focus on how the principles of PAM-L2 might complement the suggestions already made on the basis of SLM. For a general comparison of PAM-L2 and SLM, see Best and Tyler (2007). Specifically, the aims of this chapter are to:

- 1) Outline how the theoretical principles underlying PAM-L2 might apply to classroom FLA;
- 2) Identify methodological requirements for investigating L2 category acquisition in a classroom context, and;
- 3) Suggest possible avenues for incorporating PAM-L2 principles into FL learning curricula.

## **2. The Perceptual Assimilation Model of Second Language Speech Learning (PAM-L2)**

To be able to explain how PAM-L2 might be applied to FLA, first it is necessary to summarize the model. As it has been over 10 years since the publication of Best and Tyler (2007), this also provides an opportunity to elaborate on the principles of the model using more recent experimental findings.

PAM-L2 is based on the Perceptual Assimilation Model (PAM; Best, 1993, 1994, 1995). PAM was designed to account for how the native language shapes perception of consonants and vowels. Research on infant speech perception informs PAM on how a native phonology develops (e.g., Best & McRoberts, 2003; Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Tyler, Best, Goldstein, & Antoniou, 2014), and research on adult cross-language speech perception provides evidence for how prior learning of the native language influences perception (e.g., Best, McRoberts, & Goodell, 2001; So & Best, 2014; Tyler, Best, Faber, & Levitt, 2014). To test this influence, adults who are functional monolinguals are presented with contrasting phones from a never-before-heard non-native language. The participants of cross-language speech perception studies are not actively trying to learn to communicate in that language. Rather, the non-native language is used as a tool to probe the influence of native-language tuning on speech perception. PAM-L2 takes the functional monolingual as its starting point and assumes a learner who is actively acquiring an L2 in an environment where the L2 is predominantly spoken (i.e., via immersion). PAM-L2 assumes that the perceptual system is shared by all



of the learner's languages. If certain L1 phonological categories function adequately for discriminating L2 contrasts, then no additional learning is required for those contrasts. On the other hand, if the learner does not detect an L1 contrast for a given pair of L2 phonemes, then perceptual learning is required to be able to detect the L2 phonological contrast, and to build an L2 vocabulary that preserves a phonological distinction between those phonemes. How successful the learner is at detecting new phonological contrasts in the L2 is dependent on how the L2 phonemes are initially assimilated to the L1 phonological system.

PAM considers perceptual assimilation to the native phonology in two ways. First, an individual non-native phone can be categorized as a good, acceptable, or poor instance of a native category, uncategorized (i.e., it is not perceived as an instance of any one native category), or non-assimilable (i.e., it is not perceived as speech). As explained by Best (1994, pp. 261-262), the information that defines phonological category membership is only one small part of the L1 phonology, and it may differ qualitatively from the information that defines the systematic relationships *between* categories in a phonological system. Thus, PAM also considers perceptual assimilation of pairs of contrasting non-native phones and makes predictions about discrimination on the basis of the contrast assimilation type. If the non-native phones are each assimilated to a different L1 phonological category then this is termed a two-category assimilation. Discrimination is expected to be excellent because, serendipitously, the perceiver is able to detect an L1 phonological contrast between the non-native phones. When both non-native phones are assimilated to the same native category, there is no L1 phonological contrast to support discrimination, but the perceiver may be sensitive to differences in perceived phonetic goodness-of-fit to the native phonological category. If one of the non-native phones is perceived as a more acceptable instance of the L1 category than the other non-native phone, then it is a category-goodness assimilation and discrimination is predicted to be very good. If there is no difference in perceived phonetic goodness-of-fit between the two non-native phones then it is a single-category assimilation, and discrimination is predicted to be poor. For contrasts where one phone is categorized and the other is uncategorized, an uncategorized-categorized assimilation, discrimination is predicted to be very good. Discrimination should vary from poor to very good for uncategorized-uncategorized assimilations, depending on their phonetic proximity to one another and the perceived similarity to sets of native phonological categories.

A recent study has shown that non-native phones can be uncategorized in three different ways (Faris, Best, & Tyler, 2016). Focalized phones are perceived as weakly consistent with only one L1 category and clustered phones are those that are perceived as weakly consistent with multiple L1 categories. Dispersed phones are perceived as speech but are not perceived as similar to any native category. For uncategorized-categorized and uncategorized-uncategorized assimilations involving focalised and clustered phones, discrimination varies as a function of perceived phonological overlap between the sets of native categories that are weakly consistent with them, such that it is more accurate for non-overlapping contrasts than partially overlapping contrasts (Faris, Best, & Tyler, 2018). For contrasts involving dispersed phones, discrimination should be excellent for uncategorized-categorized assimilations, but for uncategorized-uncategorized assimilations discrimination should vary according to their phonetic proximity. It is only in that latter case that phonological learning from the native language would have minimal influence on discrimination. These discrimination predictions for dispersed phones are yet to be tested experimentally.

PAM-L2 uses PAM contrast assimilation types as a basis for predicting the likelihood of acquiring new L2 categories when a learner is actively acquiring the non-native language. Discrimination should improve when the contrast assimilation type changes as a result of new category acquisition (e.g., a category-goodness assimilation becomes a two-category assimilation). Best and Tyler (2007) clarified that perceptual learning could take place at multiple levels of attention focus, for example, phonological, phonetic, and gestural (see Strange, 2011, for complementary ideas about the role of attention in speech perception). For example, when each L2 phoneme in a contrast is perceived as a different L1 category (a PAM two-category assimilation), prior learning of an L1 phonological contrast serves for discrimination in the L2. Once learners begin to acquire an L2 vocabulary using those categories, they will have developed a common L1-L2 phonological category for each. If there is a discernible phonetic difference between the L1 and L2 versions, the perceived phonetic differences between them may become sharper over time. If the L1 and L2 version come to occupy separate regions of phonetic space within that L1-L2 phonological category, then the learners will have established separate L1 and L2 phonetic categories as part of a common L1-L2 phonological category. On the other hand, if the L1 and L2 versions are sufficiently similar to each other phonetically, then the learners will establish instead

a common L1-L2 phonetic category in both phonological *and* phonetic terms. Recall that in the case of a PAM two-category assimilation, the learner does not need to acquire any new phonological contrast in the L2. The likelihood of acquiring a new phonological category is low, in that case, because sensitivity to phonological contrast between existing L1 phonological categories serves perfectly well in the L2.

