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.

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. 211-236). Dept. of English, School of Communication & Culture, Aarhus University.

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-attuning 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" [ɹɑk] and "lock" [lɑk]. 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
ka [ka]	*zhuo* [t͡ʂo]
pou [poʊ]	*xiong* [ɕiʊŋ]
fu [fu]	*run* [ɹun]
lan [lan]	*zi* [tsi]
nin [nin]	*que* [t͡ɕʰue]
ting [tiŋ]	*chi* [t͡ʂʰi]
TRAINING SYLLABLES	
Familiar segment	Unfamiliar segment
ming [miŋ]	*ri* [ɹ͡i]
yao [jao]	*chun* [tɕun]
te [te]	*qiong* [tɕʰioŋ]
wa [wa]	*xue* [ɕue]
kai [kʰai]	*cuo* [tsʰuo]
lao [lao]	*zhi* [t͡ʂi]

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.

<table>
<tr><th>Pre-test</th><th colspan="3">Training</th><th>Post-test</th></tr>
<tr><td rowspan="2">Tone and Task Familiarization</td><td>Session 1</td><td>Session 2</td><td>Session 3</td><td rowspan="4">Same as Pre-test</td></tr>
<tr><td>Training
(2 blocks)</td><td>Training
(2 blocks)</td><td>Training
(2 blocks)</td></tr>
<tr><td rowspan="2">Tone identification of 96 stimuli not used in training</td><td>Training test</td><td>Training test</td><td>Training test</td></tr>
<tr><td colspan="3">• The Meaning group viewed pictures associated with each tone word. The Tone Only group viewed a fixation cross.
• Each training session contained 192 items (6 syllables x 4 tones x 2 speakers x 4 repetitions).
• The training test contained stimuli received during training.</td></tr>
</table>

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 (β=1.49, *SE* β=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 (β=-0.76, *SE* β=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 (β=-0.55, *SE* β=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.

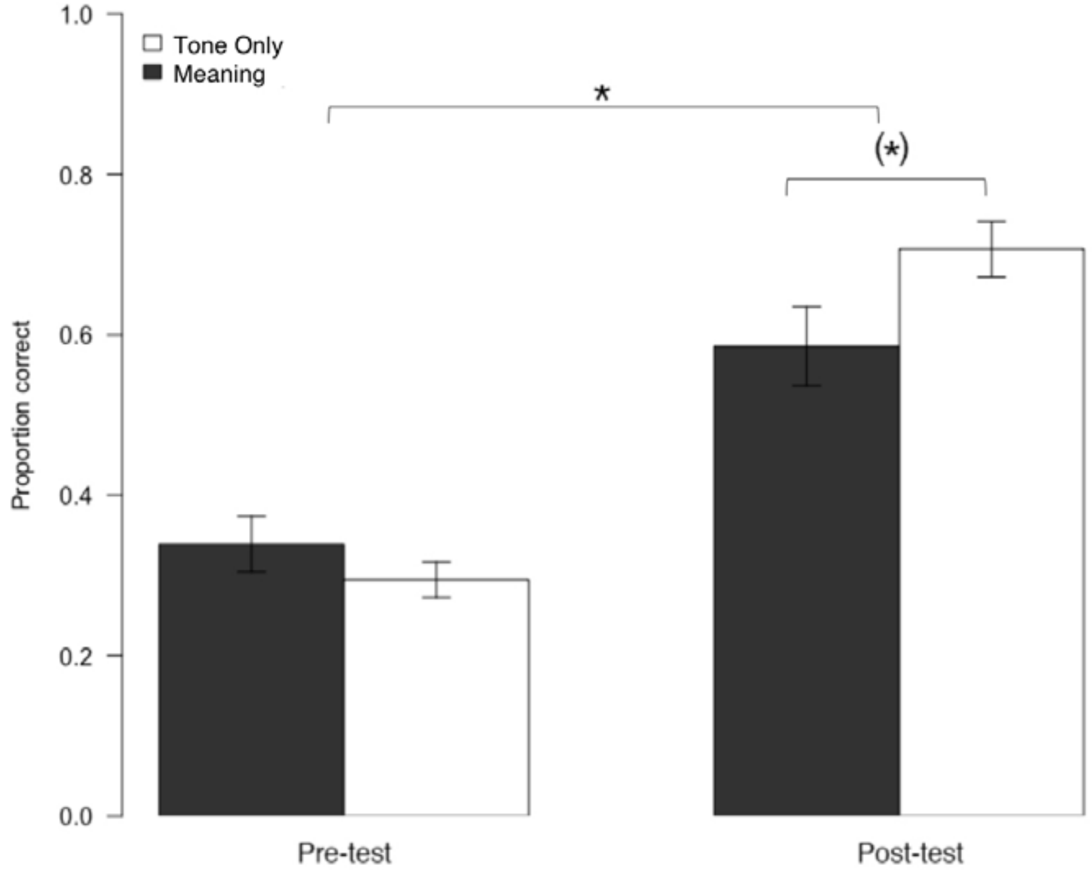

Figure 2. Mean proportion correct tone identification for Pre-test and Post-test by Training Group (Tone Only, Meaning). Asterisks denote significance (p<0.05), and asterisks in parentheses indicate marginal significance. Error bars indicate +/- 1 standard error.

