

**THE PERCEPTION, PROCESSING AND LEARNING  
OF MANDARIN LEXICAL TONE  
BY SECOND LANGUAGE SPEAKERS**

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY  
OF HAWAI'I AT MĀNOA IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

SECOND LANGUAGE STUDIES

September 2021

By

Wenyi Ling

Dissertation committee:

Theres Grüter, Chairperson

Richard R. Day

Rory Turnbull

Dustin Crowther

Haidan Wang

Keywords: lexical tone, second language processing, Mandarin, learning

## ACKNOWLEDGMENTS

This dissertation is by no means a product of just one person but was made possible by generous funding agencies, professional mentors and professors, and encouraging friends and family.

I would never have been able to complete this dissertation without the financial support of NSF Doctoral Dissertation Improvement Grant (# 1824082), Research Corporation of the University of Hawaii Graduate Fellowship Fund, Elizabeth Carr-Holmes Scholarship, Ruth Crymes Memorial Grant, Grace Ning Chinese Studies Fund and PKU-UHM Scholar Exchange Program. My sincere thanks also go to Paula Menyuk Travel Award, East-West Center Conference Scholarship and Tone Aspect Language Conference student funding for conference traveling.

I would like to express my deepest gratitude to my great advisor and mentor, Dr. Theres Grüter, for her patient guidance on research development, and warm encouragement on life in general. Her commentary, critiques and genuine enthusiasm for my project kept me engaged and carried me forward in a lot of difficult times. I am tremendously grateful to my other committee members: Dr. Richard R. Day, thank you for being a steady support for me with your wise knowledge and advise throughout my time in SLS; Dr. Rory Turnbull, thank you for providing professional comments on the part of phonetic and phonology and being such a considerate professor, Dr. Dustin Crowther, thank you for your detailed comments and great suggestion on vocabulary acquisition, and Dr. Haidan Wang, thank you for serving as the language expert and inspiring mentor in Mandarin teaching and research.

In addition to the members of my committee, I would also like to thank some great professors, who helped the development of this project at different stages. I want to send my sincere thanks to Dr. Amy Schafer, Dr. Victoria Anderson, Dr. Katie Drager, Dr. Geoffrey LaFlair, Dr. Bonnie Schwartz and Dr. Dan Isbell, who inspired me to design the experiments and taught me how to improve them. I also owe my thanks to Dr. Shiyi Lu, Dr. Yongyi Wu, Dr. Elaine Lau and Dr. Stephen Matthews, who generously offered me their lab space and helped me to recruit participants in Beijing, Shanghai and Hong Kong.

My research and life in UHM would have not been so productive and sustainable without support and encouragement from friends and local community. I thank all my friends at UHM:

Peter, Nozomi, Kirsten, Alice, Akari, Mayuko, Hitoshi, Hye Young, Crystal, Mitsuko, Parvaneh, Yu-Han, George, Haerim, Hyunwoo, Aya, Orn, Jing, Hyunah, Mari, Ivan, Lilian, Yangxi, Jonny, Katie, Chris, Elena, Kelly, Jiamin, Jiaxin, Yanxia, Xiaoxing, Rum, Vivian, and Huizhong. I especially want to send my Mahalo to Amber, whom I learnt so much from and shared a lot of happiness and sadness with, and Fred, who carefully proofread this dissertation and encouraged me to carry on. I also want to say thank you to people in the community of LAE lab, Language Acquisition Reading Group, Linguistics Beyond the Classroom and EWC student affiliate program.

Lastly, I would like to thank my family for their continued support, love, care and understanding in my long academic journey. I really appreciate the support and understanding from my parents, Guizhu Sun and Guohua Ling, who unconditionally trust my choice in life and work. I want to thank my dear husband, Prasad, and baby son, Shuka, who encouraged me through the good and bad times, enlightened me to think deeper and wider, reminded me what is the most important in life and how to love myself in their own unique ways.

## ABSTRACT

This dissertation investigates how English-speaking second language (L2) learners of Mandarin perceive, process and learn Mandarin lexical tones. While most languages use modulations in pitch (intonation) to convey meanings at the phrasal and sentential levels, a number of languages, including Mandarin, also use suprasegmental features such as pitch to encode meaning at a lexical level (lexical tone). A key challenge for speakers of a non-tonal language learning Mandarin is to acquire this new function of pitch. This includes learning to perceive tonal categories in continuous and variable acoustic input, building this information into their lexical representations, and accessing the information during real-time processing. Given the complexity of this task, it is not surprising that lexical tone has been identified as one of the most difficult domains for L2 learners to master. The precise source of this difficulty, however, is still not well understood. The goal of this dissertation is to investigate L2 learners' processing and use of tone at multiple levels, including speech perception, lexical processing, and word learning, in order to gain a broader understanding of the challenges in the L2 acquisition of Mandarin tone by native speakers of English. To this end, this dissertation draws on theoretical models and methodological paradigms from research in speech perception and psycholinguistics that has investigated the perception and processing of tone by native speakers, and extends these paradigms to investigate the acquisition and use of tone by L2 learners. This dissertation consists of three experiments. Experiment 1 employs identification and discrimination tasks well-established in categorical perception research to explore the extent to which L2 learners perceive tone categorically and the role of L2 proficiency in this process. Experiment 2 investigates how L2 learners process tone in the real-time comprehension of spoken Mandarin in a visual-world eye-tracking study and the relation between L2 learners' ability in tone perception at the

phonological level and their processing of tone at the lexical level. Finally, Experiment 3 explores the effects of manipulating the contrastive availability of tone during (novel) word learning on the use of tone in subsequent lexical processing. Results from Experiment 1 show that L2 learners tended to perceive tone less categorically than native speakers, and L2 proficiency was correlated with learners' degree of categorical perception of tone. Experiment 2 shows that L2 learners weighed tonal (vs. segmental) cues less than native speakers in lexical processing and there was a correlation between L2 learners' categorical perception of tone and their ability to use tonal (vs. segmental) cues in spoken word recognition. Results from Experiment 3 indicate that manipulating attentional focus towards tonal cues might not always be beneficial and suggests that selective focus on an individual cue may in fact be detrimental to the learning of other cues.

This dissertation provides the first direct evidence showing that L2 learners' ability to perceive tone categorically is related to their weighting of tonal cues during lexical processing, thus contributing to a better understanding of the link between processing at the phonological and lexical levels, which has been argued to be a key component in the L2 acquisition of tone (Wong & Perrachione, 2007; Cooper & Wang, 2013). Furthermore, this project tests the effectiveness of the commonly used cue-focus method for teaching words with tone in a controlled laboratory setting and raises awareness about teaching words as single entities. Overall, this dissertation contributes to research in SLA by bring conceptual insights gained from research on tone in native language processing to the study of tone in a non-native language. It also contributes to research in the fields of speech perception and psycholinguistics by probing the generalizability of findings about human language processing to L2 learners. Finally, it

contributes to evidence-based L2 instruction and curricular materials by testing the effectiveness of cue-focus training.

# CONTENTS

<b>ACKNOWLEDGMENTS</b> .....	<b>ii</b>
<b>ABSTRACT</b> .....	<b>iv</b>
<b>CONTENTS</b> .....	<b>vii</b>
<b>List of Tables</b> .....	<b>x</b>
<b>List of Figures</b> .....	<b>xii</b>
<b>Chapter 1. Introduction</b> .....	<b>1</b>
<i>The organization of this dissertation</i> .....	2
<b>Chapter 2. Background and literature review</b> .....	<b>5</b>
2.1 <i>Mandarin Chinese</i> .....	5
2.1.1 Segmental properties of Mandarin word.....	6
2.1.2 Lexical tone in Mandarin .....	7
2.1.2.1 Lexical tone notation.....	7
2.1.2.2 Tone sandhi.....	8
2.1.2.3 The mental representation of lexical tone.....	9
2.2 <i>Tone in L1 and L2 processing</i> .....	10
2.2.1 Processing tone at the phonological level.....	10
2.2.1.1 Methods for measuring categorical perception (CP) .....	12
2.2.1.2 The Automatic Selective Perception (ASP) model.....	14
2.2.1.3 Tone perception by native and naïve listeners .....	16
2.2.1.4 Tone perception by L2 learners .....	19
2.2.2 Processing tone at the lexical level .....	21
2.2.2.1 Processing tone at the lexical level by native speakers.....	22
2.2.2.2 Processing tone at the lexical level by L2 learners.....	26
2.2.2.3 The relation between tone processing at phonological and lexical levels by L2 learners.....	28
2.2.3 Tone in word learning.....	30
2.2.3.1 Factors influencing L2 tone learning .....	30
2.2.3.2 Teaching tone to L2 learners.....	33

2.2.3.3 Artificial word learning studies .....	35
2.3 Summary .....	38
<b>Chapter 3. Experiment 1: Categorical perception of lexical tone by speakers with different language backgrounds .....</b>	<b>40</b>
3.1 Research questions.....	40
3.2 Methods.....	41
3.2.1 Participants.....	41
3.2.2 Materials .....	42
3.2.3 Experimental tasks .....	45
3.2.3.1 Identification task.....	45
3.2.3.2 Discrimination task .....	45
3.2.3.3 Listening proficiency test.....	46
3.2.3.4 General procedure .....	47
3.2.4 Data analysis .....	47
3.2.4.1 Identification measurement.....	47
3.2.4.2 Discrimination measurement .....	48
3.3 Results .....	48
3.3.1 Identification task.....	48
3.3.2 Discrimination task .....	55
3.3.3 Proficiency .....	59
3.4 Discussion.....	60
<b>Chapter 4. Experiment 2: The relation between perception of tone and real-time spoken word recognition by L2 learners .....</b>	<b>65</b>
4.1 Research questions.....	65
4.2 Methods.....	66
4.2.1 Participants.....	66
4.2.2 Spoken word recognition task.....	67
4.2.2.1 Materials .....	67
4.2.2.2 Procedure .....	69
4.2.3 Identification task.....	70
4.2.4 General procedure .....	70



4.3 Results .....	70
4.3.1 Spoken word recognition task.....	71
4.3.1.1 Mouse-click data.....	71
4.3.1.2 Eye-movement data .....	75
4.3.2 Identification task.....	79
4.4 Discussion.....	81
<b>Chapter 5. Experiment 3: Learning words with tone in different cue-focus training</b>	
<b>conditions .....</b>	<b>87</b>
5.1 Research questions.....	87
5.2 Methods.....	88
5.2.1 Participants.....	88
5.2.2 Materials .....	89
5.2.3 Procedure .....	91
5.2.3.1 Pitch contour perception test (PCPT) .....	91
5.2.3.2 Training session .....	92
5.2.3.3 Test session.....	93
5.3 Results.....	94
5.3.1 Pitch contour perception test (PCPT) .....	94
5.3.2 Forced choice task.....	95
5.3.2.1 Mouse-click accuracy .....	96
5.3.2.2 Mouse-click reaction time.....	101
5.3.2.3 Eye-movement data .....	107
5.4 Discussion.....	113
<b>Chapter 6. General discussion, implications and limitations.....</b>	<b>116</b>
6.1 Summary of findings .....	116
6.2 Implications for teaching.....	122
6.3 Limitations and directions for future works .....	124
6.4 Concluding remarks.....	126
<b>APPENDICES.....</b>	<b>128</b>
<b>REFERENCES.....</b>	<b>143</b>

## List of Tables

<b>Table 1.</b> Segmental Structures of Mandarin Syllables with Examples .....	6
<b>Table 2.</b> Participants' Language Background Information .....	42
<b>Table 3.</b> Frequencies of the Morphemes that are Homophonic to the Endpoint Stimuli .....	43
<b>Table 4.</b> Mean Values of Identification Slopes for Mandarin Native Listeners, Naïve Listeners and L2 Learners for Each Tone Pair .....	50
<b>Table 5.</b> Results of Linear Mixed-effects Model for Identification Slopes .....	52
<b>Table 6.</b> Results of Comparison of Identification Slopes by Tone Pair within Each Group .....	54
<b>Table 7.</b> Results of Linear Mixed-effects Model for Discrimination.....	57
<b>Table 8.</b> Results of Comparison of Discrimination Accuracy by Tone Pair within Each Group	58
<b>Table 9.</b> Results of Generalized Linear Mixed-effects Model for Accuracy .....	73
<b>Table 10.</b> Results of Linear Mixed-effect Model for Proportion of Looks to Competitor, Aggregated over Participants (top) and Items (bottom) .....	79
<b>Table 11.</b> The Artificial Vocabulary .....	89
<b>Table 12.</b> Scheme for Generating the Pitch Contour .....	91
<b>Table 13.</b> Results of Generalized Linear Mixed-effects Model for PCPT Accuracy .....	95
<b>Table 14.</b> Results of Generalized Linear Mixed-effects Model for Accuracy .....	98
<b>Table 15.</b> Results of Generalized Linear Mixed-effects Model for Accuracy of Tone-pair Trials .....	99
<b>Table 16.</b> Results of Generalized Linear Mixed-effects Model for Accuracy of Vowel-pair Trials .....	100
<b>Table 17.</b> Results of Generalized Linear Mixed-effects Model for Accuracy of Consonant-pair Trials .....	100
<b>Table 18.</b> Results of Generalized Linear Mixed-effects Model for Accuracy of Baseline Trials .....	101
<b>Table 19.</b> Results of Linear Mixed-effects Model for Reaction Times on All Trials .....	105
<b>Table 20.</b> Results of Linear Mixed-effects Model for Reaction Times on Tone-pair Trials .....	106
<b>Table 21.</b> Results of Linear Mixed-effects Model for Reaction Times on Vowel-pair Trials...	106
<b>Table 22.</b> Results of Linear Mixed-effects Model for Reaction Times on Consonant-pair Trials .....	107
<b>Table 23.</b> Results of Linear Mixed-effects Model for Reaction Times on Baseline Trials .....	107

<b>Table 24.</b> Results of Linear Mixed-effects Model for Eye-movement Data between 200 ms after Noun Onset to Mouse Click.....	111
<b>Table 25.</b> Results of Linear Mixed-effects Model for Eye-movement Data between 200 ms and 1100 ms after Noun Onset .....	112
<b>Table 26.</b> Experimental Stimuli and English Gloss in Parentheses .....	141

## List of Figures

<b>Figure 1.</b> Adapted from Chao's (1930) Five-point Scale Theory of Mandarin Tone .....	8
<b>Figure 2.</b> Hypothetical Identification (left) and Discrimination (right) Result.....	13
<b>Figure 3.</b> Pitch Contours of /pi/ with Four Tones for Experimental Stimuli .....	44
<b>Figure 4.</b> Example of a Synthesized T2-T3 Continuum .....	45
<b>Figure 5.</b> Identification Curves Pooled across Participants and Tone Pairs .....	49
<b>Figure 6.</b> Identification Curves Pooled across Participants for Each Tone Pair .....	51
<b>Figure 7.</b> Accuracy in the Discrimination Task by Group.....	55
<b>Figure 8.</b> Discrimination Curves Pooled across Participants for Each Tone Pair.....	56
<b>Figure 9.</b> Correlations between L2 Proficiency and Identification Slope (A) and between Proficiency and Discrimination Accuracy (B) .....	60
<b>Figure 10.</b> Examples of Visual Scenes in the 3 Conditions .....	69
<b>Figure 11.</b> Accuracy (left) and Reaction Time (right) by Group and Condition .....	72
<b>Figure 12.</b> Proportion of Looks to Different AOIs over Total Looks to All Three AOIs by Condition and Group on Correct Trials .....	77
<b>Figure 13.</b> Identification Curve Averaged Across Participants and Tone Pairs by Group.....	80
<b>Figure 14.</b> Scatterplot of Identification Slope (x-axis) and Proportion of Looks to Competitor (y-axis) by Group.....	81
<b>Figure 15.</b> Examples of Triplets in the Three Training Groups.....	93
<b>Figure 16.</b> Examples of Different Trial Types.....	94
<b>Figure 17.</b> Overall Accuracy by Training Group.....	96
<b>Figure 18.</b> Overall Accuracy by Training Group and Trial Type .....	97
<b>Figure 19.</b> Overall Reaction Times for Correct Trials by Training Group .....	102
<b>Figure 20.</b> Mean Reaction Time for Correct Trials by Training Group and Trial Type .....	103
<b>Figure 21.</b> The Frequency of Raw Reaction Time data .....	104
<b>Figure 22.</b> The Frequency of Log Transformed Reaction Time Data .....	104
<b>Figure 23.</b> Differences between Mean Proportion of Looks to Target and Mean Proportion of Looks to Competitor (y-axis) by Training Group and Trial Type on Correct Trials.....	109

## Chapter 1. Introduction

Successful listening in a second language (L2) involves, among many other things, learning to identify the relevant acoustic-phonetic cues that differentiate between words in the L2, creating lexical representations and using those differentiating cues to access lexical representations during real-time comprehension. This is a particularly challenging goal to achieve when the relevant acoustic-phonetic dimensions in the L2 differ from those in the learner's native language (L1), as is the case for the L2 acquisition of Mandarin, a tonal language, by speakers of non-tonal languages like English. English-speaking learners of Mandarin must learn to (a) identify and discriminate between four different lexical tones (Yip, 2002) whose acoustic realization varies substantially across speakers and contexts, (b) establish phonological representations of tonal categories that will allow for differentiating words by tone alone, (c) integrate this information into their L2 lexical representations, and then use all relevant phonetic-phonological-lexical dimensions—segment and tone—to access those lexical representations incrementally while listening to L2 speech in real time. This is a formidable task. Indeed, while Mandarin is one of the most widely learned second/foreign languages (Goldberg et al., 2013), it is listed in “Category IV: Languages which are exceptionally difficult for native English speakers” by the Foreign Service Institute (U.S. Department of State, 2019), with the acquisition of lexical tone known to be one of the most challenging aspects for adult L2 learners of Mandarin (Shen, 1989; Wang et al., 2006).

Yet little is known about which steps in the learning task broadly outlined above may be particularly problematic and contribute to L2 learners' often long-lasting difficulties with lexical tone. The goal of this dissertation is to investigate L2 learners' processing and use of tone at several of these steps in order to gain a broader perspective and understanding of the challenges in the L2 acquisition of Mandarin tone by native speakers of English. To this end, this dissertation seeks to address three broad research questions about the L2 acquisition of lexical tone at the levels of *(1) speech perception, (2) lexical processing, and (3) word learning*:

- (1) How does language experience (native/non-native/no experience with a tonal language) affect how listeners *perceive* variation in pitch/tone on isolated syllables? (Chapter 3: Experiment 1)
- (2) How do L2 learners *process* lexical tone relative to segments in real-time spoken Mandarin word recognition? (Chapter 4: Experiment 2)
- (3) How do speakers with no tonal language experience *learn* novel words with tones under different training conditions? (Chapter 5: Experiment 3)

To address these questions, I draw on theoretical and methodological paradigms from research on native language processing in the fields of speech perception and psycholinguistics. The application of those paradigms in the field of second language acquisition (SLA) has been limited. This dissertation aims to broaden the investigation of SLA by looking at an L2 (Mandarin) that is not an Indo-European language, a domain still underrepresented in SLA research but critical in light of the large numbers of L2 learners of Mandarin worldwide. This dissertation also simultaneously contributes to research in SLA by bringing conceptual and methodological insights gained from research on tone mostly in native language processing to the study of tone in a L2, and to research in the fields of speech perception and psycholinguistics by probing the generalizability of findings about human language processing beyond native speakers.