The cases presented by Best and Tyler (2007) focused on individual contrast assimilations (e.g., two category, category goodness) to show how predictions can be made for L2 learners on the basis of cross-language perceptual assimilation by naïve perceivers. For the sake of simplicity, the same approach will be taken here. It is important to note, however, that a phonology consists of systemic relationships between all phonological categories. Successful acquisition of a given L2 contrast shows that the learner can detect an L2 phonological distinction, but to be able to conclude that a new L2 category has been acquired it is necessary to establish that the new L2 phonological category: 1) forms two-category assimilations with all other L2 phonemes, and; 2) does not form a common phonological category with any L1 phoneme. For this reason, we advocate taking a whole-system approach to the study of L2 speech learning (Bundgaard-Nielsen, Best, Kroos, & Tyler, 2012; Bundgaard-Nielsen, Best, & Tyler, 2011a, 2011b), particularly for vowels, where participants are given the opportunity to categorise all of the vowel phonemes from the L2, and where they are provided with all possible vowel labels in the categorisation task (see, e.g., Faris et al., 2016, 2018).

Since PAM and PAM-L2 define phonological differences as those that are relevant to discriminating minimally contrasting lexical items, L2 vocabulary size is likely to play a key role in guiding perceptual learning in the L2 (for studies on lexically guided perceptual retuning see, Kraljic & Samuel, 2006; McQueen, Tyler, & Cutler, 2012; Norris, McQueen, & Cutler, 2003). Best and Tyler (2009) suggested that the window of opportunity for perceptual attunement may be quite early in acquisition, prior to the establishment of a large L2 vocabulary. According to *The Vocabulary-Tuning Model of L2 Rephonologization* (Bundgaard-Nielsen et al., 2011a), an increasing vocabulary drives perceptual reattunement to the L2 phonology. This idea was supported by Bundgaard-Nielsen, Best, and Tyler (2011a, 2011b), who showed that learners with larger vocabularies had more consistent categorisation of L2 phonemes than learners with smaller vocabularies. More consistent categorisation was associated with more accurate discrimination for some assimilation

types (e.g., uncategorised-categorised), but not for others (e.g., single category). An increasing vocabulary may support perceptual learning for contrasts that are more discriminable (e.g., uncategorised-categorised and uncategorised-uncategorised), but inhibit perceptual learning for those are less discriminable (e.g., single category). This raises the question of how many words constitutes a large L2 vocabulary. For children learning their L1, vocabulary acquisition is slow up to around 50 words and rapidly increases thereafter (see, e.g., Nazzi & Bertoncini, 2003). This would seem to suggest that learners should also aim to maximise their opportunities for phonetic learning before the L2 vocabulary exceeds 50 words.

L2 learners are most likely to acquire a new L2 phonological category for contrasts where both L2 phonemes are assimilated to the same L1 phonological category but with a perceived difference in phonetic goodness-of-fit (a PAM category-goodness assimilation), or where at least one of the L2 phones is uncategorized (PAM uncategorized-categorized or uncategorized-uncategorized assimilations). For the category goodness case, Best and Tyler (2007) speculated that the more deviant sounding phone might first be established as an L2 phonetic variant of a common L1-L2 phonological space within that category (with the more acceptable phone likely residing within a common L1-L2 phonetic category). As the learners tune in to the phonetic difference between the L2 phonemes, they would recognise that the perceived phonetic difference signals a meaning difference between minimally contrasting L2 words and a new L2 phonological category would be developed. For uncategorized phones, Best and Tyler suggested that they should be relatively easy to acquire as new L2 phonological categories. However, if the L2 phonemes in an uncategorized-uncategorized assimilation are phonetically similar, it is possible that they might not be differentiated from each other. In that case, a new L2 phonological category might be established that encompasses both (undifferentiated) L2 phonemes.

When both L2 phonemes are assimilated as the same L1 category, but there is no difference in phonetic goodness-of-fit (a single-category assimilation), the learners are unlikely to acquire a new L2 phonological category. Both will be incorporated into an L1-L2 phonological category and an increasing vocabulary will reinforce the equivalence. While this may seem unoptimistic, single-category assimilations are likely to pose a particular type of difficulty for the L2 learner, even with high-quality native-speaker input. Single-category assimilations may involve cases where the L1 not only lacks a phonological distinction to assist the learner, but one

where the degree of acceptable phonetic variability of the L1 phonological category encompasses the phonological contrast in the L2. That is, certain phonetic differences serve a lexical/functional purpose (*phonological distinctiveness*), but phonological categories remain unchanged across other phonetic differences (*phonological constancy*; see Best, 2015; Best et al., 2009). Take, for instance, the bilabial plosive-implosive distinction, /b/-/ɓ/, which English native perceivers assimilate as a single-category assimilation across a number of different languages (Ma'di: Antoniou, Best, & Tyler, 2013; Zulu: Best et al., 2001; Sindhi: Fenwick, Best, Davis, & Tyler, 2017). Both [b] and [ɓ] are possible allophones of /b/ in English (Ladefoged & Johnson, 2014), so it is possible that English native perceivers have a single phonetic category that encompasses both allophones. Regardless of the explanation for why certain L2 phonemes are assimilated as single-category assimilations, it is clear that they are very difficult to discriminate. It is likely that targeted perceptual training would be needed to learn to detect the difference between L2 phonemes in a single-category contrast.