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 (β=1.5, *SE* β=0.14, $\chi^2(1)$=42.09, p<0.001) and Session B (β=0.41, *SE* β=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 (β=2.02, *SE* β=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 (β=1.2, *SE* β=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 (χ^2=2.22, p=0.14).

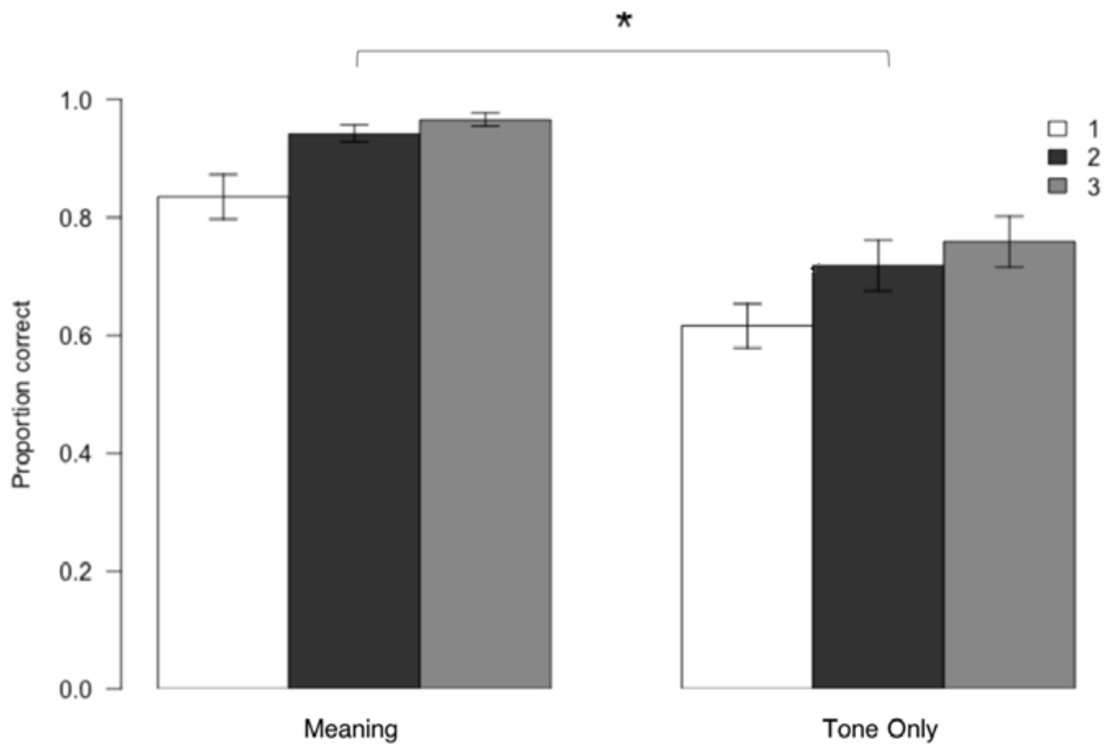

Figure 3. Mean proportion correct tone identification by training session (1-3) and training type (Meaning vs. Tone Only). The asterisk denotes significance (p<0.05). Error bars indicate +/- 1 standard error.

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 (Sereno & 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., [t͡ʂo], [ɻun], [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.