### **The organization of this dissertation**

This dissertation is composed of three parts. Part 1 (Chapter 2) presents the necessary linguistic background and review of previous research relevant to the experimental studies presented in this dissertation. Part 2 (Chapters 3, 4, and 5) presents three lab-based experiments to address the three broad research questions outlined above. Part 3 (Chapter 6) provides a general discussion of the findings from all three experiments and their implications for theory and pedagogical practice.

Chapter 2 first presents an overview of the relevant linguistic properties of words in Mandarin. Then I present a review of the previous literature on processing tone at the phonological level (2.2.1), processing tone at the lexical level (2.2.2) and the role of tone

in word learning (2.2.3) in tonal languages, focusing on the studies that are most theoretically and methodologically relevant to the experiments in this dissertation.

Chapter 3 presents the first experiment, which investigates to what extent L2 learners identify discrete tone categories from variable auditory input. This experiment was motivated by the observation that the actual realization of a phoneme in the input is full of variance due to the different pitch ranges and speech rates of different talkers, and thus listeners have to quickly and accurately group the acoustic input into the correct phonological categories to ensure successful processing. For native speakers of Mandarin, tone perception is effortless, while L2 learners, especially at the beginning stages, may find this process very challenging. Using methods from previous speech perception experiments, which mostly focused on native listeners, I tested L2 learners' performance in identification and discrimination tasks in comparison to native Mandarin listeners and naïve listeners (English speakers with no experience of any tonal languages) to investigate to what extent language experience influences the perception of tone. This experiment also addresses an additional gap in the literature by examining L2 listeners' proficiency in Mandarin as a potentially modulating factor within the L2 group.

Chapter 4 presents the second experiment, which addresses how L2 learners use tonal cues along with segmental cues in word recognition and how their ability to perceive tone categorically is related to their ability to use tonal cues at the lexical level. Previous studies showed that L2 listeners with non-tonal L1s allocate more weight to segmental cues than to tonal cues compared to native listeners in lexical access, while L2 listeners show no large differences from native listeners in identifying tones alone. These seemingly contradictory findings inspired me to explore how tones are processed at the phonological and lexical levels in this experiment. In the experiment, I examined L2 listeners' ability to perceive tone categorically, a necessary ability for success in listening to Mandarin. Specifically, I looked at how this lower-level ability influences the weight listeners allocate to tonal cues during higher-level processing of real words. To address this question, L1 and L2 speakers of Mandarin completed an identification task similar to the one in Experiment 1 and a visual-world eye-tracking task that was designed to assess their comprehension of Mandarin words during real-time listening.

Chapter 5 presents the third experiment, which focuses on learning words in a tonal language. This experiment explores whether and how presenting contrastive cues in training influences the weight learners allocate to those phonological cues (tonal and segmental cues) in word recognition. To control the distribution of tones, vowels and consonants, a set of novel words in an artificial language were created. Native English listeners with no experience of any tonal languages were trained in different conditions. Each participant then completed a word recognition task similar to the one in Experiment 2 to measure learning outcomes. I expected to see that participants trained in a given cue-focus condition would allocate more attentional focus to that cue in word processing. To control for listeners' perceptual sensitivity to pitch, an independent pitch-perception test was included at the beginning of the experiment.

Chapter 6 provides a general discussion of the findings from the three experiments. Implications for classroom pedagogy and limitations are also discussed.

The goal of this dissertation is to advance our understanding of how L2 learners perceive, process and learn lexical tone in Mandarin through a set of experiments targeting multiple steps in tone acquisition, thus contributing to a better understanding of phonetic-phonological-lexical continuities in L2 word learning. By investigating the L2 difficulties with tone learning in a controlled laboratory environment, this dissertation also provides first-hand evidence to inform evidence-based L2 instruction and curricular materials. As such, this study may have broader impacts for the instruction of Chinese, a language that is widely learned but considered to be very difficult. More broadly, the dissertation broadens the investigation of SLA by looking at an L2 that is not an Indo-European language, a domain still underrepresented in SLA research but critical in light of the large numbers of people worldwide who learn and use Mandarin as an L2.



## **Chapter 2. Background and literature review**

### **2.1 Mandarin Chinese**

The official variety of the Chinese language used in Mainland China, commonly referred to as standard Mandarin, Putonghua, or modern Chinese, is based on the dialect spoken in Beijing. According to Yip (2002), Taiwanese Mandarin also belongs to Mandarin Chinese. The broader term “Chinese” may also refer to the many dialects spoken by people in various regions of China and by overseas Chinese communities. Besides Guan Hua, the dialect spoken in Beijing and Nanjing that is most similar to standard Mandarin, the other dialects that have the most speakers and have received the most attention from researchers are Min, Yue and Wu, all of which are spoken in the southern part of China.

For modern linguists, referring to these varieties of Chinese as dialects is problematic because for the most part they are not mutually intelligible. However, due to the “sociopolitical unity of the speakers and the powerful unifying force of a shared writing system” (Wang & Sun, 2015, p. 9), native Chinese speakers generally consider them to be dialects rather than separate languages. As the official dialect, standard Mandarin is required to be learned by all Chinese citizens from kindergarten to university. From a linguistic perspective, most Chinese citizens, especially well-educated young Chinese, can be viewed as simultaneous bilinguals because they are exposed to both standard Mandarin and their local variety of Chinese in early childhood and can speak both fluently. In this sense, native speakers of Mandarin include not only Beijing locals, who account for about 1% of the overall population in China, but also many self-identified native speakers of Mandarin throughout the country. Not surprisingly, the variety of Chinese that is most often taught as a second or foreign language is standard Mandarin (not the Beijing dialect). In this dissertation, “Chinese” will be used to refer to standard Mandarin without additional clarification, and all the native speakers of Chinese included in the data analysis were self-identified native speakers of Mandarin who learned Mandarin during childhood and continue to use it regularly and fluently in daily life.

### ***2.1.1 Segmental properties of Mandarin word***

Linguistically speaking, a morpheme is the smallest unit of a language whose meaning can be identified and isolated, and a word is defined as a syntactically free form which can stand independently in a syntactic slot. In Mandarin, morphemes are generally full syllables, and they undergo virtually no morphophonemic alternation (Packard, 2015). Though studies based on dictionaries have claimed that modern Chinese words tend to be disyllabic (Cao, 2003), corpus studies examining authentic spoken language data have shown that monosyllabic words are more frequent than disyllabic ones in Mandarin. In a study by Xiao et al. (2009), though monosyllabic words accounted for only 7.56% of the word types in a corpus of 38 million words, they accounted for 54.08% of the tokens in the same corpus. In another study, Tao (2015) found that for the top 1,000 most frequently used words, monosyllabic words accounted for 72.2% of tokens, while disyllabic words accounted for only 26.3% of tokens.

The maximal segmental structure of Mandarin monosyllabic words is CGVX, where C stands for consonant, G for glide, V for vowel, and X could be either a consonant or a glide (Wee & Li, 2015). Table 1 lists the five possible segmental structures in standard Mandarin with examples. According to Shao (2001), standard Mandarin has 22 consonant phonemes and 10 vowel phonemes. All the consonants can be the onset of a syllable, except /ŋ/. Only /ŋ/ and /n/ can be the coda of a syllable.

**Table 1**

*Segmental Structures of Mandarin Syllables with Examples*

Structure	Example
V	/a/
CV	/ba/
CVX	/ban/
CGV	/kua/
CGVX	/suan/

### **2.1.2 Lexical tone in Mandarin**

In many languages, the only suprasegmental property instantiated at the word level is lexical stress, which refers to the situation where one syllable has greater intensity than its neighbors, as in the contrast between the English words *OBject* and *obJECT*.

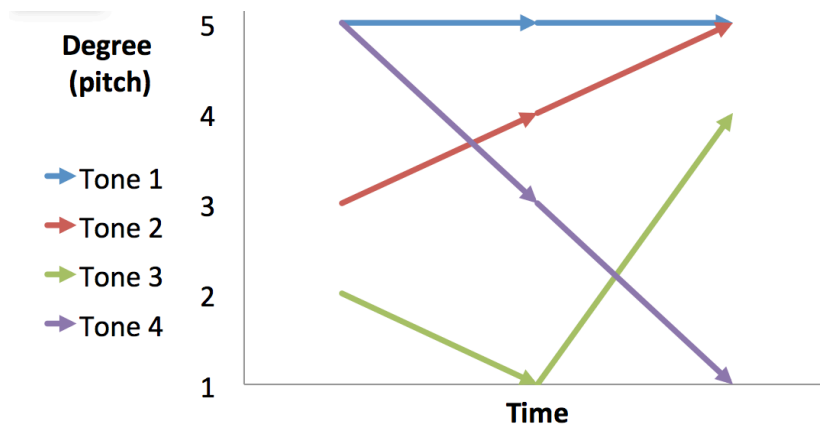
Mandarin is a tonal language. Unlike stress in English words, lexical tone is a necessary component of each Mandarin word. Different tones imposed on the same segment can change the meaning of the syllable. For example, /ma/ means *mom* if it has a flat tone, *hemp* if it has a rising tone, *horse* if it has a dipping tone, and *scold* if it has a falling tone. There are only about 400 syllables in Mandarin when tone is ignored, but there are about 1,300 when tone is included (Duanmu, 2007, 2008; Wiener & Ito, 2016). Thus, tone is critical for language processing in Mandarin.

#### **2.1.2.1 Lexical tone notation**

The tone of modern Chinese can be classified based on the five-point scale (FPS) designed by Chao (1930), which uses 1 to represent the lowest pitch and 5 to represent the highest pitch. Figure 1 shows the four Mandarin tones: Tone 1 (T1, level tone), Tone 2 (T2, rising tone), Tone 3 (T3, low dipping tone), and Tone 4 (T4, falling tone). Except for T3, which is described using three pitch points, all the tones are denoted by their beginning and end points.

**Figure 1**

*Adapted from Chao's (1930) Five-point Scale Theory of Mandarin Tone*



Though Chao's FPS was designed specifically for describing the Mandarin tone system, it has been adopted into the International Phonetic Alphabet (IPA). In IPA, the following symbols are used instead of numbers to indicate the pitch values: ˥ 'extra high', ˨˨˨ 'high', ˨˨ 'mid', ˨ 'low', and ˩ 'extra low.' For example, Mandarin T3 is transcribed as ˨˨˨. Wang (1967) introduced a feature-based binary system that uses three features of pitch height [high, central, mid] and four features of shape [contour, rising, falling, convex]. For example, Mandarin Tone 1 is [+high, -central, +mid, -contour, -rising, -falling, -convex]. Wang's model is designed to notate more complicated tonal systems than the one of Mandarin. Since this dissertation is concerned with Mandarin tone processing rather than tone typology, Chao's widely-used FPS system is easy and more straightforward to classify Mandarin tone.

#### 2.1.2.2 Tone sandhi

Tone sandhi, or the alternations that tone undergoes in certain environments, is an important independent topic in research on Mandarin tone (Wee & Li, 2015). The most widely studied case of sandhi in Mandarin is T3 sandhi, by which the first of two consecutive T3 tones change to T2 (FPS: 214 → 35). The difficulty that L2 learners experience with understanding Mandarin sentences involving tone sandhi has received increasing attention among psycholinguistic researchers in recent years (e.g., Cheng et al., 2014; Huang, 2001). However, this dissertation does not contain an in-depth discussion

of tone sandhi because it (a) uses isolated words pronounced in citation form to avoid tone sandhi and (b) focuses on the pitch perception and word processing of isolated monosyllabic words with standard tones.

### 2.1.2.3 The mental representation of lexical tone

The mental representation of lexical tone is an important topic, but it is not an easy one to discuss because many questions related to it are still not resolved (Best et al., 2019). Although the primary purpose of this dissertation is to explore tone processing by L2 learners and not to make theoretical claims about the mental representation of lexical tone, it might still be helpful to provide a brief description of some questions related to how tones are represented in the mind; doing so will also lay important groundwork for the discussion section of this dissertation.

Debates surrounding the mental representation of lexical tone are part of a broader discussion of how phonemes are represented in the mind. Since mental representations cannot be measured directly, one question that needs to be dealt with is how abstract they are. Though multiple theoretical models have been introduced, there is still no clear answer on where the mental representation of phonemes is on the spectrum of abstractness (e.g. Cutler, 2008; Pierrehumbert, 2016; Sumner & Samuel, 2005; Tenpenny, 1995). However, the perspective that I find most convincing is that the abstract mental representation of lexical tone exists within a cloud of episodic information (Pierrehumbert, 2016; Tenpenny, 1995).

Previous studies have generally found that native listeners perceive Mandarin tone more categorically than naïve listeners (e.g., Hallé et al., 2004; Peng et al., 2010; Wang, 1976; Xi et al., 2010; Xu et al., 2006). The similar patterns of categorical perception (CP) observed across different studies suggest that there is some kind of abstract canonical mental representation of tone categories, at least for native speakers. However, the mental representation of tone cannot be completely abstract and must include some episodic information, otherwise listeners would have difficulty handling the high amount of variance in the input. Previous studies show that successful listeners need to adjust their perception to different talkers. For example, successful perception of T2 and T3, which

have similar overall direction of fundamental frequency (F0), requires listeners to adjust their perception according to talker-specific pitch ranges because a T2 produced by a low-pitched talker (e.g., a male) might be similar in terms of F0 height to a T3 produced by a high-pitched talker (e.g., a female; Moore & Jongman, 1997). If listeners fail to access the episodic information related to pitch ranges stored in their minds, successful listening in Mandarin could be greatly delayed if not totally disrupted.

L2 learners' mental representations of lexical tone in tonal languages are also complicated by influences from their L1 prosodic systems. Previous researchers agree that L2 learners assimilate lexical tone in their L2s to the prosodic contrasts in their L1s, especially at beginning stages of learning (Hallé et al., 2014), while whether and how such assimilation may help or hinder learning lexical tone in an L2 tone language is mostly unanswered (Best, 2019). Theoretical models, such as the Perceptual Assimilation Model (PAM: Best, 1995; So & Best, 2008) and the Speech Learning Model (SLM: Flege, 1995) have been developed to account for cross-language perception of non-native consonants and vowels, and they have also been extended to make predictions about the cross-language perception of lexical tone. For instance, by instructing English listeners to classify Mandarin tones in terms of English intonation contours, So and Best (2008) provided evidence to support tone assimilation. However, none of the models were designed to explain the mental representation of lexical tone in word processing by non-native speakers or whether and how assimilation works in L2 lexical tone processing. As Best (2019) commented in a recent paper, many questions surrounding mental representations of lexical tone and L2 tone processing are unresolved and require more carefully designed research to explore "how [perception of lexical tones] changes developmentally in both native and non-native listeners" (Best, 2019, p. 5).

## **2.2 Tone in L1 and L2 processing**

### ***2.2.1 Processing tone at the phonological level***

Native speakers of a language generally do not realize how challenging the process of identifying phonemes can be in their mother tongue. However, anyone who has some experience with learning an L2 knows how difficult it can be to identify sounds in

spontaneous speech, especially at the beginning stages of acquisition. Although every language has a limited number of phonemes, the acoustic realizations of those phonemes could be infinite. The acoustic variants of a phoneme may differ along a variety of parameters, such as F0, voice onset time, and duration due to different talkers (e.g., male vs. female) or contexts (e.g., casual conversation vs. formal presentation). In order to process tonal information efficiently, listeners learn to pay attention to the critical features that differ between phonological categories. For example, F0 is regarded as the most important parameter to differentiate lexical tone in Mandarin, though duration, intensity, and turning points may also matter (Gandour, 1978, 1983; Jongman et al., 2006). The lack of invariance in daily speech requires listeners to develop a mechanism for identifying the phonological categories efficiently and with more perceptual sensitivity to between-category differences than within-category differences. Thus, investigating how listeners perceive phonological cues from variable acoustic signals is critical to understanding how humans process language. In experimental research, this is called categorical perception (CP).

CP is a phenomenon by which the mental categories possessed by an individual influence perception. Though this phenomenon exists in many domains of cognition (e.g., categorization of colors in vision; see Goldstone & Hendrickson, 2009), the ability to perceive sound input categorically is fundamentally important for word recognition and language processing. Without it, communication would be impossible (Schouten et al., 2003).

Experimental research on CP has tested how an individual discriminates phonemes belonging to different categories after controlling the physical differences between them (Goldstone & Hendrickson, 2009). Identification and discrimination of synthetic stimuli has long been used to investigate how listeners recognize discrete phonemes from highly variable acoustic input (Schouten et al., 2003). Earlier research mostly focused on the CP of consonants and vowels (e.g., Gerrits & Schouten, 2004; Liberman et al., 1957). Inspired by their fruitful achievements, some researches have started to use this method to investigate the perception of lexical tone (Yip, 2002).

In the following sections, I introduce the methods that have been used to measure CP of vowels, consonants, and lexical tones. The goal of Section 2.2.1 is to provide

theoretical and methodological background for Experiment 1 on the CP of Mandarin tone by L2 learners with different proficiency levels.

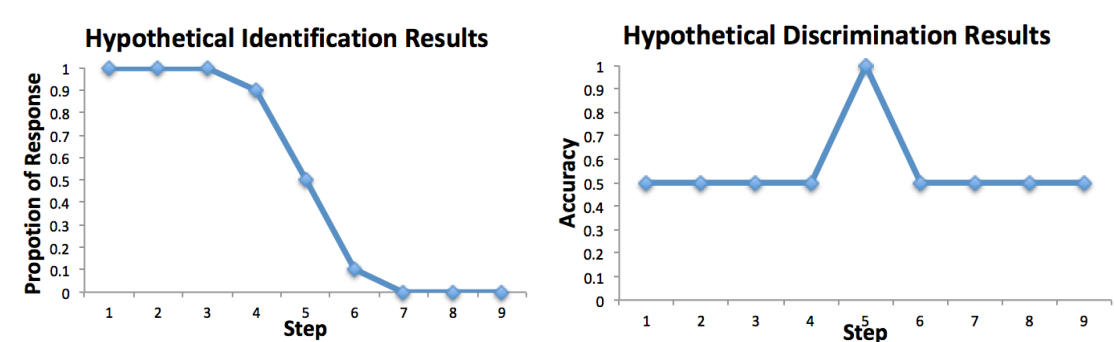
#### 2.2.1.1 Methods for measuring categorical perception (CP)

The concept of categorical perception (CP) was first introduced into the field of linguistic research by Liberman et al. (1957) to explain the phenomenon of people reducing the number and variety of sounds they hear by classifying them into discrete phonological categories for efficient comprehension. In their CP study, multiple steps of sound tokens were created from a consonant pair: /b/ and /d/, by imposing linearly changing F2 (the critically differing parameter) and controlling other parameters. The resulting tokens sounded like something in between /b/ and /d/. In the identification task, listeners were asked to label each token they heard as /b/ or /d/. In the discrimination task, three sound tokens were presented to listeners in a row, and the last sound was identical to either the first or the second sound. Listeners were instructed to say whether the last sound they heard was the same as the first or the second sound. It was assumed that if listeners perceived the tokens in the pair continua categorically, the identification curve would show a sigmoid shape with a steep slope, and the curve of discrimination accuracy would have a peak at the categorical boundary (see Figure 2 for illustration). Liberman and his colleagues also proposed the original strong form of CP according to which a listener “can discriminate the stimuli only to the extent that he can identify them as different phonemes” (p. 362). In other words, the same mechanism was assumed to be used by listeners in both discrimination tasks and identification tasks. This is the most stringent standard for CP (Francis, 2003; Gerrits & Schouten, 2004; Liberman et al, 1957).