Finally, it is worth considering how the predictions for uncategorized phones might be refined by taking into account the new uncategorized assimilations proposed by Faris et al. (2016, 2018). For contrasts involving dispersed assimilations, the predictions would be unchanged from those suggested by Best and Tyler (2007) for assimilations involving uncategorized phones. Those involving focalised or clustered assimilations, however, may have a different developmental path. If a contrast includes one of those assimilations, it means that the perceiver detects phonological similarity, albeit weakly, between the focalised or clustered phone and one or more L1 phonological categories. The likelihood of establishing a new L2 phonological category would depend crucially on the degree of perceived phonological overlap (Faris et al., 2018) between the L2 phonemes of a contrast, such that non-overlapping and partially overlapping contrasts are more likely to form separate L2 categories than completely overlapping contrasts. There is not enough experimental data available to predict specific outcomes for all contrasts involving at least one focalised or clustered phone, but it is possible to outline the range of possible outcomes. They are presented in Table 1. Note that in the cases where an uncategorised L2 phoneme is acquired as a common L1-L2 category, that would only be an optimal outcome if no other L2 phoneme had been acquired as the same L1 category.

Uncategorised-Categorised	Uncategorised-Uncategorised
<ul style="list-style-type: none"> <li>• The uncategorised L2 phoneme is acquired as a new L2 phonological category</li> <li>• The uncategorised L2 phoneme is acquired as a common L1-L2 phonological category with an L1 phoneme that is different from the categorised member of the pair</li> <li>• The uncategorised L2 phoneme forms a common L1-L2 category with the categorised member of the pair</li> </ul>	<ul style="list-style-type: none"> <li>• Each uncategorised L2 phoneme is acquired as a new L2 phonological category</li> <li>• A new L2 phonological category is acquired incorporating both L2 phonemes</li> <li>• Both L2 phonemes are acquired as a common L1-L2 phonological category with a single L1 phoneme</li> <li>• Each L2 phoneme is acquired as a common L1-L2 phonological category with a different L1 phoneme</li> <li>• One L2 phoneme is acquired as a new L2 phonological category and the other is acquired as a common L1-L2 phonological category</li> </ul>

Table 1. Possible category acquisition outcomes for contrasts involving at least one focalised or clustered uncategorised L2 phoneme.

To summarise, for PAM-L2 the likelihood of L2 phonological category acquisition is crucially dependent on how pairs of L2 phonemes are assimilated to the L1 phonological system. To have the opportunity to tune in to the phonetic differences that signal phonological contrast in the L2, learners need input that preserves those differences, and perceptual learning needs to occur prior to the establishment of a large L2 vocabulary. In the next section some of the characteristics of classroom FLA that may impact on L2 category acquisition, from a PAM-L2 perspective, will be outlined.

### 3. Factors affecting phonological category acquisition in the FL classroom

The principles of PAM-L2 were illustrated using the idealised situation of a learner previously naïve to the L2 in an immersion environment. Such a learning situation may be rare in the modern age, especially when the target

language is English, and people living in a predominantly L2 environment can vary considerably in their degree of exposure to the L2 on a daily basis (Flege & MacKay, 2004). It is useful, therefore, to consider the ways in which models such as PAM-L2 might be applied to other possibly more common L2 learning situations, such as classroom FLA. Broadly, the conditions for optimal phonological attunement, according to PAM-L2, are those where learners have an opportunity to tune in to L2 phonological contrasts prior to the acquisition of a large L2 vocabulary. The idealised L2 learner whose first and continuing exposure to the L2 is in an immersion environment with rich native-speaker input would have ample opportunity for the sort of perceptual learning that is required. However, classroom FL learners may not have the same opportunities and the context of classroom FLA may change the predicted outcomes. Here, three aspects of classroom FLA will be considered: spoken language input, written language input, and previous FL exposure. This will be followed by a reconsideration of PAM-L2 predictions as applied to classroom FLA.

### **3.1 Spoken language input**

In the FL classroom, interactions in the L2 are likely to be with foreign-accent speech. In most cases the teacher is likely to speak the target language as an L2, and with a non-native accent. Students in class will also speak to each other in class during activities, presumably with a non-native accent. This differs from the idealised immersion situation, where the learner is exposed to native-speaker input. Note that accented speech is not necessarily an impediment to the acquisition of new L2 categories. If the accented speech maintains a phonological distinction between all L2 phonemes, and native speakers unambiguously perceive them as intended, then the learners may acquire all necessary phonological distinctions in the L2. Their perception would be accented (Jenkins et al., 1995), but if a phonological distinction has been acquired, then the learners may be able to fine tune that distinction at some point in the future with exposure to rich L1 input. On the other hand, if the accented speech does not maintain certain phonological distinctions then this would clearly reduce the likelihood of learners acquiring them. Minimally contrasting words would be homophonous, this would be reinforced with an increasing vocabulary, and it may fossilize even if the learner is exposed to rich L1 input in the future. In short, it is not necessarily a lack of *native-speaker* input that may reduce the likelihood of L2 category acquisition in the classroom, but input that fails to provide clear phonological differences between L2 categories.

### 3.2 Written language input

Another key difference between L2 immersion and classroom FLA is the use of written language to teach vocabulary and grammar. Speech is transient, whereas the written word is permanent (Ehri, 1984, 1985), so written materials are an excellent resource for learners to revise and consolidate what has been learned in class. However, written materials provide the opportunity to acquire a large L2 vocabulary in a short period of time, and this may reduce the window of time available for perceptual learning of L2 phonological contrasts that are difficult to discriminate (e.g., single-category assimilations). Alphabetic writing systems or *orthographies*, also provide a (sometimes imperfect) representation of the phonology of a language. Each time learners read a word in the L2 they may be reinforcing a phonological structure for that word that is based on L1 grapheme-phoneme correspondences that have been adapted to the L2. For contrasts where learners can perceive a phonetic difference between the L2 phonemes, it is conceivable that alphabets might help learners to focus on and tune in to those phonetic differences in speech, as long as the L2 orthography signals a clear phonological difference. However, in cases where the orthography does not signal a clear phonological difference, their internal rehearsal of the pronunciation of L2 words via orthography may reinforce a perception that the L2 phonemes are equivalent rather than distinct.