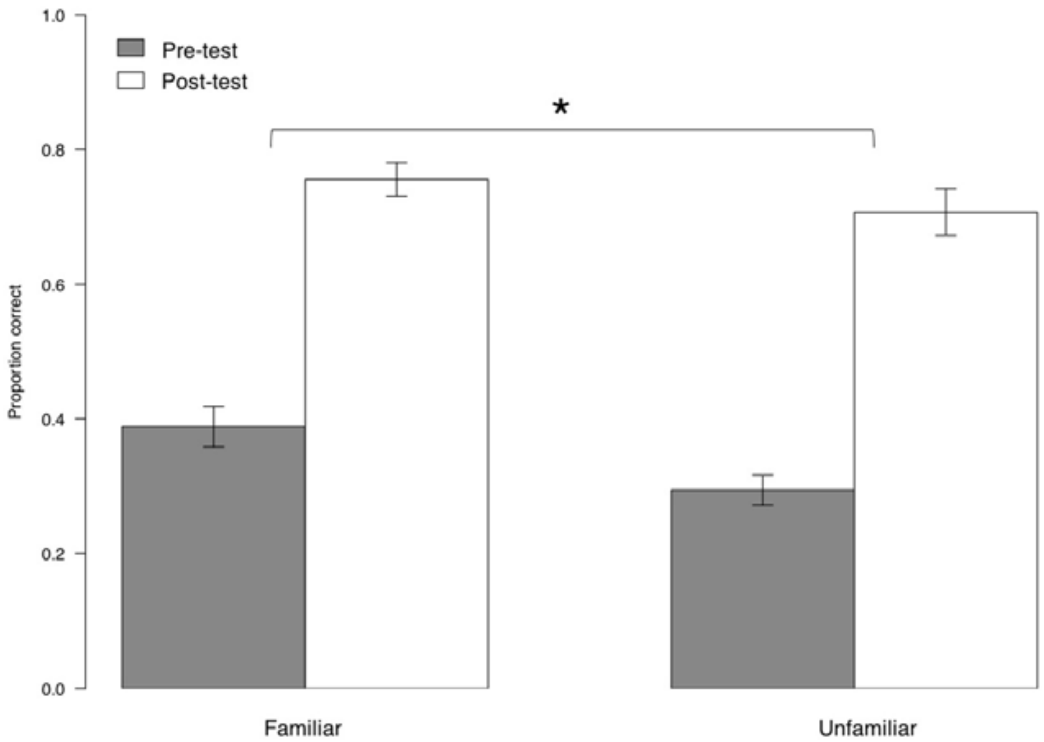

Figure 4. Mean proportion correct tone identification for Pre-test and Post-test by Segment Type (Familiar vs. Unfamiliar). The asterisk denotes significance ($p<0.05$) and error bars indicate +/- 1 standard error.

A significant main effect of Test was found (β=1.76, *SE* β=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 (β=0.34, *SE* β=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.

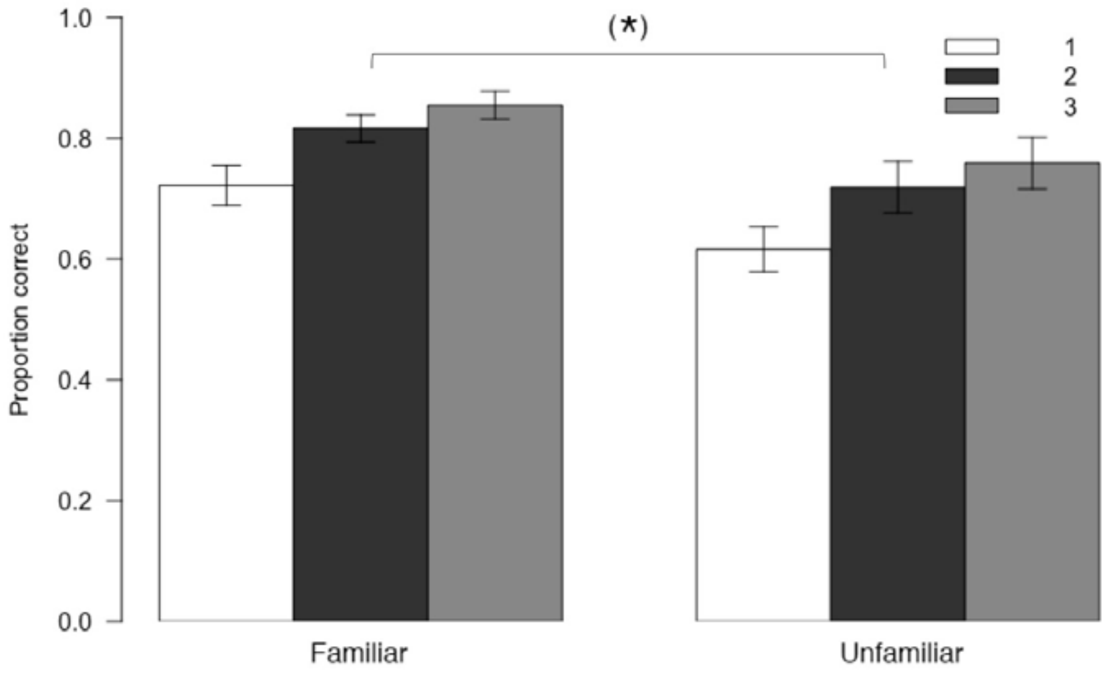

Figure 5. Mean proportion correct tone identification by training session (1-3) and training segment type (Familiar vs. Unfamiliar). The asterisk in parentheses denotes marginal significance ($p=0.06$) and error bars indicate +/- 1 standard error.