**Figure 2**

*Hypothetical Identification (left) and Discrimination (right) Results*



However, even in Liberman et al.'s (1957) own research, this standard could not be met. Their results showed that the discrimination curves predicted from identification results showed stronger CP patterns than the observed discrimination curves. On the other hand, the observed discrimination curves showed overall higher accuracy with less obvious peaks than the predicted discrimination curves, indicating that the two tasks may not measure exactly the same things. Gerrits and Schouten (2004) conducted a series of experiments to test Liberman et al.'s (1957) assumption that there is a single mechanism behind identification and discrimination. They found it was impossible to follow the stringent standard of CP. Speech perception is bidirectional, influenced both by listeners' mental representations of phonological categories (top-down) and their perceptual sensitivity to acoustic signals (bottom-up; McClelland et al., 2006; Werker & Logan, 1985). Compared with identification, discrimination is more influenced by bottom-up processing, which indicates that there is a more fundamental difference between identification and discrimination (Gerrits & Schouten, 2004).

As a result, Harnad (2003) proposed a more liberal and practical standard of CP that involves comparing the target group's performance to a baseline group's performance. In the case of lexical tone perception by native Mandarin listeners, the baseline group usually consists of naïve listeners with no experience in Mandarin (e.g., Hallé et al., 2004; Peng et al., 2010; Xu et al., 2006). There is no absolute standard of CP in Harnad's (2003) method, and one may only conclude that the target group perceives sound stimuli more or less categorically than the baseline group or another comparison group.

Aside from introducing new experimental methods into this line of linguistic research, Liberman et al.'s (1957) study also raised awareness among linguists that CP is a widespread phenomenon in L2 language processing. In order to make the concept of CP more understandable to readers, Liberman et al. (1957) introduced CP using common experiences encountered while learning an L2, where "one often has difficulty in making all the appropriate sound discriminations" (p. 358). They assumed that the difficulty of tone perception was partially due to L2 learners' less successful categorization of acoustic signals in an unfamiliar context. However, though Liberman and his colleagues raised this hypothesis, most following research has focused on speech perception by native speakers. The mechanisms involved in L2 processing might be very different from those used by native speakers due to differences in language experience. Before giving an overview of L2 research on CP, I will first introduce the Automatic Selective Perception (ASP) model (Strange, 2011), a working model that is useful for understanding differences in speech processing between native and L2 listeners.

#### 2.2.1.2 The Automatic Selective Perception (ASP) model

Strange's (2011) ASP model is a model for characterizing how native listeners and L2 learners detect reliable cues from acoustic signals for speech perception. According to Strange, the key difference between native listeners and L2 learners is their language experience. For native listeners, processing speech in their L1s is very automatic and effortless due to their lifelong "highly over-learned selective perception routines (SPR)" (p. 456), which are attuned to the most reliable acoustic-phonetic information available for speech perception and word recognition. As for L2 learners, who have already built complete L1 perception routines, they have to learn to process a new language and adjust their familiar L1 perception routines for more effective processing in an L2. Strange (2011) also defined the following terms, all of which are important to understanding the experiments in this dissertation:

### *Auditory salience and perceptual salience*

Auditory salience refers to the magnitude of the universal physiological response to acoustic signals by the general population. Perceptual salience, on the other hand, is the magnitude of the physiological response to acoustic signals that varies as a function of language experience and experimental manipulation. For example, the contrast between /r/ and /l/ is more perceptually salient to English native speakers than to Japanese native speakers, since in English /r/ and /l/ belong to two different phonemes, while Japanese does not differentiate between /r/ and /l/. Due to differences in language experience, the same language stimuli may have very different degrees of perceptual salience for native listeners and L2 learners.

### *Selective attention and attentional focus*

In the ASP model, selective attention, which is closely related to the notion of cue-weighting as described by Jusczyk (1997), refers to the automatic process by which some acoustic-phonetic cues have a greater influence on behavior response than others in language processing. For example, though Mandarin lexical tones differ along multiple parameters (duration, intensity, and pitch), pitch is the one that has the greatest influence on tone processing among native speakers (Jongman et al., 2006). According to the ASP model, native Mandarin speakers pay more selective attention to pitch than duration or intensity while processing lexical tone. Attentional focus, on the other hand, refers to listeners' goal-oriented conscious attunement to various types of information. For example, when listeners are instructed to explain how two tones differ in an experimental setting, where stimuli vary along multiple dimensions, they are more likely to place their attentional focus on tone and ignore changes on other dimensions (Tong et al., 2008). Thus, experimental manipulation can have a direct effect on the amount of attentional focus participants allocate to different types of information.

### *Short-term, long-term, and procedural memory*

Short-term memory consists of memory traces, which last a very short amount of time, and long-term memory consists of knowledge and mental representations, which usually do not fade away. For the most part, short-term and long-term memory are both

considered to be types of declarative memory. Within the ASP model, procedural memory is another important component of memory that consists of language processing routines. Native listeners, for instance, are able to process incoming linguistic information in their L1 efficiently and automatically mostly due to the automatic selective routines in their procedural memory.

### *Phonetic and phonological mode*

In the ASP model, two modes of perception are discussed that differentiate between how native speakers and L2 learners process language. Native listeners normally process incoming linguistic information in a phonological mode, which optimizes the processing of phonological contrasts instead of within-language phonetic variations. On the other hand, the phonetic mode of perception is commonly used at the beginning stages of L2 learning, where naïve listeners process cross-language and within-language variations and could fail to attend to phonological contrasts between non-native phonemes that are not distinctive in their L1s. The perceived information is stored in short-term memory for immediate use. With more exposure to the L2, learners might eventually start processing acoustic information in a more phonological mode, and as a result their L2 processing might become more native-like as proficiency increases.

#### 2.2.1.3 Tone perception by native and naïve listeners

Research going back to Wang (1976), who conducted an identification task and a discrimination task with stimuli on a continuum from T1 to T2, found evidence of CP with more sigmoid shapes of identification curves and more obvious discrimination peaks by native listeners compared to naïve listeners. However, when Abramson (1977) conducted a similar study in Thai with a continuum involving two level tones, he found no evidence of CP by native listeners of Thai. These findings indicated that tone similarity might influence native listeners' perception of tone. This assumption was tested by Francis et al. (2003) in Cantonese using stimuli varying along three continua: one ranging from a low level to a high level tone, one from a high rising to a high level tone, and one from a low falling to a high rising tone. Identification data from the last two tone

continua showed evidence of CP, but no such evidence was found in the data from the first tone continuum, suggesting that tones sharing similar pitch directions are not categorically perceived.

Realizing the influence of tone pair similarity and the absence of absolute criteria for CP, Hallé et al. (2004) created a set of experimental materials with stimuli varying along three continua—one level-contour tone continuum (T1-T2) and two contour-contour tone continua (T3-T4, T2-T4)—and compared the identification and discrimination performance of native Mandarin listeners to that of French listeners with no experience of Mandarin. The results showed that native listeners perceived all three continua more categorically with some degree of difference among continua, while French listeners' tone perception was more psychophysically based and generally non-categorical. From the perspective of the ASP model, it might be the case that the native and French listeners performed differently because they used different mechanisms to process tonal information. In the phonological mode, native listeners are able to use their knowledge of phonological categories of tone to identify tokens efficiently. As a result of their lifelong experience with the language, native listeners are more sensitive to between-category differences than to within-category differences, even though acoustically those differences could be the same in value. Without knowledge of tone, French listeners could only rely on fine-grained acoustic details and complete the tasks in a phonetic mode.

There is evidence that tone perception by native speakers is also more automatic than that by L2 learners. Xu et al. (2006) administered a series of identification and discrimination tasks to native Mandarin speakers and native English speakers with no formal experience of learning Mandarin. The stimuli consisted of both Mandarin speech samples and non-speech harmonic sound samples manipulated to form a continuum from T1 to T2. The English listeners demonstrated more sensitivity to between-category differences in the non-speech stimuli than in the speech stimuli, while the Mandarin listeners showed high sensitivity to between-category differences in both the speech and non-speech stimuli. These findings suggest that naïve listeners may construct temporary tone representations while completing a task and use them to perceive tone variation in simple non-speech stimuli, but that working memory constraints may prevent them from

employing the same strategies with more phonologically complex stimuli involving Mandarin speech samples. On the other hand, the native listeners' tone perception was more automatic, so it was less influenced by the complexity of the stimuli.

Xi et al. (2010) provided neurophysiological evidence that tone perception is more automatic in native listeners than in naïve listeners in an event-related potential (ERP) study. Though the native listeners in the study were not asked to do anything while listening to the auditory stimuli, trials containing sounds that belong to different categories (i.e., across-category stimuli) still elicited larger mismatch negativity (MMN, a brain response that occurs when one auditory token is followed by another that does not match it) in their left hemispheres (related to linguistic processing) than trials containing sounds that belong to the same category (i.e., within-category stimuli). Similar neurophysiological evidence was found with tone imposed on both speech and non-speech segments, reflecting the automatic processing of across-category contrasts regardless of linguistic environment. Since this automatic processing in native listeners is conditioned through lifelong experience with the language, the CP of tone is very stable across different tasks and stimulus types.

Naïve listeners have to rely on their sensory traces of acoustic input when discriminating between consecutive tokens because they do not have mental representations of the phonemes being tested. As Xu et al. (2006) pointed out, using acoustic details can be an effective strategy for naïve listeners to achieve high accuracy in a discrimination task. For example, Hoffmann et al. (2014) tested native Mandarin listeners and native Dutch listeners with no prior experience of learning Mandarin in perceiving several sets of T1-T4 continua with various degrees of between-step differences. They found that changing the amount of variance in the stimuli did not influence Mandarin native listeners' discrimination patterns, whereas it greatly affected Dutch native listeners' discrimination performance. This result suggests that processing by naïve listeners is mostly limited to fine-grained acoustic details, while native listeners process tone information by accessing mental categories, which gives them increased sensitivity to between-category differences. Interpreting these findings within the ASP model (Strange, 2011), naïve listeners process acoustic information in a more phonetic mode, while native listeners process information in a more phonological mode.

#### 2.2.1.4 Tone perception by L2 learners

Though naïve listeners' tone perception is generally psychophysically based and non-categorical, training can improve their performance. For instance, Wang et al. (1999) trained native English speakers with no previous experience of tonal languages to identify Mandarin tone in eight lab sessions. The results showed significant improvement on the post-test. Similarly, Lu et al. (2015) found that training improved English naïve listeners' discrimination of Mandarin tone. After training, English listeners also showed more native-like neurophysiological activation with ERP. Zhao and Kuhl (2015) trained English speakers to discriminate between tokens on a T2-T3 continuum and found significant improvement in discrimination on the post-test from pre-test, although their discrimination patterns were still different from those of the native speakers. However, the participants in the studies mentioned above were all naïve listeners who might just have been showing short-term learning effects due to having received a limited amount of training in a laboratory setting. For non-native speakers who have experience of learning Mandarin as a L2, the processing of Mandarin lexical tone might be different from naïve listeners and one cannot simply extend findings from naïve listeners to L2 learners.

Recently, some researchers have started to test L2 learners with various tasks to understand how learning experience influences their perception of tone. Wang et al.'s (2004) dichotic listening study with neurophysiological measurement provided evidence that native Mandarin listeners and advanced L2 learners display similar left-hemisphere dominance for the perception of Mandarin tone. However, in this study, the L2 listeners had started to learn Mandarin in their early life and all of them had reached high proficiency, which might not be an accurate representation of the majority of L2 learners of Mandarin. Another study testing advanced Mandarin learners by Pelzl et al. (2019) also found no significant differences between L2 learners and native Mandarin speakers in identifying Mandarin tone in isolated monosyllabic words. In a lexical decision task, however, L2 learners were much more likely to reject segmentally mismatched non-words than tonally mismatched non-words, while native speakers showed similar rejection rates in both those mismatched conditions. This suggested that tone presented more difficulty for L2 learners than for native listeners. This result was further confirmed

with neurophysiological evidence that advanced L2 learners performed much better in identifying isolated tones than in making lexical decisions (Pelzl et al., 2020).

In order to understand if and how L2 learners access different tone categories from highly variable acoustic input, synthetic tone pair continua have also been used to test their processing of tone variation. In an experiment by Shen and Froud (2016), English-speaking L2 learners with advanced Mandarin proficiency were instructed to identify and discriminate between tokens that had been manipulated such that they varied along continua from T1 to T4 and from T2 to T3. Each token was presented at the end of a carrier phrase. Native listeners and English-speaking naïve listeners participated in the same identification and discrimination tasks for purposes of comparison. Results generally showed that L2 learners had similar perception patterns to native listeners on both tasks, suggesting that L2 learners with advanced proficiency perceive tone categorically. However, in a follow-up ERP study, Shen and Froud (2019) found that native Mandarin speakers, but not L2 learners, showed neurophysiological evidence for CP of tone. L2 learners also showed larger P300 responses, suggesting that they placed more attentional focus on the processing of tone than the native speakers did. Besides differences in research design and measurement between these studies, Shen and Froud (2016) mentioned several additional factors that might have influenced the performance of the L2 learners in the earlier study. First, the stimuli were presented at the end of carrier phrases, where sentential/phrase-level intonation might have influenced processing on a phonological level (Camp & Schafer, 2016). Another point to consider is that both of the studies only tested advanced L2 learners on two tone-pair continua (T1-T4 and T2-T3). Little is known about how learners with different proficiency levels would perform on a similar set of tasks or how the inclusion of other tone pair continua might influence performance.

To address gaps in the existing research on the perception of tone variation by L2 learners of Mandarin, Experiment 1 (Chapter 3) was designed to investigate L2 tone perception by testing English-speaking learners of Mandarin with varying levels of proficiency on all six possible Mandarin tone-pair continua in both identification and discrimination tasks. Mandarin native listeners and English-speaking naïve listeners were also tested in the same experiment as comparison groups. The L2 learners' Mandarin



proficiency was assessed through a combination of self-ratings and a listening test so that I could assess how proficiency scores correlate with performance on the perception tasks. The goal of Experiment 1 was to explore how L2 learners differed from native Mandarin speakers and naïve listeners in terms of the perception of tone variation at the phonological level.

As Shen and Froud (2019) pointed out, “[i]f phonetic categorization by adult learners is less efficient and requires more attention compared to native Chinese speakers, higher-level processes such as lexical access may also be more challenging for learners” (p. 263). In the next section, I will review the literature on tone processing at the lexical level.

### ***2.2.2 Processing tone at the lexical level***

In spoken word recognition, listeners not only need to perceive phonological cues from a rapidly changing acoustic signal, but they also need to access their mental lexicon to connect the perceived phonological representations with lexical meanings. Although words are usually recognized effortlessly by native speakers, this process is in fact very complex and involves rapid incremental updating as the speech signal unfolds. The complexity of this process often becomes more obvious in an L2, where processing is less automatic and requires more cognitive resources (Strange, 2011; Xi et al., 2010).

As in an L1, word recognition in an L2 involves at least two critical steps: (a) perceiving phonological categories from acoustic input (as discussed above), and (b) using perceived phonemes to activate word candidates in the mental lexicon (Cutler, 2012). Infants are born with sensitivity to a wide variety of sound contrasts and soon tune in to those that are meaningful in the language(s) they are exposed to (Werker & Tees, 1984; Yeung et al., 2013). This leads to higher sensitivity to language-specific contrasts than to contrasts irrelevant in the L1(s). By adulthood, L1 phonological categories are deeply entrenched, allowing L1 listeners to assign acoustic input to phonological categories rapidly and without effort. According to the ASP model, the over-learned processing routines in native listeners’ procedural memory enables them to quickly use the most relevant cues for word recognition automatically (Strange, 2011).

For L2 learners, on the other hand, the challenge is not only to learn potentially new phonological categories, but also to activate and use this knowledge to identify phonological contrasts in a highly variable speech signal. As discussed in the previous section, laboratory experiments have provided ample evidence that L2 learners perceive L2 phonemic variants less categorically than native listeners do (Shen & Froud, 2019), and that they experience difficulty with phonological contrasts that do not exist in their L1(s) (Pelzl et al., 2019; Qin, 2017).

These difficulties at the phonological level in L2 speech perception might lead to further challenges at the level of lexical access during L2 word recognition. As a result of not being able to efficiently utilize all relevant phonological cues, it is likely that L2 listeners will activate a greater number of word candidates, leading to more extensive competition for selection compared to L1 processing (Cutler, 2012; Weber & Cutler, 2004). The following two sections will discuss previous studies on tone processing at the lexical level by native speakers and L2 learners to provide the relevant theoretical and methodological background for Experiment 2 (Chapter 4).

#### 2.2.2.1 Processing tone at the lexical level by native speakers

Most previous work on word processing in tonal languages has focused on the independent contributions of tonal vs. segmental information to lexical access. Early studies tended to find that tones provided a weaker cue than segments did. Using a homophone judgment task of written characters in Mandarin, Taft and Chen (1992) found that native listeners took longer and were less accurate to say ‘no’ when a pair of words differed only by tone than when they differed only by vowel quality. Cutler and Chen (1997) examined how native speakers of Cantonese (a tonal language similar to Mandarin, but with six tones) process lexical tone in spoken words. Results from a lexical decision task showed that listeners were most likely to accept a non-word as a word when the stimuli only differed in tone. A speeded same-different judgment task with the same participants showed slower and less accurate responses when the only difference between two words was in tone. Similar findings were shown with Dutch native speakers without any knowledge of Cantonese in the speeded same-different judgment task, suggesting

that the late processing of tone relative to segments is mainly due to perceptual processing rather than linguistic processing of tone distinctions. Cutler and Chen (1997) concluded that lexical tone might be a weaker cue because segmental cue from the onset consonant appeared earlier in the signal than tonal cue. Wiener and Turnbull (2016) also found that tonal information does not constrain lexical access as tightly as segmental information in Mandarin. In their experiment, native Mandarin speakers were asked to change word-like nonwords (e.g., *su3*) into words by changing a single component (e.g., consonant: *tu3*; vowel: *si3*; tone: *su4*). Results showed that native Mandarin speakers were more likely to change a tone than a vowel or a consonant to make a nonword into a word, and they were also faster in making tone changes than in making segment changes. However, all the studies discussed above used isolated syllables without much context, and daily communication provides much more context for word processing.

In a simple tone-vowel detection task with isolated syllables, Ye and Connine (1999) replicated previous research and confirmed that vowel information has a perceptual advantage over tone information. But when similar stimuli were presented at the end of an idiomatic phrase, where tone is highly predictable, tone information showed a processing advantage over vowel information in the form of shorter reaction times and higher accuracy rates. Thus, the relative contributions of tonal and segmental information in word processing are not absolute; instead, it depends on context.