### 3.3 Previous FL exposure

Whereas PAM-L2's predictions are based on an immersion learner with no previous experience with the L2, students in FL classrooms may vary greatly in their degree of previous exposure to the language. This could be in the form of prior classroom instruction, exposure to films and television, study abroad, family, or any number of other influences (Bohn & Bundgaard-Nielsen, 2009). They may also have learned to read the L2, especially if it shares the same writing system as the L1. If words had been learned in the absence of spoken input, the learner may already have applied the L1 phonology to a large vocabulary of L2 words via orthography. This is important for all models of L2 speech learning, but it is crucial for PAM-L2 because initial experience with the L2 sets the trajectory for perceptual learning. A learner with previous L2 experience may already have acquired a category-goodness contrast as a single common L1-L2 category, for example. Fossilisation may already have begun to occur, due to an increasing L2 vocabulary, making it more difficult to acquire the



less acceptable version as a new L2 phonological category than PAM-L2 would generally predict. Furthermore, predicting category acquisition is made even more difficult when learners have been exposed to a different range and variety of regional and foreign accents (Bohn & Bundgaard-Nielsen, 2009). Differences in prior FL experience are not only a problem for evaluating PAM-L2's category acquisition predictions in a classroom context. It is also important to consider from a pedagogical standpoint because learners may benefit from different classroom learning experiences as a function of their prior exposure.

### **3.4 PAM-L2 category acquisition predictions in an FLA context**

Taking these three aspects of classroom FLA into account, it is now possible to outline how the PAM-L2 predictions might change for the classroom FL context. There is no change to the predictions for two-category assimilations. L2 learners have all the phonological sensitivity they need from the L1 to learn vocabulary with L2 phonemes that form two-category assimilations with all other L2 phonemes. Category-goodness assimilations are less likely to be acquired in the classroom than via immersion, especially if there is little phonetic difference between the L2 phonemes in L2-accented spoken input, or the student internally rehearses the words on the basis of their written form. Perceptual learning of the phonetic distinction may be curtailed by rapid vocabulary acquisition, especially via the written medium. If single-category assimilations are unlikely to be acquired in the immersion case, then they are even less likely to be learned in the classroom! Acquisition of dispersed uncategorised L2 phonemes is unlikely to be inhibited by classroom FLA. It is even possible that they may be learned successfully, and very rapidly, if there is sufficient spoken input and the L2 orthography provides unambiguous information about the phonological contrasts involving that L2 phoneme. As it is not yet clear how clustered and focalised L2 phonemes are acquired in the immersion case, it is difficult to specify how classroom FLA would alter their development. It is possible that the reduced exposure to native speaker input in FLA versus L2 acquisition might increase the likelihood of acquiring the uncategorised L2 phoneme as a common L1-L2 phonological category, whereas the provision of unambiguous orthographic information might provide a focal point for the learner to tune in to the phonetic properties and establish a new L2 phonological category.

#### **4. Testing phonological category acquisition in the FL classroom**

There are certain methodological requirements for testing PAM-L2 predictions. As with any study testing PAM, it is essential to include a categorisation task with goodness rating. For example, in cross-language speech perception studies, participants are usually presented with a consonant or vowel in a carrier frame (e.g., consonant + /a/ or /h/ + vowel + /bə/) and they select a native category label that best matches the non-native phone. Next, they rate how acceptable a version the non-native phone is of the category that they chose. Without including the goodness rating step, it is not possible to distinguish between single-category and category-goodness assimilations. To test PAM-L2 predictions, it is necessary to determine whether a new L2 phonological category has been acquired, but Best and Tyler (2007) did not address the question of how category acquisition can be determined. As there is no direct link between perception and production for PAM-L2, learners' productions of the L2 category does not necessarily provide accurate information about whether that category has been acquired in perception. For example, an L2 learner could be trained to articulate a pair of contrasting L2 phonemes without necessarily being able to discriminate them. In recent work in our lab, we have had participants complete two categorisation tasks with the same L2 speech stimuli, one using L1 labels and the other using L2 labels, (Faris, Best, & Tyler, in preparation; San & Tyler, in preparation). By comparing categorisation across the two tasks it should be possible to infer whether a new category has been acquired. For example, an L2 phoneme that is uncategorised in the L1 and categorised in the L2 would seem to be a clear case of L2 category acquisition. We chose to use separate categorisation tasks for each language, rather than a single task with labels from both languages, because results of a single task would be difficult to interpret. If an L2 phoneme has been acquired as a common L1-L2 category, a given participant may only ever choose either the L1 or L2 label, or a mixture of both. Categorising in terms of L2 labels only would also be inadequate because the researcher would be unable to exclude the possibility that all L2 phonemes had been acquired as common L1-L2 categories. Ideally, tests of categorisation should be accompanied by tests of discrimination. If an L2 category has been acquired, and, for example, a contrast that was category goodness at the initial stage of learning has become a two-category assimilation, then discrimination should have improved.

One challenge with this approach is that categorisation is a metalinguistic task. Labelling in terms of L1 phonological categories requires phonemic awareness, which is learned as a by-product of alphabetic literacy (Ehri, 1984, 1985; Read, Zhang, Nie, & Ding, 1986; Tyler & Burnham, 2006). Performing a categorization task can be challenging even in the L1, especially for vowel labels in English where the orthography does not provide a one-to-one mapping between graphemes and phonemes (see, e.g., Faris et al., 2018; for a discussion, see Flege, 2003). Asking L2 learners to perform a metalinguistic task with L2 labels is even more challenging, and the data obtained may be inherently noisier than data obtained using L1 labels. Clearly, learners cannot perform an L2 categorisation task until they reach fairly high levels of proficiency in the L2. To be able to make any conclusions about category acquisition in a cross-sectional study, it is essential to include a control group of naïve listeners to establish a baseline for perceptual assimilation among the L1-speaking population. Studies could employ a categorisation task using L1 labels for the control group, to determine perceptual assimilation patterns prior to L2 learning, and a combination of L1 and L2 categorization tasks with the learner group to probe L2 category acquisition.

