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.

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^h 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.

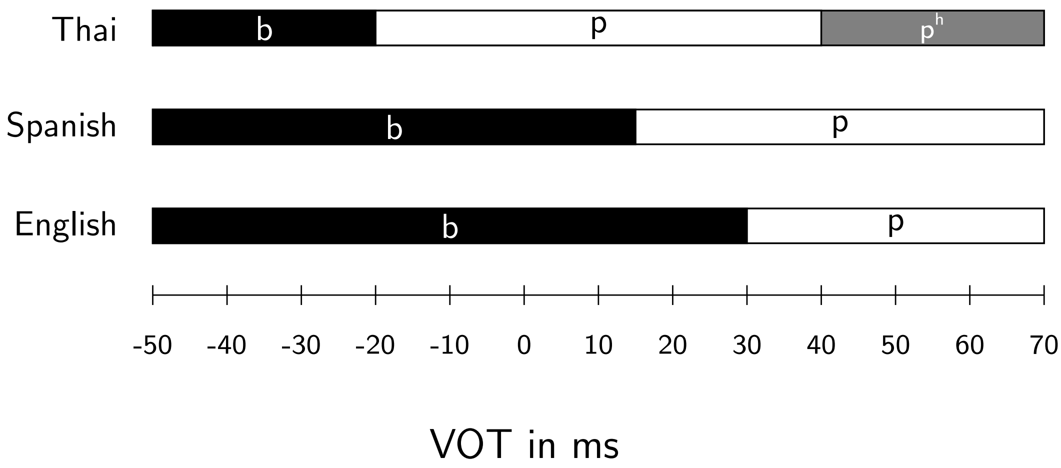

Figure 1. VOT boundaries across three languages. Adapted from Abramson & Lisker, 1970.

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. /pʰ/, 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, & Hennessy, 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, ʈ, ȶ/, 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.

Lab.	Lam.- dent.	Apic.- Alv.	Apic- postalv.	Lam.- postalv.	Vel.
p	t̪	t	ʈ	ȶ	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.	Glot.
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.