As pointed out by Liu and Samuel (2007), the tasks employed in many early studies are mostly sensitive to sub-lexical processing, where listeners can respond without accessing lexical meaning. In tasks requiring listeners to access lexical representations, previous research has found more similar roles for tonal and segmental cue. Malins and Joanisse (2010) conducted a real-time spoken word recognition experiment using the visual world eye-tracking paradigm (VWP). A basic assumption underlying this paradigm is that the probability of fixation on any given object in a visual scene reflects the corresponding item's level of activation in the mental lexicon (Tanenhaus et al., 2000). Previous studies have consistently found strong cohort competition between words sharing the same initial segments (e.g., *beaker* vs. *beetle*; Allopenna et al., 1998; Tanenhaus et al., 1995) and comparatively weaker and delayed competition between words sharing the same rhyme (e.g., *beaker* vs. *speaker*; Allopenna et al., 1998).

In Malins and Joanisse's (2010) study, each visual scene consisted of a target item (e.g., *chuang2* 'bed'), a phonological competitor item, and two phonologically unrelated distractor items. There were four kinds of phonological competitors: (a) cohort competitors (e.g., *chuan2* 'ship'), (b) rhyme competitors (e.g., *huang2* 'yellow'), (c) segmental competitors (e.g., *chuang1* 'window'), and (d) tonal competitors (e.g., *niu2* 'cow'). Participants were instructed to click on the picture corresponding to the word they heard. If spoken word recognition is similar in Mandarin and Indo-European languages, then strong cohort competition and weaker rhyme competition should be observed. Of critical interest were the segmental competitor (c) and cohort competitor (a) conditions because the tonal divergence of segmental competitors from targets and the segmental divergence of cohort competitors from targets arguably occur at roughly the same point in time (see the Appendix in Malins & Joanisse, 2010). If native speakers process tonal and segmental information simultaneously, similar patterns of competition should be found between the two conditions. The tonal condition was included to see whether tone alone can trigger competition.

Although the results showed no evidence of rhyme competition or tonal competition, strong cohort competition was observed. Crucially, the eye gaze patterns did not differ between the cohort and segmental conditions, leading Malins and Joanisse to conclude that "tonal and segmental information are accessed concurrently and play comparable roles" (p. 407) in word recognition by Mandarin native speakers. A follow-up study (Malins & Joanisse, 2012) using an ERP paradigm supported the conclusion that tonal and segmental cues were both accessed as soon as they were available. The results also suggested that there are potentially different underlying processing mechanisms for tonal vs. segmental information, with less persistent and more left-lateralized activation in the segmental condition than in the cohort condition.

So far, I have discussed two lines of research on the role of tone relative to segments, each leading to somewhat different conclusions. In experiments encouraging sub-lexical processing and not requiring access to lexical meanings, tone appears to be a weaker cue than segments (Cutler & Chen, 1997; Ye & Connine, 1999). In spoken word recognition tasks, which require full lexical access, tone and segments appear to be on more equal footing (Malins & Joanisse, 2010, 2012).

A third line of research focusing specifically on the interaction between tonal and segmental cues has provided evidence of a dynamic relationship between different cues. This research draws attention to the limitations of isolating individual cues in experimental contexts, thereby emphasizing the importance of processing tonal and segmental information simultaneously in successful word recognition. In a form priming task by Sereno and Lee (2015), no priming effect was found when the prime and the target matched only in tonal content, and a weak effect was observed when they matched only in segmental content. However, a much larger priming effect was found when the prime and the target matched in both segmental and tonal content (see also Lee, 2007), indicating that the segment and tone as one unit, rather than its individual components, is what plays a critical role in word processing.

The predominantly monosyllabic structure of Mandarin words makes it possible to process segmental and tonal information together quickly and efficiently as a single unit. In an ERP study, Zhao et al. (2011) investigated how native Mandarin listeners weigh segmental and tonal information while processing words. In the experiment, participants were asked to compare two pictures presented sequentially on a computer screen and decide whether they belonged to the same semantic category. A word was also presented auditorially between the two pictures. This word either (a) partially mismatched the second picture for onset, rhyme or tone, or (b) completely mismatched the second picture in terms of both tonal and segmental information (i.e., the entire syllable). The results showed that all three partial mismatch conditions triggered N400 effects of similar time course and amplitudes, while the syllabic mismatch condition led to earlier and stronger N400 effects. In ERP studies, an N400 effect signals a violation of expectations during language processing. Based on these results, Zhao et al. (2011) argued that holistic syllable-based processing is the most important type of processing in Mandarin.

Using the Garner speeded classification paradigm (Garner, 1974, 1976; Garner & Felfoldy, 1970), Tong, Francis and Gandour (2008) investigated the interaction between tonal and segmental cues from a different angle. In this task, participants have to classify stimuli according to a specific dimension (e.g., consonant, tone, or vowel) while ignoring variation along other dimensions. Results from native Mandarin speakers showed that variation in vowel or consonant quality interfered more in the classification of stimuli by

tone than the reverse, which led the authors to conclude that the processing of tone was not independent, but instead integrated into the processing of segments.

Further evidence comes from another speeded classification experiment conducted by Lin and Francis (2014). Using stimuli consisting of legitimate words in both English and Mandarin, they found that L1 Mandarin listeners showed symmetric interference between consonant and tone, regardless of whether the stimuli were presented in an English or Mandarin context, while L1 English listeners did not show any interference from the non-target dimension in either direction. Lin and Francis suggested that native Mandarin listeners process tonal and segmental cues in a more integrated manner regardless of the ambient language, while English listeners with no experience of tonal languages process them more separately.

These findings are consistent with the notion of selective perception routines (SPRs, see 2.2.1.2) as proposed in Strange's (2011) ASP model: L1-Mandarin listeners have developed a routine for perceiving and processing tonal cues along with segmental cues during lexical access through lifelong experience with a tonal language. This SPR constitutes such a highly over-learned pattern that even when the instruction and stimuli are in English, native Mandarin listeners are unable to inhibit the automatic processing of tone. On the other hand, English listeners unaccustomed to processing pitch at a lexical level might treat pitch as intonation and process it separately from segmental content. Thus framed within the ASP model, the task for L2 learners is to inhibit their entrenched L1 SPRs and learn new SPRs that optimize use of the most reliable linguistic cues in the L2. A key goal of Experiment 2 is to gain a better understanding of the extent to which L1-English learners of Mandarin are able to go about this process in the context of using tonal and segmental cues during lexical processing.

#### 2.2.2.2 Processing tone at the lexical level by L2 learners

Though lexical tone is notoriously difficult for speakers of non-tonal languages to learn, previous work has provided ample evidence that listeners with non-tonal language backgrounds are sensitive to pitch contrasts (Hallé et al., 2004) and that accuracy of tone identification can be improved significantly with even short-term training (Cooper &

Wang, 2013; Wang, 2013; Wang et al., 1999; Wong & Perrachione, 2007). For example, Wong and Perrachione (2007) demonstrated that L1-English learners of Mandarin with limited exposure to Mandarin are able to identify tones in syllables with above-chance accuracy after short-term training in a laboratory setting. Tone-bearing syllables in different phonetic contexts and spoken by different native speakers of Mandarin were used to train participants. Results showed great improvements in accuracy from the pre-test to the post-test, and these improvements also carried over to new stimuli involving different segments and produced by different talkers. Most importantly, a retention test showed that these improvements were still observable after six months.

The same study also found that training improved non-native listeners' identification of words with tone. English native speakers with no previous exposure to a tonal language were also trained to associate an image of an object with a pseudo-word with tone in a manner comparable with L2 word learning in a classroom setting. After each of the four sessions, participants took a quiz in which they needed to select the picture corresponding to the word they heard. Overall, participants' identification accuracy increased from session to session and they had more difficulty in learning tones than segments. The data also showed large individual differences, suggesting that learning success was associated with listeners' aptitude for pitch perception and previous musical experience.

Although listeners with little or no experience with a tonal language seem to improve a lot after even short-term tone perception training, there is also abundant evidence that L2 learners with intermediate to advanced proficiency still show persistent difficulty with processing of tone at the lexical level (Pelzl et al., 2019, 2020; Qin, 2017). A recent study by Pelzl et al. (2019) directly addressed this gap between perception and lexical processing of tone in an L2. They found that although English-speaking learners of Mandarin were as successful as native Mandarin speakers at identifying tone in isolated syllables in a tone identification task, the same learners were less likely than native speakers to correctly reject non-words in a lexical decision task when the non-words and words differed only by tone, thereby indicating a “disconnect between L2 abilities to categorize tones as phonetic objects and abilities to utilize those categories as lexical cues” (p. 69). A third task using an ERP paradigm provided consistent neurophysiological

evidence that learners are more sensitive to segmental mismatches than to tonal mismatches, suggesting that learners have more difficulty using tonal cues than segmental cues during lexical processing. These findings constitute important evidence for a dissociation between phonological and lexical processing ability related to tone in an L2 based on group-level analyses comparing L2 learners to L1 speakers on different tasks. One area that has not yet been explored is how these abilities relate to each other at the level of the individual learner—a question that I aim to address in Experiment 2.

### 2.2.2.3 The relation between tone processing at phonological and lexical levels by L2 learners

Many factors can contribute to the dissociation between phonological and lexical processing of tone by L2 learners. At a phonological level, although L2 learners are able to achieve native-like accuracy in identifying the four standard Mandarin tones in isolated syllables (Pelzl et al., 2019), L2 listeners may have more difficulty handling the lack of invariance of tone realization in actual speech, where they have to assign a specific token quickly to a phonological tone category. As discussed in the previous section, Shen and Froud (2016) administered an identification task involving T1-T4 and T2-T3 pairs to advanced learners of Mandarin and native Mandarin speakers, with results showing that the two groups had similar patterns of performance. However, in a follow-up ERP study (Shen & Froud, 2019), L2 learners showed weaker electrophysiological evidence of categorical perception of tones than native Mandarin speakers did. Shen and Froud suggested that this less efficient lower-level of phonetic categorization by L2 learners may lead to more processing difficulty in higher-level tasks, such as lexical access. More specifically, L2 learners may rely less on tonal cues than native listeners do at the lexical level. This would be consistent with Pelzl et al.'s (2019) finding that L2 learners are more likely to reject segmentally mismatched non-words than tonally mismatched non-words, while native listeners reject them at equal rates. This pattern of results indicates that, compared to native listeners, L2 learners allocate less weight to tone relative to segments.

In addition to the difficulty of drawing phonological information from acoustic input, the top-down influence of highly over-learned L1 automatic SPRs may also make L2



learners less likely than native speakers to use tonal cues during word recognition. In English, suprasegmental features are rarely lexically distinctive. Although some English words differ from each other only in stress, English listeners do not appear to use this suprasegmental information during lexical access, presumably because such cases are exceedingly rare (Cutler, 1986). In Mandarin, on the other hand, tone is a key component of lexical representation and is indispensable in word processing.

The question that arises, then, is whether long-term exposure and learning experience lead to increased use of tonal cues in an L2. Zou et al. (2017) addressed this question using a classification task with beginning and advanced Dutch learners of Mandarin. A group of Dutch speakers who were naïve listeners of Mandarin were also included as controls. Results showed that, like the naïve Dutch control group, beginning learners with 8–20 months of learning experience were less accurate at classifying stimuli based on tone alone than Mandarin native speakers. The advanced learners with 3–14 years of learning experience showed significantly higher accuracy than both the Dutch control group and the beginning learners, and did not differ significantly from the native Mandarin speakers. These findings indicate that the weight attributed to tonal cues increases at higher levels of proficiency. However, this study only investigated processing at a phonetic-phonological level in a task that did not involve lexical access, and therefore we still do not know whether advanced L2 learners use tonal and segmental cues to the same extent as native listeners during word recognition. It remains possible that, during a demanding and high-level cognitive task such as lexical access, attention to tonal cues remains limited even in advanced learners.

Experiment 2 explores this possibility with a visual-world eye-tracking experiment inspired by Malins and Joanisse (2010) that investigates how L1 and L2 speakers of Mandarin process tonal and segmental cues during real-time word recognition. Previous studies have reported that L2 listeners have more trouble than native listeners when it comes to utilizing all relevant phonological cues during word processing due to the difficulty that learners have with identifying phonemes in the L2 (Broersma, 2012; Broersma & Cutler, 2008, 2011; Qin, 2017). Evidence from recent studies shows that L2 learners are more likely than native speakers to activate multiple homophones or word

candidates that partially overlap with the target words (Broersma & Cutler, 2008, 2011; Dijkstra et al., 2000; Marian & Spivey, 2003; Weber & Cutler, 2004).

Mandarin words are composed of both lexical tone and segments. If listeners are not able to use tonal cues along with segmental cues, they might encounter more difficulty during word processing, because they could activate extra candidates that share the same segment (but not the same tone). Until now, however, no study has directly addressed this assumption in a time-sensitive experiment that approximates natural word processing in daily life. Experiment 2 is designed to investigate the differences between L2 learners and native Mandarin speakers in spoken word recognition and explore the relation between tone processing at the phonological and lexical levels.

### ***2.2.3 Tone in word learning***

In the two previous sections, I discussed how L2 learners process tone differently from native listeners at the phonological and lexical levels, which may account for the persistent difficulties learners encounter in acquiring tone. The next question one might ask is how we, as educators, can help L2 learners learn lexical tone in words more efficiently. Experiment 3 is designed to address this question by investigating the effectiveness of a popular teaching method—cue-focus training—in tonal word learning. In the remainder of this Chapter, I review literature on L2 tone learning.

#### **2.2.3.1 Factors influencing L2 tone learning**

Word learning or vocabulary acquisition is one of the earliest and most essential tasks that a person encounters in learning a language. As the building blocks of language, words involve both form and meaning. Previous research suggests that lack of vocabulary knowledge is one of the main reasons for the difficulties in L2 (Loewen, 2020; Loewen & Sato, 2017). One aspect of vocabulary knowledge is the link between form and meaning (Webb, 2020). In all languages, words are composed of consonants and vowels, while in some languages, lexical tone is also a critical component of lexical representation (Yip, 2002). As reviewed above, speakers of non-tonal languages (e.g., English) find it difficult

to learn and process words in a tonal language (e.g., Pelzl et al., 2019, 2020; Wong & Perrachione, 2007), which might be related to their knowledge of tone and the automatization of using tonal along with segmental information to access lexical meanings of words. A number of factors, including L1 experience, musicality, auditory ability, and learning mode, are all likely to contribute to success in learning tone.

### *L1 influence*

Previous studies have shown that the L1 prosodic system can affect how learners perceive tone information and which features they attend to. Unlike Mandarin-learning infants, adult learners already have a complete L1 system. In L2 learning, adults with fully developed cognitive abilities are likely to analyze an L2 in the same way they process their mother tongues. However, in the long run, the tendency of learners to map L2 sounds onto L1 phonological categories might slow down the development of mental representation of tone in the L2. As with other suprasegmentals (e.g., intonation), lexical tone is realized through F0, duration, and amplitude. Thus, adult naïve listeners are likely to assimilate unfamiliar Mandarin lexical tone to their native lexical tone or L1 intonation categories (e.g., Hao, 2014; So & Best, 2008). So and Best (2008) tested Australian English speakers who had neither learned Mandarin nor received formal music training in a forced categorization task. They found evidence of assimilation, in that Mandarin T1 was categorized as the “Flat pitch” category, T2 to the “Question” category, and T4 to the “Statement” category. T3, which has a contour that is less familiar to English listeners, was assigned to the “Uncertainty” and the “Question” categories at statistically equal rates.

Since all adult L2 learners have an intimate familiarity with their L1 prosodic system, its influence might be unavoidable in L2 tone learning. Thus, taking learners’ L1 experience into consideration may not only be important, but also necessary in predicting their success at learning tone in an L2.

### *Musicality and basic auditory ability*

Previous studies have suggested that music and tonal language experience are closely related and that music experience can positively influence the processing of lexical tones. Native speakers of non-tonal languages who happen to be musicians tend to perceive

lexical tone more accurately than non-musicians with similar L1 backgrounds. Alexander et al. (2005) tested the ability of English-speaking musicians and English-speaking non-musicians to identify and discriminate between the four standard lexical tones in Mandarin imposed on five different syllables. Results showed that musicians achieved significantly higher accuracy than non-musicians, suggesting that musical experience might help learners perceive tone. Music experience can also assist in the learning of words with lexical tone. In the artificial-word-learning experiments by Wong and Perrachione (2006), native English speakers were trained to use pitch patterns to identify novel vocabulary items. They found that learners' ability to perceive pitch patterns and their previous musical experience were both predictive of their success with learning tonal words. In another word-learning experiment done by Cooper and Wang (2013), L1-English musicians were found to achieve higher accuracy in learning words with Cantonese tones than L1-English non-musicians. Furthermore, L1-English musicians who had not received any explicit training in tone perception performed similarly to L1-English non-musicians who had been trained to identify Cantonese tones while learning new words.

However, an empirical study by Bowles et al. (2016) showed that the correlation between musicality and success with tone learning might be mediated by basic auditory ability. In their study, native English speakers with no previous tonal language experience completed a series of tasks measuring pitch ability, musicality, L2 aptitude, and general cognitive ability. They also completed a six-session Mandarin pseudo-word learning task. Results showed that pitch ability and musicality were stronger predictors for tonal word learning performance than L2 aptitude and general cognitive ability. However, compared to musicality, basic auditory ability was much more strongly correlated with learning outcomes. Bowles et al. (2016) also pointed out that those two abilities are likely to be highly correlated because musicians are more likely to have a natural sensitivity to pitch than members of the general population.

### 2.2.3.2 Teaching tone to L2 learners

While manipulation of the factors listed above is mostly beyond learners' control, teachers and educators are always seeking ways to improve learning outcomes by focusing on external factors that can be manipulated, such as different teaching and training methods. Applied linguistics and pedagogical studies have reported that a variety of teaching methods, such as visualization of tone contours (Liu et al., 2011), using color or number coding (Godfroid et al., 2017), music (Lin, 1985), hand gestures or other body movements (Tsai, 2011), and instruction focusing on pitch direction and pitch height (He et al., 2016) can all be effective in the perceptual learning of tone. Common to all of these methods is that they try to draw learners' attentional focus to tone as a minimally contrastive feature. This aligns with a number of different theoretical models of language learning. For instance, the Noticing Hypothesis (Schmidt, 1990) claims that "noticing is necessary for intake" (p. 141). If learners do not pay attention to the contrastive features of a cue, they cannot internalize the information and learn the cue. According to the ASP model (Strange, 2011), the perceptual salience of a cue is influenced by speakers' linguistic experience, but experimental manipulation can potentially reallocate attentional focus. Finally, the Competition Model (MacWhinney, 2005, 2012) supports cue-focus training more directly by arguing that presenting contrastive forms can increase the relative strength of a cue in acquisition. According to MacWhinney (2005), "[c]ue availability is defined as the presence of the cue in some contrastive form" (p. 53), and it can be used as a predictor of the relative strength of a cue (MacWhinney, 2012, p. 222).