Significant effects of Session A (β=0.91, *SE* β=0.11, $\chi^2(1)$=34.26, p<0.001) and Session B (β=0.27, *SE* β=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 (β=0.53, *SE* β=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 (χ^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.

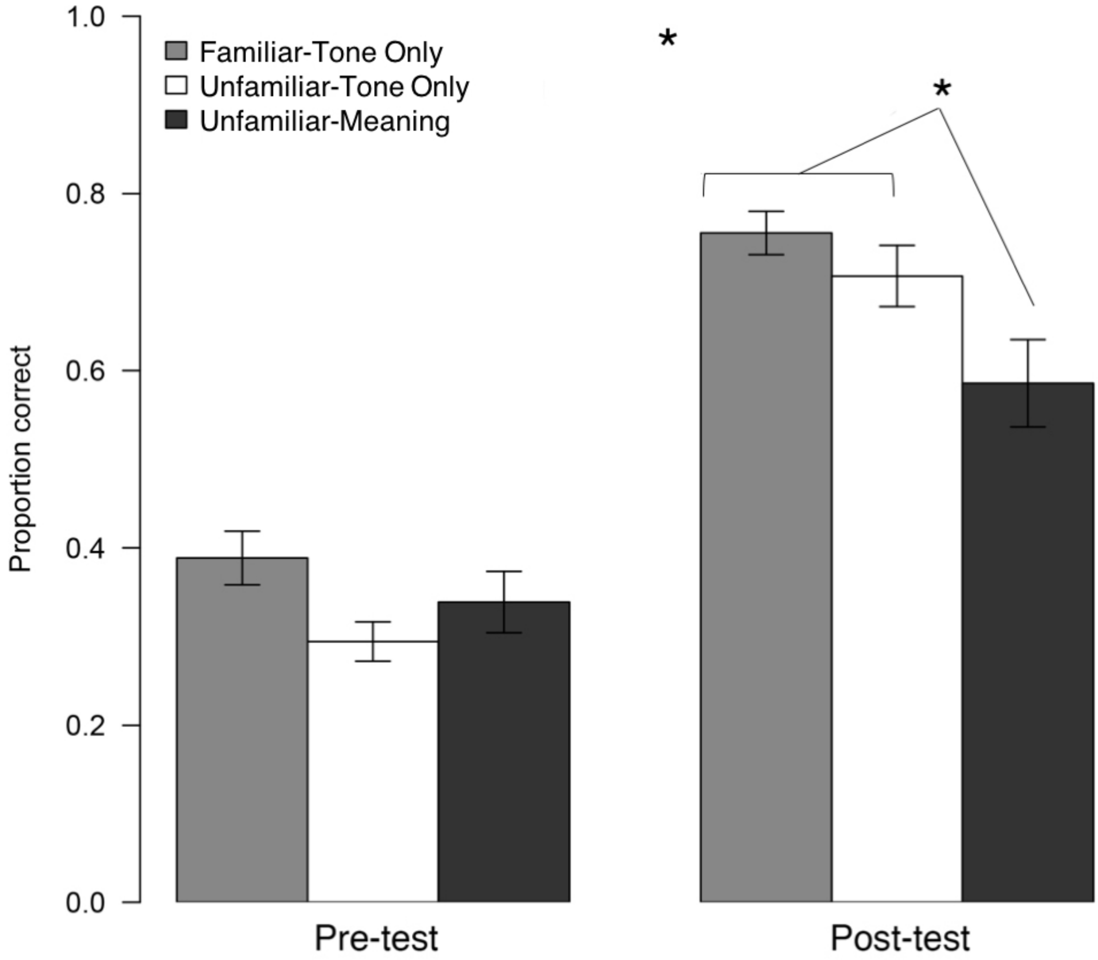

Figure 6. Mean proportion correct tone identification for Pre-test and Post-test by Group (Familiar-Tone Only, FTO; Unfamiliar-Tone Only, UTO; Unfamiliar-Meaning, UM). The top, centre asterisk denotes a significant pre- to post-test difference. Error bars indicate +/- 1 standard error.

A significant main effect of Test (β=1.52, *SE* β=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 (β=0.46, *SE* β=0.21, $\chi^2(1)$=4.7, p=0.03), along with a significant Test x Group A interaction (β=0.87, *SE* β=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 (β=0.89, *SE* β=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 (β=0.35, *SE* β=0.18, $\chi^2(1)=3.806$, $p=0.051$)[1]; 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 [poʊ] vs. [ɕiʊŋ]).

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

---

1 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, Ettlinger, Vytlacil, & 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 *17th 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. Nusbaim (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., Vytlacil, 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

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

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.