It could be argued that differences or changes in discrimination accuracy could be used instead as evidence for L2 category acquisition. Improvements in discrimination are certainly evidence of learning, but they do not provide direct evidence for category acquisition. For example, an improvement in discrimination could occur because a new L2 category has been formed for both L2 phonemes in an uncategorised-uncategorised assimilation, or because both were acquired as common L1-L2 categories with different L1 phonological categories. Without a categorisation task, it is not possible to differentiate these alternatives. For some researchers and educators, showing an improvement in discrimination may be sufficient, especially to evaluate a targeted classroom intervention. However, the benefit of using a theoretical framework, such as PAM-L2, is that it can *explain* why certain contrasts are easier to learn than others. Classroom interventions based on PAM-L2 would be focused on supporting the detection of new L2 phonological contrasts, so that all L2 phonological contrasts become two-category assimilations, which should then lead to improved discrimination. Without a theory that links perceptual learning of phonological contrast with discrimination accuracy, it may be difficult to explain why discrimination of certain L2 contrasts improves more than others. If using PAM-L2 as a framework, researchers should aim to include

a range of perceptual assimilation types (e.g., single category, category goodness, and uncategorised-categorised) to provide opportunities for observing differences in category acquisition, and variable improvements in discrimination.

It is also necessary to account for each learner's experience with the L2 prior to participating in the experiment. To overcome the problem of learners' prior experience, the best solution is to include only those participants with no prior exposure to any language other than the L1 (i.e., a functional monolingual), and to follow their language development longitudinally. However, such strict inclusion criteria may be an impediment to many researchers, especially those working in English FL classrooms. Some researchers may choose to compromise by including participants without prior formal instruction or immersion experience, but who may have had incidental exposure through television, film, or a short holiday in an environment where the L2 is spoken. Others may have no choice but to include participants with a range of prior learning experiences. The more prior experience the participants have had with the L2, the greater the requirement to also take into account covariates that might affect the initial state of the L2 phonology prior to formal instruction (e.g., extended time spent in a country where the L2 is spoken, L2 vocabulary size, watching television or movies in the L2 that have not been dubbed). Even if covariates are taken into account in statistical analyses, it is important to acknowledge the consequences of relaxing the inclusion criteria for testing theoretical predictions. Comparisons of learner groups and monolingual controls, or performance over time in a single learner group, are likely to be heavily influenced by experience of the learners prior to entering the classroom. Failure to observe category acquisition could be due to fossilisation, and successful acquisition in class could equally be due to the perceptual learning trajectory that was set by prior L2 exposure. Some of the variability in the initial state of the learner may be taken into account by forming subgroups of participants based on their individual perceptual assimilation patterns (see, e.g., Tyler, Best, Faber, & Levitt, 2014). That is, if some participants perceive a given contrast as a single-category assimilation, while others perceive it as category goodness, then the latter subgroup might be more likely than the former to acquire a new L2 category. Should that turn out to be case, then this would possibly open up new avenues for assessing students prior to formal instruction, and tailoring learning experiences to their specific needs.

## 5. Possible ways to incorporate PAM-L2 principles into FL learning curricula

The detection of phonological contrast is important for communication in the L2. Just as intelligibility is important for communication in L2 production (Munro, 2008), the inability to distinguish L2 phonemes can have implications for processes of L2 word recognition that extend beyond the homophony of minimal-pair words (see Cutler, 2012). Recognising words in continuous speech involves processes of competition between candidate words. For example, the English phrase “ship inquiry” (/ʃɪpɪŋkwaiəri/) contains the candidates *ship*, *shipping*, *ping*, *pink*, *ink*, *inquire*, *inquiry*, *choir*, *why*, and *wire*, among others, and quite a few fragments of words that are partially activated and excluded as the words unfold (e.g., *shipwreck*, *include*, *quiet*) (Norris, 1994). It is not difficult to imagine how the pool of candidates would increase if certain contrasts were not able to be discriminated. If a learner was unable to discriminate English /s/-/ʃ/ and /i/-/ɪ/, then the candidates *see*, *she*, *seep*, *sip*, *sheep*, *sipping*, *seeping*, and *pea* would be added to the list, along with many additional word fragments. A further consequence of this is that unresolved competition between candidate words lasts longer in L2 listening than in L1 listening (Weber & Cutler, 2004). This means L2 comprehension involves a higher cognitive load than L1 comprehension when the L2 user cannot discriminate certain L2 contrasts. Proficient L2 users may be able to use prior knowledge of communicative situations to reduce cognitive load (Tyler, 2001), but if L2 learners were able to acquire phonological contrasts early in L2 acquisition, then their L2 vocabulary should support a more efficient word recognition system.

To incorporate PAM-L2 principles in the classroom, perceptual assimilation to the L1 needs to be taken into account in FL curricula, and learners need to have opportunities for tuning in to the phonetic differences that signal phonological contrast in the L2 prior to the acquisition of a large L2 vocabulary. Suggestions already provided by Piske (2007) resonate with this idea – students should be exposed to high quality input and there should be opportunities for perceptual training. Below are some suggestions that elaborate on those ideas from a PAM-L2 perspective, and which also consider the issue of vocabulary acquisition.