Focusing learners' attention on the contrastiveness of a cue such as tone thus not only seems intuitively helpful, but is also broadly supported by existing theoretical models. It is important to note, however, that very few studies have directly tested this assumption. Nevertheless, many researchers in applied linguistics accept it as a premise and focus on the examination of different types of cue-focus training methods in the learning of tone. For example, Lin (1985) created drill materials consisting of Chinese words written on a musical scale and using pitch levels to assist in the learning of tone contours. Post-test results showed improvement in tone identification.

In a study by Liu et al. (2011), L1-English first-year learners of Mandarin were trained to learn Mandarin tone on syllables in three different learning conditions—

contour + pinyin, number + pinyin, and contour only—in a classroom setting. All three of these learning conditions were versions of cue-focus training, and the study was based on the assumption that such training could direct learners' attention to the critical features of tone and thus lead to better learning. Learning outcomes were measured in terms of decreases in error rates on two identical tone judgment tasks conducted as pre- and post-tests. Results showed that the contour + pinyin condition was associated with greater error reduction than the other two conditions.

Based on earlier work finding that English native speakers are more sensitive to pitch height than to pitch contour (e.g., Gandour, 1983), He et al. (2016) compared the effectiveness of pitch-height-focused instruction and pitch-direction-focused instruction in a classroom setting. Although Mandarin native listeners are known to have a preference for using pitch direction for tone perception (e.g., Gandour, 1983), the English speakers in He et al.'s study learned tone more effectively with pitch-height-focused instruction than with pitch direction-focused instruction.

Findings from those three studies are of potential pedagogical relevance, but they do not constitute direct support for the assumption that cue-focus training is effective in principle in the learning of tone since all conditions focused directly on contrasts between different tones. Furthermore, additional confounds between groups were difficult to control since the studies were conducted in already existing classrooms.

Based on Liu et al.'s (2011) positive results with dual visual representation, Godfroid et al. (2017) compared the effectiveness of five multimodal methods (three single-cue methods: number, color and pitch contour; two dual-cue methods: color + number, color + pitch contour) of tone contrastive training for Mandarin tone perception in a more controlled experimental setting. Results from a pre-test as well as both immediate and delayed post-tests showed that all training methods were effective, with instruction involving pitch contours and numbers being more beneficial than instruction involving colors. Also, dual-cue methods were no more effective than single cue-methods. These findings align well with those of Liu et al. (2011) and Lin (1985) in that they suggest that training can improve learners' perception of tone. However, none of these studies have tested the assumption that cue-focus training is effective in the learning of tone in comparison to some baseline condition where no contrastive cue(s) is focused during the

training. Moreover, the training studies mentioned above only examined the learning effect at a phonological level.

Though previous studies have provided ample evidence that the accuracy of tone identification by listeners with non-tonal language backgrounds can be greatly improved by even short-term training (Godfroid et al., 2017; Liu et al., 2011; Wang, 2013; Wang et al., 1999; Wong & Perrachione, 2007), L2 speakers' long-term difficulty with tone processing appears to lie at the lexical level (Pelzl et al., 2019, 2020; Qin, 2017). In Pelzl et al.'s (2019) study, the accuracy of tone identification by English-speaking learners of Mandarin was similar to that of native listeners, while learners were less successful than native listeners at rejecting non-words in a lexical decision task when the non-words and words differed only by tone, indicating a “disconnect between L2 abilities to categorize tones as phonetic objects and abilities to utilize those categories as lexical cues” (p. 69). As a lexical cue, the most important and fundamental function of tone that learners need to acquire is how to use tonal cues along with segmental cues to recognize words. As suggested by the Competition Model (MacWhinney, 2005, 2012), presenting contrastive forms of the tonal cue (e.g., /pa/-rising vs. /pa/-falling) might increase its relative strength during acquisition, which is the implicit assumption underlying cue-focus training. If this assumption is correct, listeners trained in a cue-focus condition should outperform those trained in a non-cue-focus condition in learning words with tone. The goal of Experiment 3 is to test this prediction in an artificial word learning task in a controlled laboratory setting.

### 2.2.3.3 Artificial word learning studies

Vocabulary learning is a complicated process, involving many known and unknown confounding factors. As discussed above, individual differences related to L1 language experience, musicality, and basic auditory pitch ability can all influence learning outcomes. Other factors such as training materials, classroom settings, and teaching styles might also affect the final result. To better control the influence of confounding factors, researchers use the artificial language learning (ALL) paradigm as an important tool for studying the principles of languages and language learning in an experimental setting

(Ettlenger et al., 2016; Hayakawa et al., 2020). In the remainder of this section, I will briefly review the benefits of the ALL paradigm in language learning research along with some relevant studies that have used this paradigm in the context of learning lexical tone.

The artificial language learning (ALL) paradigm refers to the experimental paradigm where participants learn language-like stimuli created for a specific research purpose in a laboratory setting and are then tested on what they have learned (Ettlenger et al., 2016). The words used as ALL stimuli are usually called pseudowords (Poltrock et al., 2018; Wong & Perrachione, 2007) or novel words (Quam & Creel, 2017). The ALL paradigm was first used by Esper (1925) to examine biases in word learning. In his study, participants were presented with pairs of words accompanied by pictures of different-colored shapes. After training, participants were asked to name the pictures presented to them. The artificial words were bi-morphemic with the first morpheme representing a color and the second one representing a shape. Learning outcomes were measured in terms of error rates in the production of the first and second morphemes. Esper found that participants often re-segmented the stimuli into two consonant-vowel-consonant morphemes, which he saw as a reflection of a learning bias against morphemes with complex structure.

Recent studies on tone learning in words use the ALL paradigm to explore whether tone can be learned by speakers of non-tonal languages and which factors influence learning outcomes (Cooper & Wang, 2013; Poltrock et al., 2018; Quam & Creel, 2017; Wiener et al., 2020; Wong & Perrachione, 2007). In Wong and Perrachione's (2007) study, for example, native English speakers with no previous experience in tonal languages were trained to learn English pseudowords with three pitch signatures resembling three Mandarin lexical tones: T1, T2, and T4. The study was designed to investigate whether native English speakers were able to use pitch to identify words and to determine which factors (e.g., musicality or basic pitch ability) influence the learning of tones. Instead of using actual Mandarin words, Wong and Perrachione used pseudowords with segments that would be familiar to native English speakers. By using segments familiar to participants, the researchers increased the learnability of the words and encouraged participants to focus on pitch information rather than segmental information. For similar reasons, T3—the most difficult tone in Mandarin—was excluded



from testing. By using pseudowords that do not exist in any language, researchers could eliminate the confound of participants' prior exposure to the test items. The results showed that native English speakers were able to learn words with tones and that individual learning success was correlated with learners' basic pitch ability. This outcome suggested that there is phonetic-phonological-lexical continuity in word learning by non-native speakers, a finding which has been corroborated by Cooper and Wang (2013) in their Cantonese word learning project. Artificial languages have also been used to test the efficacy of different training methods. For example, in Wiener et al. (2020)'s study, native Mandarin speakers and L2 learners were trained to learn a Mandarin-like artificial language to examine the benefits of explicit instruction and high variability phonetic training (HVPT). Results showed that HVPT interacted with explicit instruction and improved learners' production.

Besides providing better control of confounding factors in an experiment, another benefit of using the ALL paradigm is that learning outcomes can be quickly tested after the training phase, which is difficult to accomplish in studies conducted in actual classrooms. For instance, Poltrock et al. (2018) used the ALL paradigm to study how Cantonese-, Mandarin-, and French-speaking adults learn pairs of Cantonese-based pseudowords that differ from each other in terms of consonants, vowels and tones, and their experiment had a testing phase that immediately followed the training phase. In each training trial, participants learned a pair of words and their picture associations. Though all pairs of words were contrastive in Cantonese, some of them were not contrastive in Mandarin or French. In each test trial, two pictures were presented on the screen and participants were instructed to look at the picture corresponding to the word they heard, while eye movements were recorded as the measurement for analysis. Results showed that all three groups performed at above chance levels in learning pseudowords. Cantonese speakers outperformed Mandarin and French speakers on all three contrasts, and French speakers performed the worst among the three groups on tones. This finding suggested that L1 experience with lexical tone positively influences the learning of novel words with tones.

Besides eye movements, mouse clicks are also commonly used to measure learning outcomes in ALL experiments designed for adult participants. Quam and Creel (2017),

for example, examined how Mandarin-English bilinguals process tone differently in a Mandarin context vs. an English context by using those two measurements in the test phase. In the first experiment, Mandarin-English bilinguals and English monolinguals were trained to learn pairs of novel words with Mandarin-compatible segments that differed from each other either in their vowels or in their tones, and both groups were tested in an alternative choice test of those tone and vowel minimal pairs. To ensure that participants would be able to learn a sufficient number of words in one session, Quam and Creel (2017) carefully controlled the phonological properties of the words they used in the learning materials by making the segments either English-like or Mandarin like. Results showed that Mandarin-English bilinguals outperformed English monolinguals, especially with minimal tone pairs, and bilinguals' accuracy scores were significantly correlated with Mandarin dominance. In the second experiment, tones were added to English-like segments, and the results showed no effect of Mandarin dominance. In all, Quam and Creel's (2017) study showed that within-word language context influenced tone processing by Mandarin-English bilinguals.

These studies demonstrate that the ALL paradigm is a useful tool for controlling the properties of the experimental materials and allowing training and testing to occur in a single session when resources are limited. In Experiment 3, I used the ALL paradigm to investigate the effectiveness of cue-focus training, which has become a popular teaching method in language classrooms despite the fact that its utility has never been tested under experimental conditions.

### **2.3 Summary**

From the literature review above, we see more research is required if we want to achieve a better understanding of how L2 learners perceive, process, and learn Mandarin lexical tones, which are an important but challenging feature of a widely used and widely learned language. The goal of this dissertation, is to bridge the gaps in the previous literature by investigating L2 learners' processing of tone at the phonological and lexical levels and exploring the effectiveness of cue-focus training in the learning of tonal words by native English speakers. This dissertation is composed of experiments designed to test

three broad research questions related to the L2 acquisition of lexical tone at the levels of *(1) speech perception, (2) lexical processing, and (3) word learning*:

- (1) How does language experience (native/non-native/no experience with a tonal language) affect how listeners *perceive* variation in pitch/tone in isolated syllables? (Chapter 3: Experiment 1)
- (2) How do L2 learners *process* lexical tone relative to segment in real-time spoken Mandarin word recognition? (Chapter 4: Experiment 2)
- (3) How do speakers with no tonal language experience *learn* novel words with tones under different training conditions? (Chapter 5: Experiment 3)

## **Chapter 3. Experiment 1: Categorical perception of lexical tone by speakers with different language backgrounds<sup>1</sup>**

### **3.1 Research questions**

As mentioned in Chapter 2, previous studies on categorical perception (CP) of lexical tone have mostly focused on the perception of a small number of tone pairs (e.g., T1-T4 and T2-T3 in Shen & Froud, 2016) by naïve listeners (e.g., Hallé et al., 2004). Little is known about how L2 learners perceive all six possible Mandarin tone pairs compared to native speakers and naïve listeners or how learners' L2 proficiency influences their perception of tone. Experiment 1 was designed to advance our understanding of L2 tone perception by testing English-speaking learners of Mandarin with various proficiency levels in both identification and discrimination tasks with all six Mandarin tone pairs. Mandarin native speakers and English native speakers with no experience of any tonal languages (naïve listeners) were also tested in the same tasks for comparison. More specifically, this study was designed to explore the following three questions:

- (1) To what extent do L2 learners of Mandarin perceive lexical tone categorically in comparison to native listeners and naïve listeners?
- (2) How does L2 proficiency affect tone identification and discrimination within the L2 group?
- (3) How do different groups of listeners perceive different lexical tone pairs?

Based on previous studies showing that training can improve tone perception in speakers of non-tonal languages (Lu et al., 2015; Shen & Froud, 2016, 2018; Wang et al., 1999; Wang et al., 2004), I predicted that L2 learners would have a pattern of performance that is between those of the native and naïve listener groups, and that they would show stronger evidence of CP than the naïve listeners on both tasks. Since proficiency is a measurement of overall language ability, I also expected that there would

---

<sup>1</sup> A preliminary short report on L1 data of this study appeared in the TAL proceedings: Ling, W. & Schafer, A. J. (2016). Tone pair similarity and the perception of Mandarin tones by Mandarin and English listeners. *Proceedings of the 5th International Tonal Aspects of Language (TAL)*, Buffalo, NY.

be a positive correlation between proficiency scores and degree of CP for the L2 learners. Lastly, since not all tone pairs have equally distinct pitch contours (e.g. T2 and T3 have similar pitch contours while T1 and T3 have very different contours), the three groups of listeners might perceive different tone pairs differently.

## **3.2 Methods**

### ***3.2.1 Participants***

Thirty-one native Mandarin speakers from the community in and around East China Normal University in Shanghai, China took part in this experiment. None of them had any professional music experience. They had all spent at least some time learning English as an L2, but they also reported that Mandarin was their dominant language used in daily life. Thirty-two native English speakers studying at the University of Hawai'i at Mānoa, USA participated in this experiment. They all verified that they had no experience with either learning tonal languages or practicing music professionally. Twenty-seven native English-speaking L2 learners of Mandarin were also recruited from the two universities mentioned above. As with the participants in the other two groups, none of them reported having any professional music experience. All the L2 participants had started learning Mandarin after the age of 15 years (range: 15–40), and none of them reported having any exposure to Chinese earlier in life. L2 proficiency was assessed through a listening proficiency test adapted from the Chinese Standard Exam (Hanyu Shuiping Kaoshi or HSK; Confucius Institute in Atlanta, 2017), which is described in more detail below, in addition to participants' self-ratings of their Chinese speaking (CS), listening (CL) and reading (CR) skills on 5 point scales (1 = poor, 5 = good). I then calculated a composite proficiency score because listening and speaking abilities are closely related to each other (Demir, 2017) and previous studies found strong evidence of correlation between listening and reading abilities, both of which are receptive skills and require decoding of phonological cues (Devine, 1968; Hagtvet, 2003; Hastuti & Kalim, 2019; Song et al., 2016). Except for one L2 learner who felt uncomfortable taking the listening proficiency test, all L2 learners completed the test and scored from 0.2 to 1. Among all 27 L2 learners, 6 were tested in China and 21 were tested in the US. Since the six tested in China had

similar self-rating scores (CS:  $M = 2.33$ ,  $SD = 1.21$ ; CL:  $M = 2.67$ ,  $SD = 1.37$ ; CR:  $M = 2.50$ ,  $SD = 1.22$ ) and listening test scores ( $M = 0.86$ ,  $SD = 0.19$ ) to those tested in the US (CS:  $M = 2.29$ ,  $SD = 1.42$ ; CL:  $M = 2.10$ ,  $SD = 1.18$ ; CR:  $M = 2.24$ ,  $SD = 1.37$ ; listening test:  $M = 0.72$ ,  $SD = 0.29$ ), I combined their data for all further analyses (see Table 2).

**Table 2.**

*Participants' Language Background Information*

	Age (years)	Gender (male)	AO (years) Birth	CS	CL	CR	Listening test
Native Mandarin speakers (n = 31)	18–40	4	NA	NA	NA	NA	NA
Native English speakers (n = 32)	18–50	11	NA	NA	NA	NA	NA
L2 learners of Mandarin (n = 27)	18–40	20	21.22 (5.76)	2.30 (1.35)	2.22 (1.22)	2.30 (1.32)	0.78 (0.24)

*Note.* values are means (standard deviations shown in parentheses), except for age and gender. AO = age onset of learning Mandarin; CS = Chinese speaking; CL = Chinese listening; CR = Chinese reading.

All participants reported having normal or adjusted-to-normal auditory and visual ability. Participants were rewarded for their time with course credit or a small amount of money. Two additional participants were tested, but their data were excluded from analysis because of their early exposure to tonal languages (Vietnamese: 1; Cantonese: 1).

### 3.2.2 Materials

Two syllables, /pa/ and /pi/, were used to create experimental materials for both the identification and discrimination tasks; /kwo/ was used for the practice stimuli. All combinations of the /pa/ and /pi/ segments with the four Mandarin tones correspond to

existing words in Mandarin, as illustrated in Table 3. Since it is very common to have several homophones for one syllable in Mandarin, I decided to follow Hallé et al. (2004) in reporting only the frequencies of the most common homophone for each syllable. Table 3 also reports the log frequencies of the most frequent homophone for each syllable + tone pair based on data from Google (Huang, 2005). Chi-square tests showed no significant differences across the four tones ( $X^2 = 0.05$ , *NS*). After data collection, a technical error was discovered in synthesizing stimuli with some tokens of /pa/.<sup>2</sup> For this reason, only the data from /pi/ items were included in the analyses.

**Table 3**

*Frequencies of the Morphemes that are Homophonic to the Endpoint Stimuli*

Syllable	Tone			
	Tone1	Tone2	Tone3	Tone4
/pa/	6.95 (八 “eight”)	6.24 (拔 “pull”)	7.14 (把 “take”)	6.33 (爸 “daddy”)
/pi/	6.45 (逼 “oppress”)	6.48 (鼻 “nose”)	7.66 (比 “compare”)	6.90 (必 “must”)

*Note.* For each endpoint, the most frequent morpheme is shown in parentheses, along with an English gloss. The word frequency is in log frequency. The Chinese characters and their approximate English glosses are in parentheses.

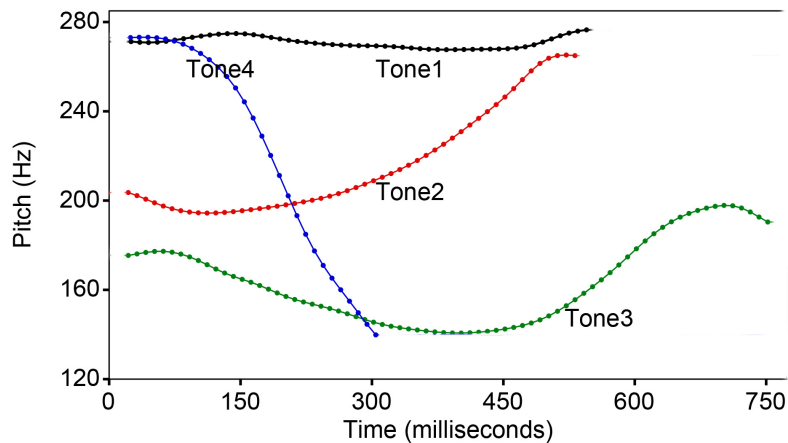
To avoid effects of co-articulation and tone sandhi, the sound stimuli were constructed from isolated syllables produced by a female native Mandarin speaker. The stimuli were recorded at 44.1kHz as mono sound on a laptop using Praat (Boersma & Weenink, 2013). The speaker read each stimulus three times at the same speed and in a

<sup>2</sup> Some synthesized tokens of /pa/ were found missing after data collection. To ensure the completeness of tokens along the synthesized spectrum, only perception data of /pi/ tokens are reported here.

clear voice. One token of each was selected based on loudness and sound quality. Figure 3 shows the pitch contour of /pi/ with all four tones.