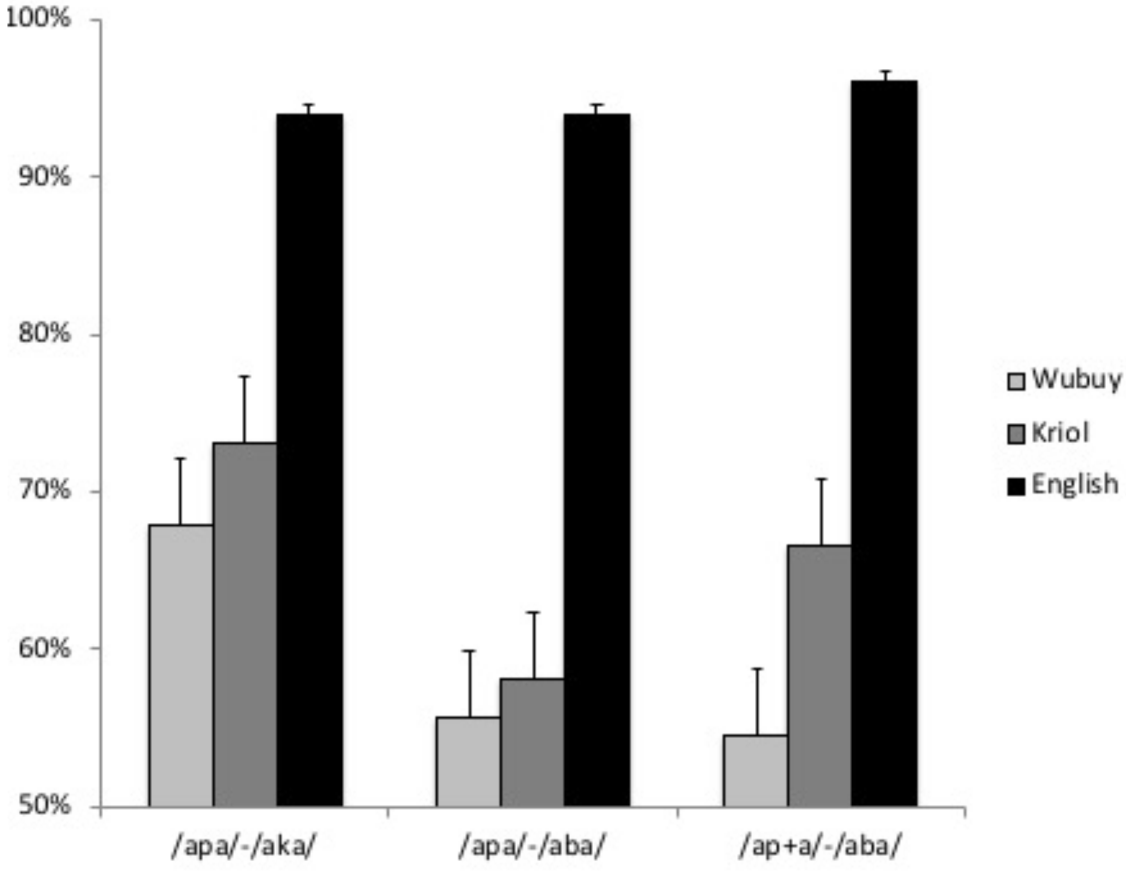

Figure 2. Mean discrimination accuracy for Wubuy, Kriol and English speakers in Study 1. 50% indicates chance performance. Error bars indicate S.E.

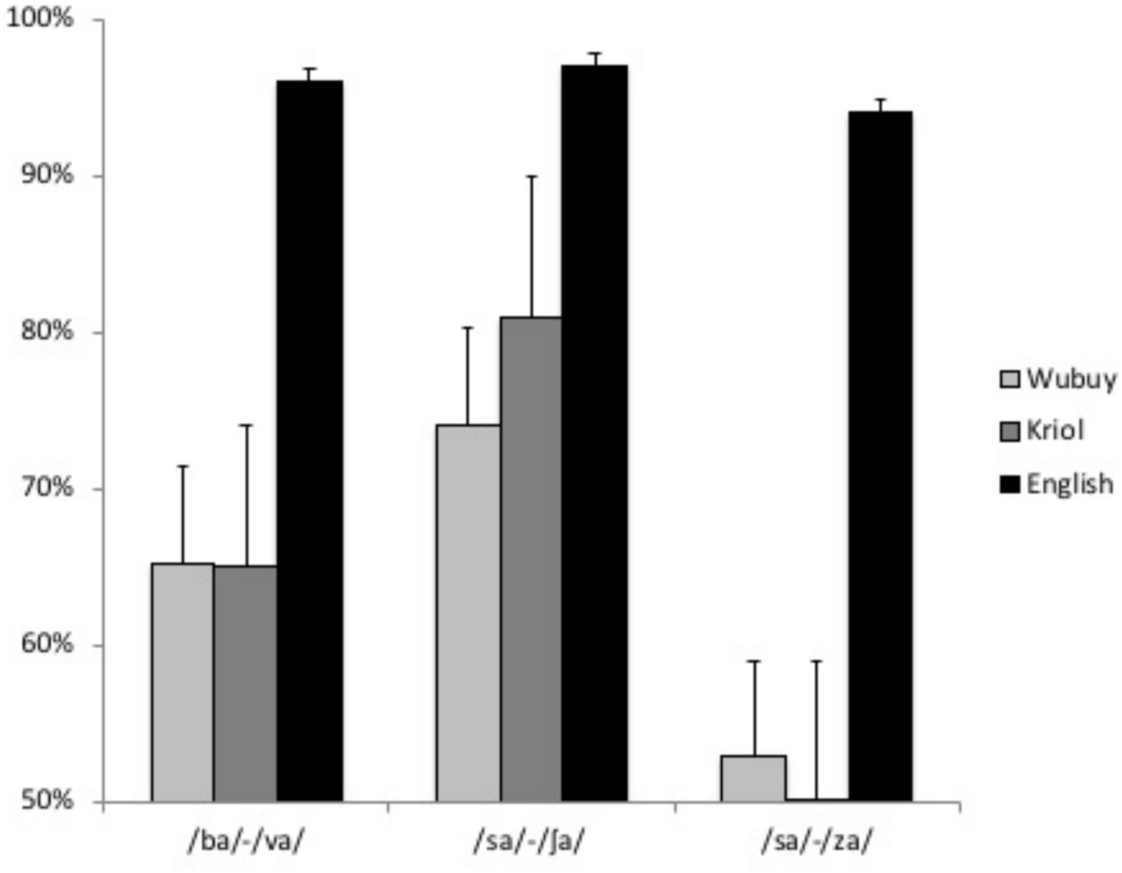

Figure 3. Mean discrimination accuracy for Wubuy, Kriol and English speakers in Study 2. 50% indicates chance performance. Error bars indicate S.E.

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 [±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: Theoretical 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*.