### 5.1 Ensure students are exposed to L2 phonological contrast

Exposure to rich and varying speech from native speakers may be important for L2 acquisition (Piske, 2007), but this may not always be

possible to provide in an FL classroom. Students in class need to practise communicating with each other in the L2 to be able to use the language in real-life situations, so accented speech certainly cannot be avoided entirely. Nevertheless, teachers should ensure that students have as much exposure as possible to L2 speech that unambiguously preserves a phonological contrast between all phonemes in the target accent of the L2. They should also explain the importance of learning to perceive the differences between L2 phonemes for ease of L2 word recognition. If teachers reliably produce a phonetic difference between L2 phonemes (i.e., that native speakers of the L2 detect as a phonological contrast), then they can confidently model pronunciation for the class and supplement their exposure with audio(visual) materials of authentic native speaker productions. On the other hand, if they do not reliably produce a distinction, then audio(visual) materials should be used, and they should avoid modelling pronunciation of words with confusable L2 phonemes. To be clear, it is not L2-accented input that needs to be avoided – it is input that fails to provide clear phonetic differences between contrasting L2 phonemes. The beginner stages of learning are crucial for perceptual learning. The acquisition of L2 vocabulary that does not preserve L2 phonological contrast may set the learner on a trajectory that is difficult to remediate at a later stage of learning.

### **5.2 Provide opportunities for perceptual training**

Time in class should be devoted to perceptual training of single-category, category-goodness, uncategorised-categorised, and uncategorised-uncategorised assimilations (see also Piske, 2007). There is a long history of high variability training studies that have shown improvements in identification of minimal pair words when feedback is provided and the stimuli are spoken by multiple speakers (e.g., Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Iverson, Pinet, & Evans, 2012; Logan, Lively, & Pisoni, 1991). High variability perceptual training could be conducted in class, or as self-study using a computer-based approach. Additionally, activities designed to draw students' attention to phonological contrasts in the context of L2 pronunciation teaching may be adapted for this purpose (for reviews see Gurzynski-Weiss, Long, & Solon, 2017; Mora & Levkina, 2017). The crucial time for perceptual training is at the early stages of learning, prior to the establishment of a large L2 vocabulary.

For effective perceptual training, it is necessary to know how L2 consonants and vowels are assimilated by the students. In classrooms where all students have the same L1, existing cross-language speech perception

studies may provide sufficient information about how the L2 phonemes are likely to be assimilated. However, there are individual differences in perceptual assimilation, and many classrooms have students from diverse backgrounds. To obtain a clear picture of how students' L1s might influence L2 speech perception, teachers may consider including tests of perceptual assimilation as part of the initial student diagnostic tests that are often used to gauge a student's level. This would allow perceptual training to be tailored to the individual beginner student's needs and identify areas for remediation for more advanced learners.

### **5.3 Take perceptual assimilation into account when introducing L2 vocabulary**

To the extent that it is possible without becoming artificial, early vocabulary should preferentially include words that are easily discriminable using common L1-L2 categories, and words involving uncategorised phones. Words involving single-category assimilations, or the less-good phoneme of a category-goodness assimilation, should be introduced slowly and incorporated into perceptual training regimes. This should give students an opportunity to tune in to the phonetic differences that signal the phonological contrast before the vocabulary becomes too large. In addition to learning words for meaning and context, students should be given frequent opportunities to compare the pronunciations of groups of words containing one L2 phoneme with other groups of words containing a contrasting L2 phoneme. Obviously, there are many other factors that determine the order that vocabulary is introduced, but with an awareness of perceptual assimilation as a factor when designing curricula, it may be possible to delay the introduction of many words to allow more time for perceptual learning before the vocabulary becomes too large.

### **5.4 Delay introduction of orthography and/or teach the phonetic alphabet**

Students whose L1 is alphabetic are likely to apply L1 grapheme-phoneme correspondences when reading L2 words (e.g., Escudero, Simon, & Mulak, 2014; Hayes-Harb, Nicol, & Barker, 2010), which may inhibit optimal L2 phonological development. Delaying the introduction of orthography for as long as possible should increase the window of time available for tuning in to the phonetic differences that define L2 phonological contrasts. Also, delaying the introduction of orthography may be key to managing the rate of vocabulary growth. Many students may be frustrated if they

are not given the spelling for newly acquired vocabulary so it would be important to explain the importance of tuning in to phonological contrasts and how delaying spelling may support that. An additional solution may be to introduce an orthography that provides a one-to-one correspondence between phonemes (or allophones) and graphemes, such as the International Phonetic Alphabet (IPA). This would promote an awareness of phonological differences that are difficult to perceive, and it may provide a point of focus to help students to learn the phonetic differences between L2 phonemes. Vocabulary would still need to be introduced slowly, but once the student has learned a certain number of words using the phonetic script (e.g., 50 words), the L2 orthography could be introduced for words already learned. Any new words learned subsequently would be acquired with both the phonetic script and the L2 orthography to ensure that the learner is aware of the correct phonological form. Teaching IPA at the beginner stage would also open up possibilities for tracking perceptual assimilation over time, because IPA symbols could be used instead of regular orthography and keywords in L2 categorisation tasks. They could also be used in perceptual training tasks to focus attention at the phonemic level rather than using identification of minimal-pair words, which requires the acquisition of new vocabulary, and as diagnostic tests to track students' phonological development.

## **6. Summary and conclusions**

PAM-L2 bases its predictions about L2 category acquisition on the pattern of perceptual assimilations of L2 phonemes to the L1 phonological system at first contact with the L2. For optimal L2 perception, the learner needs to detect a phonological contrast between each L2 phoneme and all other L2 phonemes. This can be achieved using existing L1 phonological categories, which become common L1-L2 categories, or by establishing new L2-only phonological categories. A new L2 phonological category is most likely to be acquired for the less-good version of a category-goodness assimilation, or for an uncategorised L2 phoneme. The likelihood of acquiring a new L2 phonological category is crucially dependent on the learner having opportunities for perceptual learning at an early stage of language acquisition. The perceptual learning should occur prior to the establishment of a large L2 vocabulary, especially for contrasts where learners have poor discrimination accuracy (i.e., single category assimilations and contrasts involving overlapping uncategorised L2 phonemes). If learners can already detect a clear phonetic difference between contrasting L2 phonemes



(i.e., contrasts involving non-overlapping or dispersed uncategorised L2 phonemes) then perceptual learning should be rapid. An increasing L2 vocabulary and unambiguous grapheme-phoneme correspondences may support further attunement, provided that the spoken input preserves the phonological properties of the contrast.