**Figure 3**

*Pitch Contours of /pi/ With Four Tones for the Experimental Stimuli*



The eight selected tokens of /pi/ and /pa/ were used to construct six tone pairs (T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, and T3-T4) for each syllable. Six 9-step tone continua were generated by equalizing duration within the continuum and linearly changing the pitch and intensity between the two endpoints (see Figure 4) using the PSOLA method (Moulines & Laroche, 1995) in Praat. The F0, intensity and duration for each experimental token were modeled using the following functions:

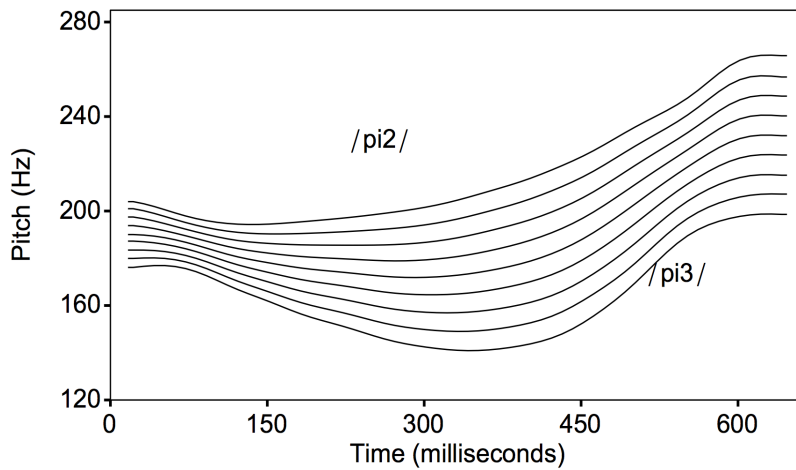
$$\begin{aligned} \text{Pitch:} \quad & P_i = (1 - i/8) * P_{\text{SoundA}} + (i/8) * P_{\text{SoundB}} \text{ (Hz)} \\ \text{Intensity:} \quad & \text{rms}_i = (1 - i/8) * P_{\text{SoundA}} + (i/8) * P_{\text{SoundB}} \text{ (dB)} \\ \text{Duration:} \quad & L_i = (L_{\text{SoundA}} + L_{\text{SoundB}}) / 2 \text{ (ms)} \\ & (i = \text{step } 0, 1, \dots, 8 ; P = \text{pitch, rms} = \text{intensity, } L = \text{duration}) \end{aligned}$$

Six 4-step tone continua for /kwo/ were created in a similar manner for use in the practice trials.



**Figure 4**

*Example of a Synthesized T2-T3 Continuum*



### **3.2.3 Experimental tasks**

#### **3.2.3.1 Identification task**

Stimuli were blocked by tone pair to reduce listeners' memory load and avoid uninformative answers (e.g., the middle step of the T2-T4 continuum has a flat pitch, which is similar to T1). Each block started with information about the upcoming tone pair and two exposure trials with labeled endpoints. Blocks were counterbalanced across participants. All stimuli were presented three times in random order within each block, resulting in a total of 324 trials (2 segments \* 6 tone pairs \* 9 steps \* 3 presentations). An initial practice block was used to familiarize participants with the task using the T1-T2 continuum with the syllable /kwo/. Participants were asked to classify each token by pressing the appropriate tone-number key on the keyboard (e.g., 1 or 2 in block T1-T2), followed by the space bar to initiate the next trial at their own pace. Participants were encouraged to guess if unsure.

#### **3.2.3.2 Discrimination task**

Discrimination was tested in an AXB task using steps two intervals apart, e.g. step 1 vs. step 3. In each trial, participants were asked to indicate whether token X matched

token A or token B by pressing the “f” and “j” keys, respectively. As in the identification task, all participants started with a practice block followed by six testing blocks, divided by tone pair. There were a total of 336 experimental trials (2 syllables \* 6 tone pairs × 7 step-pairs × 4 AXB combinations: AAB, ABB, BBA, BAA). Following common practice, the inter-stimulus interval (ISI) was set to 500 milliseconds (van Hessen & Schouten, 1992; Xu et al., 2006). The keyboard was deactivated while the sound files played, forcing participants to listen to the complete AXB recording before responding. This task was self-paced and took about 35 minutes to finish.

### 3.2.3.3 Listening proficiency test

In order to explore the possible correlation between L2 learners’ Mandarin proficiency and their performance in the two experimental tasks, proficiency information for each participant was collected through self-ratings and a simple listening proficiency test (see Appendix A1 for the version with Chinese instructions and Appendix A2 for the version with English instructions). In the online questionnaire, participants were asked to rate their reading, listening and speaking ability for both Mandarin and English on a 5-point scale. The Mandarin listening test was created from the first 20 items of the listening section of the HSK second level test (Confucius Institute in Atlanta, 2017), a boundary test used to differentiate beginners from intermediate learners. There were two parts of the listening test, each of which had 10 items. Part 1 involved listening to short sentences. For each item, participants needed to decide whether the sentence they heard is compatible with the picture or not. Part 2 involved listening to 10 short conversations. Participants were required to put a series of pictures in order to match the conversations they heard. Native Mandarin speakers were given the version of the test with instructions in Mandarin and L2 learners were given the version with instructions in English. A correct answer for each item was given 5% and the maximum score on the task was 100% (5 \* 20).

### 3.2.3.4 General procedure

First, all participants completed a questionnaire about their basic demographic information, language background, music experience and self-rated Mandarin skills. Then they took the identification and discrimination tasks, the order of which was counterbalanced across participants. There was a short break between these two tasks. L2 learners also completed a Mandarin listening proficiency test at the end of the experimental session.

### 3.2.4 Data analysis

To investigate the effect of language learning experience and tone pair similarity on participants' tone perception, I obtained measurement of tone perception for each participant based on two characteristics: slope of identification and discrimination accuracy. Steeper identification slopes indicate higher degrees of CP.

#### 3.2.4.1 Identification measurement

Responses in the identification task were coded as choice of the first tone (1) versus the second tone (0) in any given block. For example, in the T1-T2 block, T1 was coded as 1, and T2 as 0. For each group and tone pair, a generalized mixed-effects logistic regression model was used to obtain identification slopes (van Hessa & Schouten, 1992; Xu et al., 2006):

$$glmer(\text{Response} \sim \text{Step} + (\text{Step} | \text{Participant}), \text{family} = \text{binomial}(\text{link} = \text{"logit"}))$$

This produced model-based intercepts and regression coefficients (in log odds) for each tone pair and participant. The regression coefficients represented the value of the identification slopes.

#### 3.2.4.2 Discrimination measurement

Responses in the discrimination task were coded as accurate (1) or inaccurate (0). Mean accuracy scores for each participant and tone pair were calculated for further analysis.

### 3.3 Results

Data from six participants were excluded because they were biased toward selecting the first sound (Sound A) in the AXB discrimination task (1 native listener, 1 L2 learner and 1 naïve listener) or low overall accuracy in the discrimination task (1 L2 learner and 2 naïve listeners). The exclusion criterion in both cases was that values were more than 2.5 standard deviations away from the respective group mean. Data from the remaining 30 native, 25 L2 and 29 naïve listeners were included in the following analyses.

#### *3.3.1 Identification task*

Figure 5 shows the identification curves for each group collapsed over tone pairs. Visual inspection indicates that the L2 learners' identification curves were steeper than those of the naïve English listeners, but shallower than those of the native listeners.

**Figure 5**

*Identification Curves Pooled across Participants and Tone Pairs*

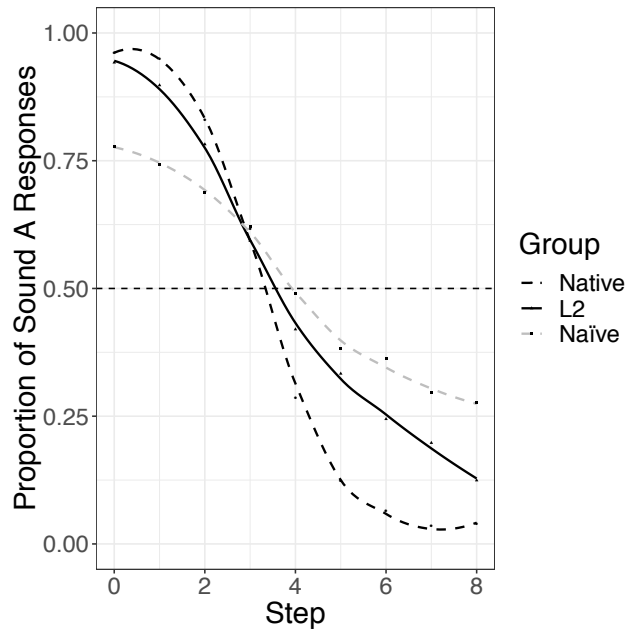


Table 4 shows the value of identification slopes by group for each tone pair. Collapsed by tone pair, native listeners had the steepest identification curves ( $M = -1.69$ ,  $SD = 0.35$ ) and naïve listeners had the shallowest identification curves ( $M = -0.48$ ,  $SD = 0.21$ ), while L2 learners were in-between ( $M = -0.98$ ,  $SD = 0.38$ ). This pattern remained true for each individual tone pair (see Table 4).

**Table 4**

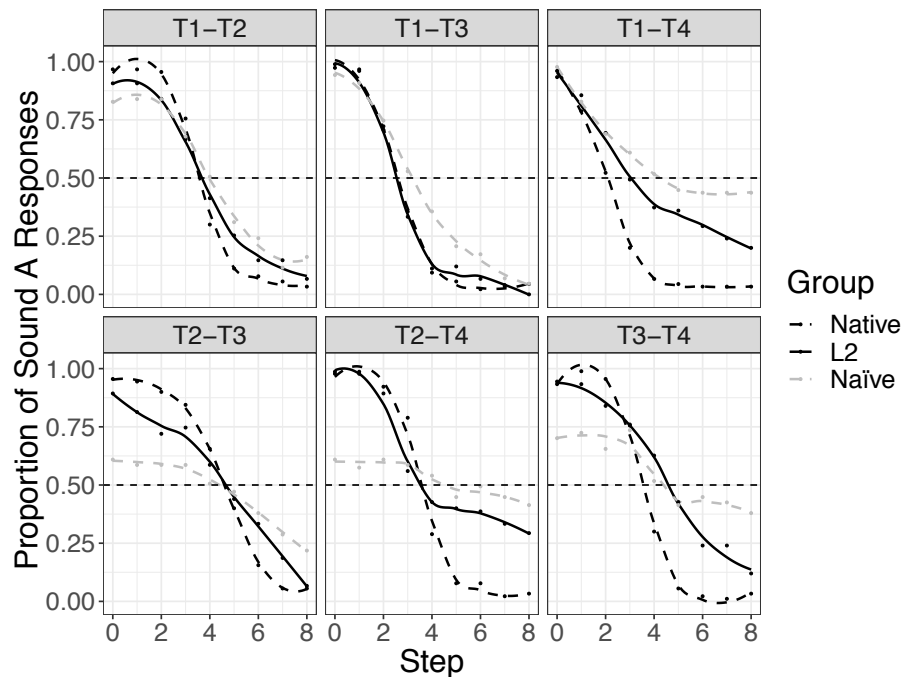
*Mean Values of Identification Slopes for Mandarin Native Listeners, Naïve Listeners and L2 Learners for Each Tone Pair*

Tone pair	Native listeners (N = 30)	L2 learners (N= 25)	Naïve listeners (N = 29)
T1-T2	-1.74	-1.15	-0.69
T1-T3	-1.81	-1.78	-1.01
T1-T4	-1.32	-0.77	-0.37
T2-T3	-1.43	-0.64	-0.33
T2-T4	-1.78	-0.72	-0.18
T3-T4	-2.02	-0.82	-0.28
Collapsed by tone pairs	-1.69	-0.98	-0.48

Figure 6 shows the identification curves for each group and tone pair. For most tone pairs, the L2 learners' identification curves were steeper than those of the naïve English listeners but shallower than those of the native listeners, except for T1-T3, where the L2 learners' identification curve was similar to that of the native listeners but steeper than that of the naïve listeners. For both groups of non-native listeners, especially the L2 learners, the identification curves for T1-T2, T1-T3 and T1-T4 were steeper than those for T2-T3, T2-T4 and T3-T4, suggesting that tone pairs involving the level tone (T1) were more likely to be categorically perceived than tone pairs made of two contour tones.

**Figure 6**

*Identification Curves Pooled across Participants for Each Tone Pair*



*Note.* The lines represent the proportion of Sound A responses at each step (0–8).

In order to compare the degree of CP between groups statistically, I fitted the data to mixed-effect linear regression models. I used a forward-fitting strategy to construct the models. I started with a simple model that included the identification slopes as a dependent variable and the intercepts for participants as a random effect; then I incrementally added group, tone pair and their interaction to the model as fixed effects. Maximal random effect structures justified by the design were also attempted, and reduced if convergence problems arose (Barr et al., 2013). Models were compared using the *anova()* function in R in order to determine the best fitting model. The factor group was simple-coded to examine a possible main effect, and L2 learners were set as the reference level. I also simple-coded tone pair (6 levels) to detect a potential main effect; T1-T2 served as the reference level. All analyses were carried out in R (R Core Team, 2014) and R studio using the *lmerTest* package (Kuznetsova et al., 2017) based on *lme4* package (Bates et al., 2015). The final model contained group, tone pair and their interaction as fixed effects, as well as intercepts for participants as a random effect.

**Table 5***Results of Linear Mixed-effects Model for Identification Slopes*

Formula (lmer): Slope~Group*TonePair +(1 Participant)				
	Fixed Effects			
	b	SE	t	p
Intercept	-1.04	0.03	-30.48	< .001
Native	-0.70	0.09	-8.29	< .001
Naïve	0.50	0.09	5.89	< .001
T1-T3	0.34	0.06	5.84	< .001
T1-T4	0.71	0.06	12.21	< .001
T2-T3	0.73	0.06	12.56	< .001
T2-T4	0.64	0.06	10.96	< .001
T3-T4	0.49	0.06	8.45	< .001
Native: T1-T3	-0.56	0.14	-3.88	< .001
Naïve: T1-T3	-0.32	0.14	-2.18	.03
Native: T1-T4	-0.51	0.14	-3.56	< .001
Naïve: T1-T4	-0.37	0.14	-2.53	.01
Native: T2-T3	-0.75	0.14	-5.24	< .001
Naïve: T2-T3	-0.45	0.14	-3.14	.002
Native: T2-T4	-1.02	0.14	-7.11	< .001
Naïve: T2-T4	-0.22	0.14	-1.53	.13
Native: T3-T4	-1.17	0.14	-8.14	< .001
Naïve: T3-T4	-0.23	0.14	-1.62	.11

The results (Table 5) showed that the identification slope for the L2 group collapsed over tone pair (main effect) was significantly steeper than the one for the naïve listeners ( $b = 0.50$ ,  $t = 5.89$   $p < .001$ ) but shallower than the one for the native listeners ( $b = -0.70$ ,  $t = -8.29$ ,  $p < .001$ ), consistent with the visual inspection of identification curves in Figure 6. The model output also showed that the identification slope collapsed over group for T1-T2 was significantly different from the ones for the other five tone pairs ( $|b| > .34$ ,



$p < .001$ ) and there were eight significant interaction effects, indicating that the main effect of tone pair was qualified by an interaction with group. To examine RQ3 about how different groups perceive different tone pairs with various degrees of pitch similarity, I therefore conducted separate analyses within each group (*lmer(Slope ~TonePair+(1|Participant))*). Tone pair was treatment-coded and I reran the model by changing the reference level to compare each tone pair to the other five. Table 6 presents the model output. The ranking for the steepness of the identification slopes for the tone-pair continua was  $T3-T4 = T1-T3 = T2-T4 = T1-T2 > T2-T3 = T1-T4$  for the native listeners,  $T1-T3 > T1-T2 > T3-T4 = T1-T4 = T2-T4 = T2-T3$  for the L2 learners, and  $T1-T3 > T1-T2 > T1-T4 > T2-T3 = T3-T4 = T2-T4$  for the naïve listeners. Basically, all three groups tended to perceive T1-T3 and T1-T2 more categorically than T1-T4 and T2-T3.

**Table 6***Results of Comparison of Identification Slopes by Tone Pair within Each Group*

	Native		L2		Naïve
T3-T4 vs. T1-T3	-2.02 vs. -1.81 ( <i>b</i> = 0.21, <i>p</i> = .08)	T1-T3 vs. T1-T2	-1.78 vs. -1.15 ( <i>b</i> = 0.63, <i>p</i> < .001)	T1-T3 vs. T1-T2	-1.01 vs. -0.69 ( <i>b</i> = 0.31, <i>p</i> < .001)
T1-T3 vs. T2-T4	-1.81 vs. -1.78 ( <i>b</i> = 0.03, <i>p</i> = .80)	T1-T2 vs. T3-T4	-1.15 vs. -0.82 ( <i>b</i> = 0.33, <i>p</i> < .001)	T1-T2 vs. T1-T4	-0.69 vs. -0.37 ( <i>b</i> = 0.32, <i>p</i> < .001)
T2-T4 vs. T1-T2	-1.78 vs. -1.74 ( <i>b</i> = 0.04, <i>p</i> = .71)	T3-T4 vs. T1-T4	-0.82 vs. -0.77 ( <i>b</i> = 0.04, <i>p</i> = .63)	T1-T4 vs. T2-T3	-0.37 vs. -0.33 ( <i>b</i> = 0.09, <i>p</i> = .29)
T1-T2 vs. T2-T3	-1.74 vs. -1.43 ( <i>b</i> = 0.31, <i>p</i> = .01)	T1-T4 vs. T2-T4	-0.77 vs. -0.72 ( <i>b</i> = 0.05, <i>p</i> = .58)	T2-T3 vs. T3-T4	-0.33 vs. -0.28 ( <i>b</i> = 0.05, <i>p</i> = .58)
T2-T3 vs. T1-T4	-1.43 vs. -1.32 ( <i>b</i> = 0.11, <i>p</i> = .35)	T2-T4 vs. T2-T3	-0.72 vs. -0.64 ( <i>b</i> = 0.08, <i>p</i> = .37)	T3-T4 vs. T2-T4	-0.28 vs. -0.18 ( <i>b</i> = 0.10, <i>p</i> = .22)

*Note.* Data show means of different tone pairs within each group, as well as the coefficient (*b*-value) and its significance (*p*-value) in each comparison.