While PAM-L2's predictions were formulated with an immersion context in mind, there are no qualitative differences in the predictions when they are applied to the FL classroom. However, the likelihood of new category acquisition would be generally lower in the classroom than in an idealised immersion environment because of fewer opportunities for perceptual learning of L2 phonological contrast prior to the acquisition of a large L2 vocabulary. All of the suggestions made by Piske (2007) would certainly improve the likelihood of category acquisition in the classroom. Applying the principles of PAM-L2, the likelihood may be further improved by: 1) ensuring that learners are exposed to clear phonological differences for all L2 contrasts; 2) providing perceptual training at the beginner level for single-category, category-goodness, uncategorised-categorised, and uncategorised-uncategorised assimilations; 3) taking perceptual assimilation into account when introducing new vocabulary, and; 4) managing the introduction of written forms of words.

Optimal L2 phonological acquisition is a desirable outcome for the L2 learner, but the theoretically inspired suggestions made here clearly do not take into account the practicalities of classroom-based FLA. It is acknowledged that some of the suggestions may be impossible to achieve in certain FL contexts. For example, some language schools may enrol students for only short periods of time, they may not offer classes for absolute beginners, or they may have limited hours of face-to-face teaching, making it impractical to introduce time-consuming activities such as perceptual training and teaching IPA script. Given the widespread teaching of English throughout the world, it may be that these suggestions are more feasible for other L2 target languages where students are more likely to be naïve at the onset of learning. Before these theoretical suggestions can be put into practice, more research is required to test whether perceptual training prior to large vocabulary acquisition results in improvements over the longer term. Researchers at universities with a mandatory foreign language requirement would seem to be well placed to conduct such a study.

Even without direct empirical evidence for the specific suggestions made here, it is clear from previous research that perceptual assimilation to the L1 has a strong influence on L2 speech perception. Some L2 contrasts

are easy to discriminate at the onset of FL learning, while others are very difficult indeed. The input that students receive in the classroom (and outside of class) is crucial for setting their perceptual learning trajectories. Even if it is not possible to implement curriculum-based strategies to support phonological acquisition, students should be made aware of those contrasts that are likely to cause difficulty in the L2. Motivated students who seek opportunities to “train their ears” outside of the classroom may benefit from less effortful L2 comprehension when they progress to an advanced level of L2 acquisition.