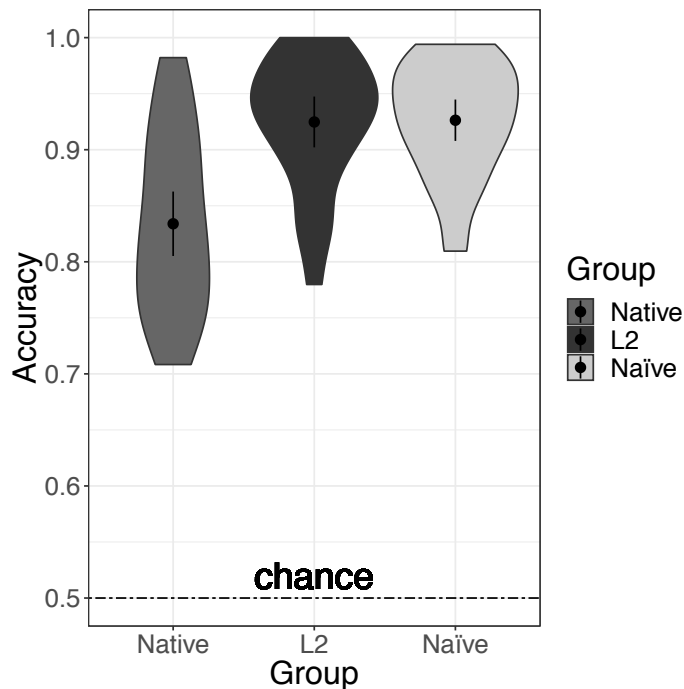
In sum, the native listeners showed the most CP in the identification task. The naïve listeners showed the least evidence of CP and the L2 learners consistently patterned between the native and the naïve listener groups. This overall pattern is suggestive of developing mental representations of tone categories among the L2 learners. This will be further investigated below by examining the contribution of L2 proficiency as a predictor in the model. With regard to differences between different tone-pair continua, all three groups tended to perceive tone pairs with different contours (T1-T3) or different starting points (T1-T2) more categorically than tone pairs with similar contours (T2-T3) or with similar starting points (T1-T4).

### 3.3.2 Discrimination task

Overall, the L2 learners ( $M = 0.92$ ,  $SD = 0.06$ ) had rates of discrimination accuracy that were very similar to those of the naïve listeners ( $M = 0.93$ ,  $SD = 0.06$ ), and both the L2 learners and the naïve listeners were more accurate than the native listeners ( $M = 0.83$ ,  $SD = 0.08$ ). Figure 7 shows violin plots of the accuracy data collapsed over tone pair. The L2 learners and the naïve listeners had similar accuracy distributions, while the native listeners had a wider distribution of accuracy scores.

**Figure 7**

*Accuracy in the Discrimination Task by Group*



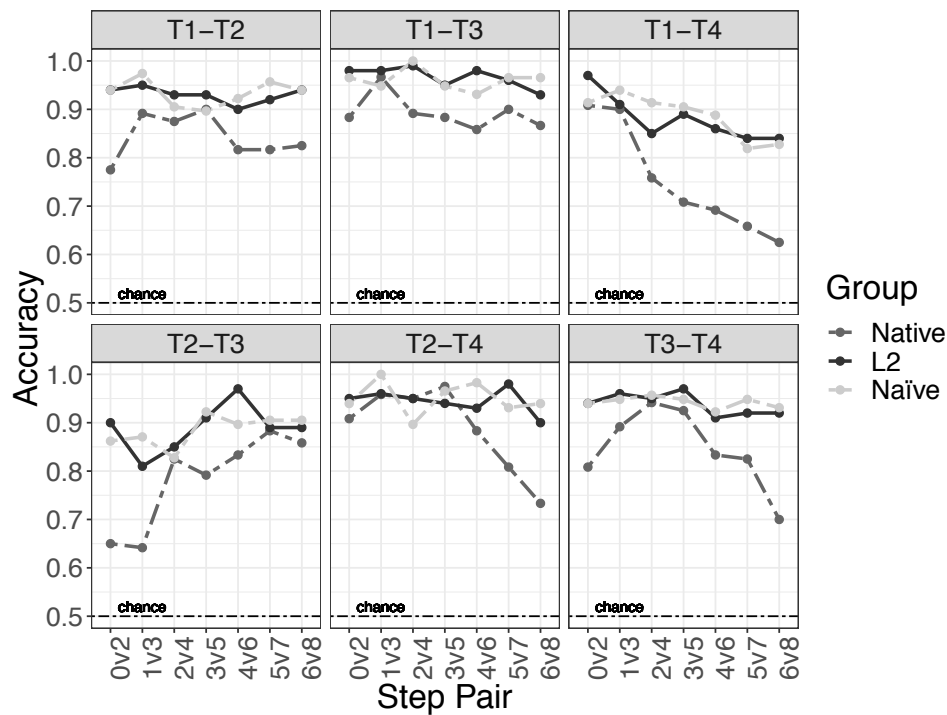
*Note.* Black dots are the group means and error bars show the 95% confidence intervals for each group. The width of the violin plot shows the density of data points.

Figure 8 shows the discrimination accuracy data by group and tone pair. The data from the L2 learners patterned more closely with the data from the naïve listeners than with the data from the native listeners across all tone pairs. The native listeners also had lower accuracy at most step-pairs across the continua than the other two groups; their

accuracy scores were only comparable to those of the L2 learners and the naïve listeners for certain step-pairs that might cross categorical boundaries. This suggested that the native listeners had decreased sensitivity to within-category differences, which contributed to their overall lower accuracy relative to the L2 learners and the naïve listeners. The native listeners were more sensitive to between-category difference than within-category difference.

**Figure 8**

*Discrimination Curves Pooled across Participants for Each Tone Pair*



For statistical analysis, I again began with a simple model with discrimination accuracy as the dependent variable and added predictors incrementally. The predictor coding for the final model was the same as for the identification analysis. The output (see Table 7) showed that the L2 learners were significantly more accurate than the native listeners overall ( $b = -0.09, p < .001$ ); however, the difference between the accuracy scores for the L2 learners and the naïve listeners was not statistically significant ( $b = 0.002, p = .93$ ). The accuracy rates collapsed over groups for T1-T2 were also

significantly different from those for T1-T3, T1-T4, T2-T3 and T2-T4 but not those for T3-T4. No interaction effects reached significance.

**Table 7**

*Results of Linear Mixed-effects Model for Discrimination*

Formula (lmer): Accuracy~Group*TonePair +(1 Participant)				
	Fixed Effects			
	b	SE	t	p
Intercept	0.90	0.01	128.68	< .001
Native	-0.09	0.02	-5.28	< .001
Naïve	0.002	0.01	0.09	.93
T1-T3	0.04	0.01	3.69	< .001
T1-T4	-0.06	0.01	-6.13	< .001
T2-T3	-0.05	0.01	-4.86	< .001
T2-T4	0.03	0.01	2.48	.01
T3-T4	0.01	0.01	0.67	.50
Native: T1-T3	0.01	0.03	0.50	.61
Naïve: T1-T3	-0.01	0.03	-0.39	.69
Native: T1-T4	-0.04	0.03	-1.68	.09
Naïve: T1-T4	0.003	0.03	0.13	.90
Native: T2-T3	-0.02	0.03	-0.71	.48
Naïve: T2-T3	-0.008	0.03	-0.31	.76
Native: T2-T4	0.03	0.03	1.21	.23
Naïve: T2-T4	0.003	0.03	0.12	.91
Native: T3-T4	-0.005	0.03	-0.20	.84
Naïve: T3-T4	0.000	0.03	0.002	.998

In order to explore how different groups perform in discriminating different tone pairs, I performed analyses within each group (*lmer(Accuracy~TonePair+(1|Participant)*).

Table 8 shows the means and model outputs for the tone pair comparisons within each

group. The accuracy ranking for the tone pair continua was  $T1-T3 = T2-T4 > T3-T4 = T1-T2 > T2-T3 = T1-T4$  for the native listeners,  $T1-T3 = T2-T4 = T3-T4 = T1-T2 > T2-T3 = T1-T4$  for the L2 learners, and  $T1-T3 = T2-T4 = T3-T4 = T1-T2 > T1-T4 = T2-T3$  for the naïve listeners. Though there were differences in whether or not certain tone-pair distinctions were statistically significant across the three groups, the overall ordering of the tone pairs was similar. All three groups had higher accuracy with T1-T3 and T2-T4 than with T2-T3 and T1-T4. In other words, listeners tended to perceive tone pairs with similar beginning points (T1-T4) or similar contours (T2-T3) less accurately than the other tone pairs.

**Table 8**

*Results of Comparison of Discrimination Accuracy by Tone Pair within Each Group*

	Native		L2		Naïve
T1-T3 vs. T2-T4	0.89 vs. 0.89 ( $b = -0.05$ , $p = .81$ )	T1-T3 vs. T2-T4	0.97 vs. 0.94 ( $b = -0.02$ , $p = .17$ )	T1-T3 vs. T2-T4	0.96 vs. 0.95 ( $b = -0.01$ , $p = .55$ )
T2-T4 vs. T3-T4	0.89 vs. 0.85 ( $b = -0.04$ , $p = .04$ )	T2-T4 vs. T3-T4	0.94 vs. 0.94 ( $b = -0.006$ , $p = .73$ )	T2-T4 vs. T3-T4	0.95 vs. 0.94 ( $b = -0.01$ , $p = .60$ )
T3-T4 vs. T1-T2	0.85 vs. 0.84 ( $b = -0.003$ , $p = .86$ )	T3-T4 vs. T1-T2	0.94 vs. 0.93 ( $b = -0.009$ , $p = .61$ )	T3-T4 vs. T1-T2	0.94 vs. 0.93 ( $b = -0.01$ , $p = .60$ )
T1-T2 vs. T2-T3	0.84 vs. 0.78 ( $b = -0.06$ , $p = .004$ )	T1-T2 vs. T2-T3	0.93 vs. 0.89 ( $b = -0.04$ , $p = .01$ )	T1-T2 vs. T1-T4	0.93 vs. 0.89 ( $b = -0.05$ , $p = .006$ )
T2-T3 vs. T1-T4	0.78 vs. 0.75 ( $b = -0.03$ , $p = .10$ )	T2-T3 vs. T1-T4	0.89 vs. 0.89 ( $b = -0.009$ , $p = .61$ )	T1-T4 vs. T2-T3	0.89 vs. 0.88 ( $b = -0.002$ , $p = .88$ )

*Note.* Data shows means of different tone pairs within each group, as well as the coefficient ( $b$ -value) and its significance ( $p$ -value) in each comparison.

In sum, the L2 learners performed similarly to the naïve listeners in the discrimination task. Both the L2 learners and the naïve listeners achieved higher accuracy than the native listeners. Visual inspection of the discrimination curves indicated that this was due mostly to the fact that the native listeners had lower accuracy on step-pairs that were likely to be within a tone category. Since the naïve listeners did not have knowledge of Mandarin tones and had to rely on acoustic differences to complete the task, it is understandable that they did not show any evidence of CP. Unexpectedly, the L2 learners behaved very similarly to naïve listeners, with no evidence of CP, which was also inconsistent with their identification results, where the L2 learners patterned between the other two groups. I will return to this unexpected difference between the two tasks in the discussion section (3.4).

### 3.3.3 Proficiency

First, I examined the correlation between the different proficiency measurements, i.e., the scores from the listening task and the self-ratings of Chinese listening (CL), speaking (CS) and reading (CR) ability. Since the data were not normally distributed and the sample size was small (27 L2 learners), I used the Kendall test to measure the correlations between different measurements of proficiency. The results showed that there was at least a medium-sized correlation between all the self-rating scores (CS & CL: 0.81; CS & CR: 0.67; CL & CR: 0.62). I then calculated an overall self-rating score for each individual by adding up the three individual scores and dividing by 15 ( $M = 0.43$ ,  $SD = 0.24$ , range: 0.20–1.00)<sup>3</sup>. The overall self-rating scores and listening test scores were also significantly correlated ( $\tau = 0.61$ ,  $z = 3.97$ ,  $p < 0.01$ ). By averaging those two scores, I obtained a composite proficiency score for each participant ( $M = 0.58$ ,  $SD = 0.23$ , range: 0.10–1.00).

Results from another Kendall test showed that this composite proficiency score was significantly correlated with the identification slopes for the L2 group ( $\tau = -0.39$ ,  $z = -2.71$ ,  $p = .007$ ; Figure 9A), such that participants with higher proficiency scores tended

---

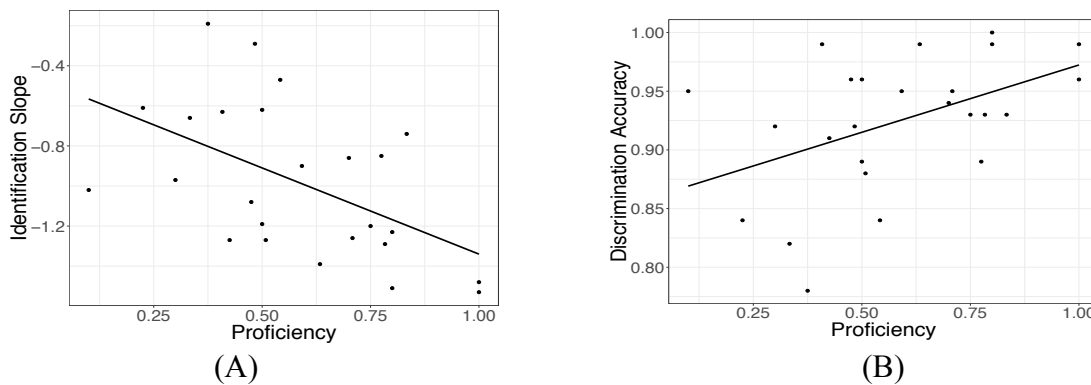
<sup>3</sup> The maximum rating in each category is 5 points and there are three categories: CL, CS and CR.

to have steeper identification slopes. This finding suggested that L2 proficiency was significantly correlated with learners' degree of CP of tone, which has not been explored in previous studies.

The composite proficiency scores were also significantly correlated with the discrimination accuracy scores ( $\tau = 0.30$ ,  $z = 2.04$ ,  $p = .04$ ; Figure 9B), such that participants with higher proficiency had higher discrimination accuracy.

**Figure 9**

*Correlations between L2 Proficiency and Identification Slope (A) and between Proficiency and Discrimination Accuracy (B)*



### 3.4 Discussion

The goals of this study were (a) to investigate how L2 learners of Mandarin perceive lexical tone in comparison to native Mandarin listeners and naïve listeners (RQ1), (b) to explore how L2 proficiency influences L2 learners' identification and discrimination of tone (RQ2) and (c) to examine how the degree of similarity between different tone pairs influences perception of lexical tone (RQ3). In this study, categorical perception (CP) was envisioned as a relative concept rather than an absolute one. For this reason, I adopted a relatively liberal standard for identifying CP (Harnad, 2003), by comparing the perception of L2 learners to that of native Mandarin listeners and English-speaking naïve listeners.

For RQ1, I found that the L2 learners behaved differently from the other two groups. Compared to the naïve listeners, the native listeners had steeper identification slopes in



perceiving pitch and lower accuracy for discriminating within-category pitch tokens, both of which suggested that the perception of tones was more categorical for native listeners. In contrast, the naïve listeners' identification curves were more linear and they had similar rates of accuracy across all step-pairs on the discrimination task. These results are consistent with previous studies that have found that naïve listeners' perception is more psychophysically based and non-categorical compared to that of native listeners (Francis, 2003; Hallé et al., 2004; Peng et al., 2010; Wang, 1976; Xu et al., 2006). The L2 learners, who had a variety of Mandarin proficiency levels, patterned between the other two groups on the identification task and showed a similar pattern to the naïve listeners on the discrimination task. One explanation for this inconsistent pattern of results might be that the two tasks had different requirements.

In the current identification task, listeners were asked to identify individual auditory stimuli by grouping them into one of the two tonal categories in each block. Though there was a brief exposure session to inform listeners about the two choices at the beginning of each block, the listeners had to access their knowledge of tone representations to complete the task. For the naïve listeners, it might not be an easy task, because they did not have mental representations of tones and had to access the short-term memory traces they had built in the exposure session, which might not be reliable. However, the L2 learners could access the mental tone categories that they had established during their learning experience. As shown in the identification data, the L2 learners had steeper identification slopes than the naïve group, though they were shallower than those of the native group. This indicated that L2 learners accessed their developing mental representations of tone for tone identification, though the representations might not be as robust as those of native listeners.

In the discrimination task, participants had to decide whether the middle sound matched the first sound or the last sound in each AXB sequence. Unlike the identification task, which required participants to access their mental tone categories, the discrimination task could be completed by simply comparing consecutive tokens without any reference to tone categories. As a result, the naïve listeners had higher rates of discrimination accuracy than the native speakers, which stood in stark contrast to their performance on the identification task. This could be due to the native listeners' reduced sensitivity to

within-category differences, which led to overall lower discrimination accuracy. Although the native listeners may very well have been trying to focus on the acoustic differences between consecutive tokens, which is the most effective strategy for succeeding in the discrimination task, it is possible that the perception routines of accessing mental tone categories were so automatic that the native listeners could not inhibit them. This automaticity might be a result of lifelong native language use, which enables native listeners to efficiently and effortlessly extract phonetic information and identify phonological categories for word recognition.

As for the L2 learners, since long-term learning experience may be able to change the way they perceive lexical tone (Wang et al., 1999; Wang et al., 2004; Wayland & Guion, 2003), I expected to see a clear difference between the discrimination patterns of the L2 learners and the naïve listeners, like the one that had already been observed in the identification data. However, the discrimination results showed no significant differences between the naïve listeners and the L2 learners, indicating that the two groups made use of similar processing mechanisms. Unlike identification, discrimination does not require listeners to access their mental tone categories, and it might be the case that L2 learners, like naïve listeners, mostly rely on short-term memory traces of acoustic differences between consecutive tokens. Thus, accessing mental tone categories might be more task-dependent and less automatic for L2 learners than for native listeners.

However, we should be cautious about interpreting the unexpected findings with respect to the L2 learners' performance on the discrimination task. Many factors, such as the length of the inter-stimulus-interval (ISI) and step sizes used when preparing the audio stimuli, could have influenced the results. For example, the relatively long ISI used by Shen and Froud (2016) might explain why their L2 learners had rates of discrimination accuracy that were similar to those of the native listeners and higher than those of the naïve listeners. Although Shen and Froud (2016) set the ISI at 500 milliseconds, all the test stimuli were presented at the end of a Chinese phrase, which made the actual time intervening between the presentations of the two critical tokens much longer than 500 milliseconds. Previous studies have shown that longer ISIs benefit participants in tasks that involve labeling, whereas shorter ISIs are advantageous in tasks that involve comparisons based on short-term sensory traces (van Hoesen & Schouten,

1992). For naïve listeners, labeling unfamiliar lexical tones is a difficult task and their discrimination accuracy could be greatly influenced by the length of the ISIs. However, in the present study, the naïve listeners could still use their sensory traces of the small acoustic differences between consecutive tokens with the 500ms ISI. A longer ISI might also make L2 learners process pitch in a more phonological mode, where accessing mental tone representation becomes necessary. This might also lead to more obvious discrimination accuracy peaks, such as those observed in Shen and Froud's (2016) study. Aside from ISIs, some other methodological differences between Shen and Froud's (2016) study and the present project that might have contributed to the different findings include the number of participants, the tone types used and the proficiency levels of the L2 learners. However, the main focus of the present study was to investigate differences in tone perception between the L2 learners and the other two groups rather than the factors that can influence their tone perception.

For RQ2 about how L2 learners' proficiency affects their performance, results showed that L2 proficiency was positively correlated with both identification slopes and discrimination accuracy. The positive correlation between L2 proficiency and identification slopes suggested that learners with higher L2 proficiency tended to perceive tones more categorically. This finding is consistent with Shen and Froud's (2016) finding that L2 learners with advanced proficiency have native-like identification slopes. L2 learners' proficiency scores were also positively correlated with their discrimination accuracy scores, suggesting that learners with higher L2 proficiency might also have higher sensitivity to pitch differences. This result was consistent with Bowles et al.'s (2016) findings that perceptual sensitivity to pitch differences plays an important role in learning a tonal language.