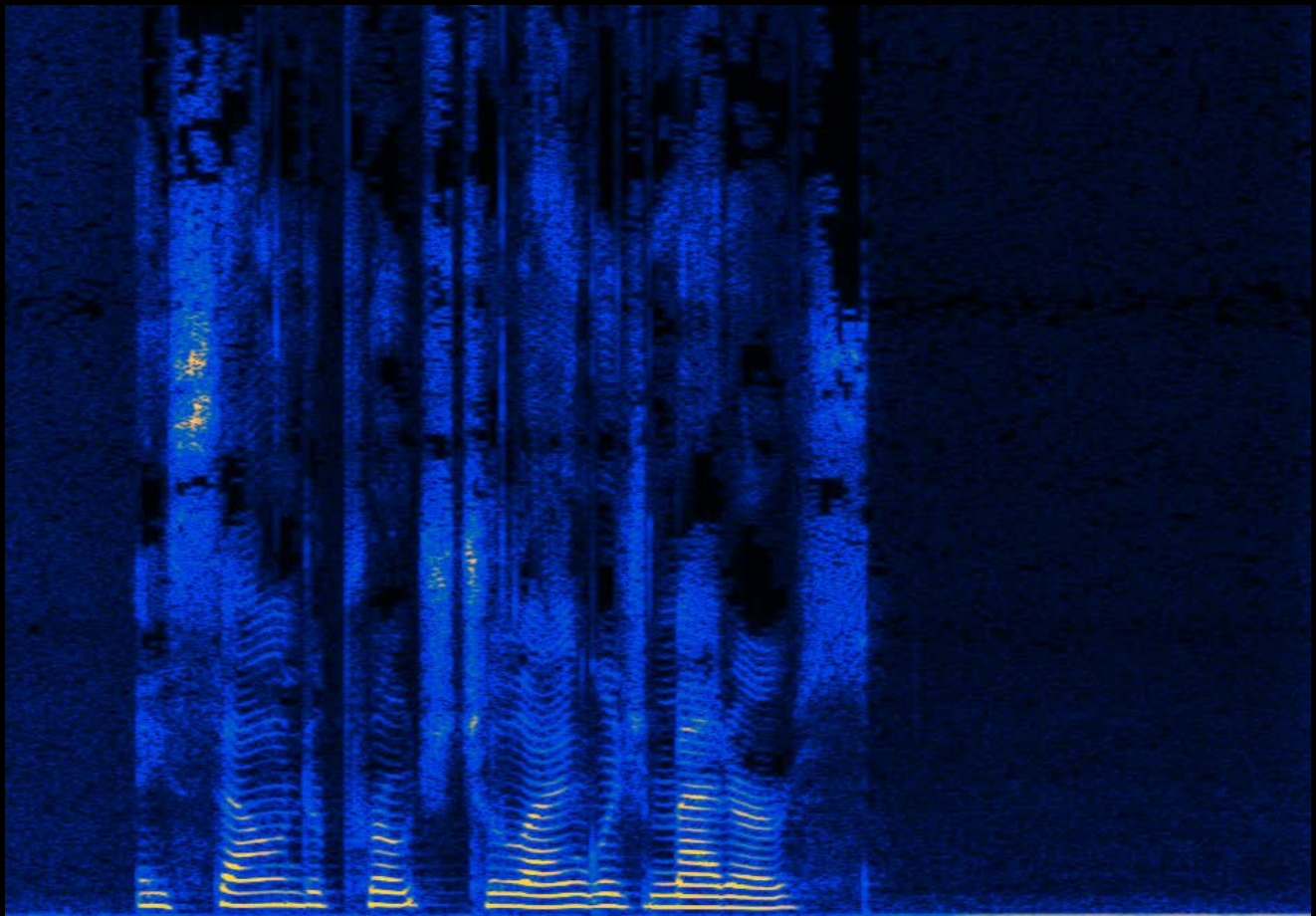
### References

- Antoniou, M., Best, C. T., & Tyler, M. D. (2013). Focusing the lens of language experience: Perception of Ma'di stops by Greek and English bilinguals and monolinguals. *Journal of the Acoustical Society of America*, 133(4), 2397-2411.
- Best, C. T. (1993). Emergence of Language-Specific Constraints in Perception of Non-Native Speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 289-304). Dordrecht: Springer Netherlands.
- Best, C. T. (1994). Learning to perceive the sound pattern of English. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 9, pp. 217-304). Norwood, NJ: Ablex.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Baltimore: York Press.
- Best, C. T. (2015). Devil or angel in the details? Perceiving phonetic variation as information about phonological structure. In J. Romero & M. Riera (Eds.), *Phonetics-phonology interface: Representations and methodologies* (pp. 3-31).
- Best, C. T., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, 46(2-3), 183-216.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, 109(2), 775-794.
- Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior & Development*, 18(3), 339-350.

- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13-34). Amsterdam: John Benjamins.
- Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native-and Jamaican-accented words. *Psychological science*, *20*(5), 539-542.
- Bohn, O.-S., & Bundgaard-Nielsen, R. L. (2009). Second language speech learning with diverse inputs. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 207-218). Clevedon, UK: Multilingual matters.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*, 977-985.
- Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C., & Tyler, M. D. (2012). Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics*, *33*, 643-664.
- Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011a). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, *33*, 433-461.
- Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011b). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, *32*, 51-67.
- Cutler, A. (2012). *Native listening*. Cambridge, MA: MIT Press.
- Ehri, L. C. (1984). How orthography alters spoken language competencies in children learning to read and spell. In J. Downing & R. Valtin (Eds.), *Language awareness and learning to read* (pp. 119-147). New York: Springer-Verlag.
- Ehri, L. C. (1985). Effects of printed language acquisition on speech. In D. R. Olson, N. Torrance, & A. Hildyard (Eds.), *Literacy, language, and learning: The nature and consequences of reading and writing* (pp. 333-367). Cambridge, UK: Cambridge University Press.
- Escudero, P., Simon, E., & Mulak, K. E. (2014). Learning words in a new language: Orthography doesn't always help. *Bilingualism: Language and Cognition*, *17*, 384-395.
- Faris, M. M., Best, C. T., & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America*, *139*(1), EL1-EL5.
- Faris, M. M., Best, C. T., & Tyler, M. D. (2018). Discrimination of uncategorized non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics*, *70*, 1-19.
- Faris, M. M., Best, C. T., & Tyler, M. D. (in preparation). Phonological category acquisition and discrimination of L2 vowels by learning in an immersion setting: A longitudinal study.

- Fenwick, S. E., Best, C. T., Davis, C., & Tyler, M. D. (2017). The influence of auditory-visual speech and clear speech on cross-language perceptual assimilation. *Speech Communication, 92*, 114-124.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-276). Baltimore: York Press.
- Flege, J. E. (2003). Methods for assessing the perception of vowels in a second language. In E. Fava & A. Mioni (Eds.), *Issues in clinical linguistics* (pp. 19-44). Padova, Italy: University Press.
- Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26*(1), 1-34.
- Gurzynski-Weiss, L., Long, A. Y., & Solon, M. (2017). TBLT and L2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis? *Studies in Second Language Acquisition, 39*(2), 213-224.
- Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech, 53*, 367-381.
- Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics, 33*(1), 145-160.
- Jenkins, J. J., Strange, W., & Polka, L. (1995). Not everyone can tell a “rock” from a “lock”: Assessing individual differences in speech perception. In D. J. Lubinski & R. V. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 297-325). Palo Alto, CA, USA: Davies-Black Publishing.
- Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*(2), 262-268.
- Ladefoged, P., & Johnson, K. (2014). *A course in phonetics* (7th ed.). Stamford, CT: Cengage Learning.
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874-886.
- McQueen, J. M., Tyler, M. D., & Cutler, A. (2012). Lexical retuning of children’s speech perception: Evidence for knowledge about words’ component sounds. *Language Learning and Development, 8*(4), 317-339.
- Mora, J. C., & Levkina, M. (2017). Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition, 39*(2), 381-399.
- Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 193-218). Amsterdam: John Benjamins.
- Nazzi, T., & Bertoni, J. (2003). Before and after the vocabulary spurt: Two modes of word acquisition? *Developmental Science, 6*(2), 136-142.

- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189-234.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204-238.
- Piske, T. (2007). Implications of James E. Flege's research for the foreign language classroom. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning. In honor of James Emil Flege* (pp. 301-314). Amsterdam: John Benjamins.
- Read, C., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition*, 24, 31-44.
- San, N., & Tyler, M. D. (in preparation). Acquisition and discrimination of vowels by learners varying in foreign-language experience.
- So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, 36(2), 195-221.
- Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39, 456-466.
- Tyler, M. D. (2001). Resource consumption as a function of topic knowledge in nonnative and native comprehension. *Language Learning*, 51, 257-280.
- Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, 71(1), 4-21.
- Tyler, M. D., Best, C. T., Goldstein, L. M., & Antoniou, M. (2014). Investigating the role of articulatory organs and perceptual assimilation in infants' discrimination of native and non-native fricative place contrasts. *Developmental Psychobiology*, 56, 210-227.
- Tyler, M. D., & Burnham, D. K. (2006). Orthographic influences on phoneme deletion response times. *Quarterly Journal of Experimental Psychology*, 59(11), 2010-2031.
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1), 1-25.



*School of Communication & Culture  
Aarhus University  
Denmark*

*E-ISBN: 978-87-7507-440-2*