As for RQ3 about how the degree of similarity between different tone pairs influences perception of lexical tone, the results were consistent across groups on both the identification and discrimination tasks. On the identification task, all three groups perceived T1-T3 and T1-T2 the most categorically and T1-T4 and T2-T3 the least categorically among the six tone pairs included in the study. On the discrimination task, all three groups achieved higher discrimination accuracy with T1-T3 and T2-T4 than with T2-T3 and T1-T4. Thus T1-T3, the pair with tones that differed in terms of contours,

beginning points and end points, was associated with the steepest slopes on the identification task and the highest accuracy rates on the discrimination task. Conversely, T2-T3, the pair with the most similar contours, and T1-T4, the pair with the most similar beginning points, were associated with the shallowest slopes on the identification task and the lowest accuracy rates on the discrimination task. Although there were some small discrepancies between the groups on the two tasks, overall I observed similar patterns for all three groups, regardless of their language experience.

In the end, unlike previous studies that have used a combination of identification and discrimination tasks to test categorical perception, the results of the current experiment suggested that identification tasks might be more appropriate than discrimination tasks for testing the degree of CP of lexical tone by L2 learners. If resources are limited, identification tasks similar to the one used in the current study could be a good choice for measuring listeners' CP of lexical tone.

However, the discrimination task in the current study did show one important difference between native listeners and L2 learners that is sometimes reported in other studies, namely that L2 learners appear to process lexical tone in a less automatic manner. In a recent study, Pelzl et al. (2019) found that even though advanced learners of Mandarin were as successful as native listeners at identifying tone in isolated syllables, they had trouble correctly rejecting non-words that differed from real words only in tone, thereby indicating a “disconnect between L2 abilities to categorize tones as phonetic objects and abilities to utilize those categories as lexical cues” (p.11, Pelzl et al., 2019). This disconnect might be due to relatively low levels of automaticity in the accessing of tone categories by L2 learners. Compared to native listeners, L2 learners have less robust mental tone categories, as shown in the identification task, and less automatic access to those categories for tone perception, as shown in the discrimination task, both of which might lead to L2 learners placing less weight on tonal information than on segmental information during word recognition (e.g. Pelzl et al., 2019, 2020; Qin, 2017). Experiment 2 was designed to explore this possible continuity of tone processing at the phonological level and at the lexical level.

## **Chapter 4. Experiment 2: The relation between perception of tone and real-time spoken word recognition by L2 learners<sup>4</sup>**

### **4.1 Research questions**

The goal of Experiment 2 is to examine the relation between learners' difficulties with lexical tone at the phonological level in L2 speech perception and the challenges they encounter when using lexical tone at the level of lexical access during L2 word recognition. More specifically, I aim to examine how much weight L1 and L2 speakers of Mandarin put on tonal cues relative to segmental cues for lexical access during real-time listening, as well as to what extent they perceive tone categorically. This will allow me to then examine to what extent the two are related. To this end, I report findings from a visual world eye-tracking experiment designed to investigate the role of tonal cues in Mandarin word recognition, as well as from a tone identification task similar to the one in Experiment 1. Based on the findings from these two tasks, I will address the following two research questions:

- RQ1: How do L1 and L2 listeners weight tonal cues relative to segmental cues in Mandarin spoken word recognition?
- RQ2: How does L2 learners' use of tonal cues in spoken word recognition relate to their ability to perceive tone categorically?

Previous studies showed that L2 learners had more trouble than native speakers when it came to utilizing all relevant phonological cues during word processing due to the difficulty that learners had with identifying phonemes in the L2 (Broersma, 2012; Broersma & Cutler, 2008, 2011; Qin, 2017). Mandarin words are composed of both lexical tones and segments, and L2 learners have been reported to have persistent difficulty with processing lexical tone (Pelzl et al., 2019, 2020; Qin, 2017). Thus, I predicted that I would find evidence that L2 listeners attend less to tonal cues than to segmental cues during Mandarin spoken word recognition compared to L1 listeners, and

---

<sup>4</sup> An article based on this experiment has been published: Ling, W., & Grüter, T. (2020). From sounds to words: The relation between phonological and lexical processing of tone in L2 Mandarin. *Second Language Research*. doi: 10.1177/0267658320941546

a positive correlation between L2 learners' ability to perceive tone categorically and their use of tonal cues in spoken word recognition.

## 4.2 Methods

### 4.2.1 Participants

A total of 30 native and 34 L2 speakers of Mandarin participated in this study. Data from 5 L2 participants was excluded due to exposure to Chinese in childhood (4) or professional music experience (1), leaving data from 29 L2 speakers for analysis. Native speakers (21 female, mean age: 25.6 years, range: 20–36) were recruited from among the international student community at the University of Hawai'i at Mānoa ( $N = 20$ ), as well as at Peking University ( $N = 10$ ). All native speakers reported being born in Mainland China and self-identified as native speakers of Mandarin.

L2 learners (11 female, mean age = 24.4 years, range: 19–41) were recruited at the University of Hawai'i at Mānoa ( $N = 3$ ), as well as at Peking University, University of Hong Kong and Chinese University of Hong Kong ( $N = 26$ ). All L2 learners self-identified as native speakers of English and started to learn Mandarin after age 12 (mean age of onset: 20.0 years,  $SD = 4.5$ ). To ensure basic familiarity with the Mandarin tonal system and the vocabulary used in the experimental materials, only participants who were taking or had taken 3rd-year Chinese classes (or above) in the U.S., or intermediate/advanced classes in China, were admitted to the study. L2 proficiency was assessed through self-ratings of speaking ( $M = 2.8$ ,  $SD = 1.1$ ), listening ( $M = 2.9$ ,  $SD = 1.1$ ) and reading ( $M = 3.0$ ,  $SD = 1.0$ ) skills on a 5-point scale, as well as through a listening proficiency test adapted from the Hanyu Shuiping Kaoshi (HSK or Chinese Standard Exam) level 4 (Confucius Institute in Atlanta, 2017). The HSK level 4 test is a boundary test to differentiate intermediate to advanced proficiency. There were two parts of the listening test. Part 1 had 10 items, involved listening to short sentences. For each item, participants needed to decide whether a statement is true or false according to the sentences they heard. Part 2 involved listening to 15 short conversations. For each item, participants needed to choose one correct answer from four options according to the conversations they heard. All participants were given the test with instructions in Chinese.

The complete test is provided in appendix B. Correct answer for each item was given 4% and the maximum score on the task was 100% (4 \* 25). One L2 learner did not complete the listening proficiency test. The remaining 28 scored from 40% to 100% ( $M = 76.9\%$ ,  $SD = 17.5\%$ ). Information on music experience and language experience was collected by questionnaire. The study protocol was approved by the Institutional Review Board at the University of Hawai'i, and participants were compensated with extra course credit or a small amount of money.

#### **4.2.2 Spoken word recognition task**

##### 4.2.2.1 Materials

Linguistic stimuli in the visual-world eye-tracking experiment consisted of 12 sets of 5 monosyllabic words. All words were easily imageable common nouns, and were composed of a consonant onset and a rhyme (see Appendix C for a complete list of stimuli). Each set consists of (i) a target (e.g., *gou3* 'dog'); (ii) a segmental competitor (SC: *gou1* 'hook'), which completely matches the target in segmental content but differs in tone; (iii) a rhyme competitor (RC: *shou3* 'hand'), which matches the target in segmental and tonal content of the rhyme but differs in onset; (iv) a vowel competitor (VC: *dou4* 'bean'), which matches the target in segmental but not tonal content of the rhyme and also differs in onset; and (v) a distractor (*qiu2* 'ball'), which does not share either segmental or tonal content with the target. Due to the limited number of natural words forming such sets of five, I was unable to fully control the tone pairs in the stimuli. However, the number of different tone types were approximately balanced across targets, the three types of competitors, and distractors. The SC and VC competitors were included to allow for the critical comparison between competitors that differ from the target in tone only (SC) versus competitors that differ in both tone and segmental (onset) content (VC). Rhyme competitors (RC), which share tone and vowel but not onset segmental content with the target, were included to assess potential late co-activation due to overlapping rhymes (Allopenna et al., 1998).

In order to examine potential differences in word frequency between stimulus types, I used word frequency indices ( $\log_{10}W$ ) from the SUBTLEX-CH corpus (Cai & Brysbaert,

2010). A one-way ANOVA showed no significant differences between targets, the three types of competitors, and distractors ( $F(4,55) = 1.58, p = .19$ ). Since indices of frequency in an L1 corpus may not be fully reflective of frequency experienced by L2 learners, I also used HSK vocabulary level as an index of word difficulty for L2 learners (level 1 = easiest to level 6 = most difficult). Words not listed in the HSK vocabulary were given a value of 7 (5 words). L1 word frequency ( $\log_{10}W$ ) and HSK level index correlated moderately ( $\tau = -0.50, p < 0.001$ ), indicating some consistency between the two values. A one-way ANOVA with HSK level as the dependent variable showed no significant differences between stimulus types ( $F(4,55) = 1.24, p = 0.30$ ) either. Differences in word frequency were thus unlikely to greatly influence looks to targets, competitors, and distractors in the present study.

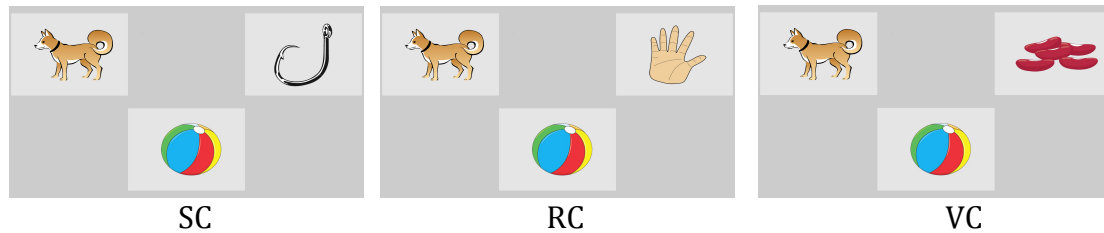
Visual scenes contained three areas of interest (AOIs): the target, one of the three competitors (SC, RC, or VC), and a distractor (Figure 10). Participants saw each target in three conditions (SC, RC, or VC), for a total of 36 experimental trials. The location of different AOIs was rotated across conditions. The order of items was pseudo-randomized, and interspersed with 54 filler trials. Fillers were constructed from the same 60 words used in the experimental trials, but words constituting competitors or distractors in experimental trials acted as targets in filler trials. Fillers were created to approximately balance the occurrence of each image as a named vs. unnamed referent. All participants were presented with the same 90 trials in two blocks, with 18 experimental trials and 27 fillers in each block. Block order was counterbalanced across participants. Three initial practice trials similar to filler trials familiarized participants with the task.

The auditory stimuli were produced by an adult female native speaker of Mandarin at a slow speed in a sound-proof booth at 44.1k Hz and recorded in Praat (Boersma & Weenink, 2016). Each noun was spoken preceded by the carrier phrase *qing3xuan3...* ('please choose...') three times in citation form with full realization of tones. One token of each noun was selected according to intensity and sound quality. Nouns were then extracted and concatenated with the same token of the carrier phrase for all items, with noun onset 2,042 ms after onset of the carrier phrase. Average duration of target nouns was 658 ms (range: 429–913). Two native speakers of Mandarin confirmed the naturalness of the concatenated sentences.



**Figure 10**

*Examples of Visual Scenes in the 3 Conditions*



*Note.* SC = segmental competitor. RC = rhyme competitor. VC = vowel competitor. The locations of the areas of interest (AOIs) were rotated in the actual materials.

#### 4.2.2.2 Procedure

Prior to the visual-world eye-tracking experiment, participants completed a self-paced vocabulary familiarization task followed by a naming test, in order to ensure understanding of all vocabulary and word-image associations. In the familiarization task, participants were presented with all 60 words (12 sets \* 5 words) in random order in PsychoPy (Peirce, 2007). Each word was presented auditorily once together with its corresponding image, Chinese characters and English gloss. Participants were instructed to take their time to familiarize themselves with each word and image before pressing the space bar to move on. They were told they would be tested in a naming task afterwards. In order to make sure participants paid attention and were familiar with the word-image associations, 10 words were selected and used as test trials in the naming task. To ensure L2 learners were familiar with the tested words, I asked three Chinese instructors to check the stimuli and select the 10 words their students might have most difficulty with. For the 10 test items in the naming task, all 5 words not listed in HSK level 1-6 were included in addition to another 5 words selected as difficult by instructors. In each test trial, participants saw an image and after 500 ms they heard a beep and were required to name the picture. Participants needed to produce 9/10 words with correct pronunciation of segments, as judged by the experimenter, to pass the naming task; otherwise they were asked to repeat the familiarization task and naming test until they met criterion. This step was included to ensure listeners paid attention and were sufficiently familiar with the

vocabulary used in the experiment. All native listeners and 11 L2 learners passed the naming task the first time. 15 L2 learners repeated the familiarization task once and 3 L2 learners repeated it twice. No feedback was provided during the naming task.

After passing the naming task, participants proceeded to the main part of the spoken word recognition experiment. The experiment was conducted on an SMI RED250 eye-tracker sampling at 250 Hz (for participants tested in Hawai‘i), or a mobile REDn Scientific eye-tracker sampling at 60 Hz (for participants tested in China). Participants were instructed to click on one of the three images in the scene after listening to the auditory instruction *qing3xun3* (‘Please choose’) + NOUN, preceded by 1,500 ms preview of the visual scene. Mouse clicks and eye fixation were recorded through SMI ExperimentSuite software. Fixation data were binned into 20 ms samples. Preliminary analyses showed no differences in the structure of the data from the two eye-trackers, thus all data was combined for further analysis.

#### ***4.2.3 Identification task***

The identification task was the same as the one in Experiment 1 (see 3.2.3.1), except only stimuli with /pi/ were included.

#### ***4.2.4 General procedure***

Before coming to the lab, all participants completed a web-based questionnaire to collect information on basic demographics, language background, music experience, and self-ratings of speaking, listening and reading ability in both Mandarin and English. In the lab session, all participants completed the visual-world eye-tracking experiment followed by the identification task. L2 learners additionally completed the listening proficiency test at the end.

### **4.3 Results**

Data from a total of 2,124 trials (59 participants, 36 experimental items) on the eye-tracking experiment was first inspected for valid mouse-click responses. Trials with no

mouse click (L1: 1, L2: 1) and trials in which the participant clicked on an image before noun onset (L1: 4) were excluded, as were trials in which the timing of the click exceeded 3 *SDs* of the group's average RT (L1: 16, L2: 17). The remaining data were inspected for track loss. Trials with more than 16% (=  $M+3SD$ ) missing sample points were excluded (L1: 25, L2: 4). In all, a total of 3.2% of the data (68/2124 trials; L1: 4.3%, L2: 2.1%) was discarded.

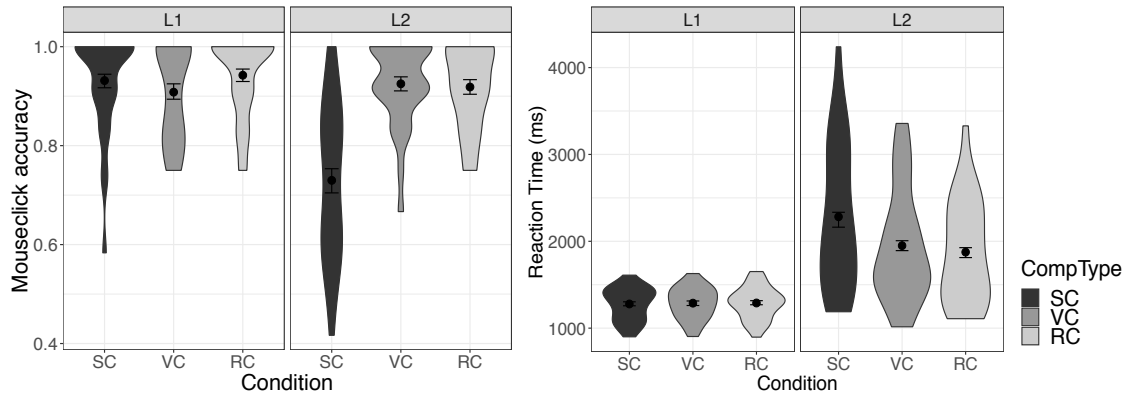
### ***4.3.1 Spoken word recognition task***

#### 4.3.1.1 Mouse-click data

Participants' accuracy in selecting the named target is illustrated in Figure 11 (left). While the L1 group showed similar accuracy rates across conditions, accuracy in the L2 group was substantially lower in the SC condition than in the other two. For statistical analysis, accuracy data were submitted to a generalized linear mixed effect model. This and all subsequent statistical analyses were conducted in R (version, 3.6.0, R Core Team, 2019), using the lme4 package (version 1.1-21, Bates et al., 2015). Fixed effects included Group (L1, L2; contrast-coded and centered), Condition (SC, RC, VC; simple coded with VC as reference level) and their interactions. Maximal random effect structures justified by the design were attempted, and reduced if convergence problems arose (Barr et al., 2013). Model comparisons were carried out using the *anova()* function to identify the best-fitting model.

**Figure 11**

*Accuracy (left) and Reaction Time (right) by Group and Condition*



*Note.* The dot represents the mean in each condition and group. Error bars indicate one standard error by participant in each condition and group. SC = segmental competitor. RC = rhyme competitor. VC = vowel competitor.

Table 9 presents the output of the best-fitting model. The significant negative estimate for group ( $b = -0.72, p = .01$ ) indicates lower accuracy overall in the L2 than in the L1 group. The significant negative estimate for SC ( $b = -1.33, p = .03$ ) indicates overall lower performance in the SC than in the (reference-level) VC condition. In other words, participants were less accurate in selecting the named target when there was a competitor that differed only in tone versus a competitor differing in both tone and segmental content. Importantly, this effect interacted with Group ( $b = -2.04, p < .001$ ), prompting follow-up analyses within each group separately. Within the L1 group, there were no significant differences in accuracy across conditions (all  $p > .4$ ). In the L2 group, on the other hand, accuracy in the SC condition was significantly lower compared to the VC ( $b = -2.24, p < .001$ ) and RC ( $b = -3.10, p = .002$ ) conditions, with no significant differences between the latter two ( $b = 0.86, p = .49$ ).

**Table 9***Results of Generalized Linear Mixed-effects Model for Accuracy*


---

Formula (glmer):  $accuracy \sim group * condition + (1 | participant) + (1 + condition | item)$

---

	Fixed Effects			
	b	SE	z	p
Intercept	3.04	0.23	13.07	< .001
Group	-0.72	0.28	-2.59	.01
RC	0.18	0.89	0.20	.84
SC	-1.33	0.61	-2.18	.03
Group x RC	-0.68	0.45	-1.51	.13
Group x SC	-2.04	0.40	-5.09	< .001

---

Analyses of reaction time (RT) in trials with correct mouse clicks (Figure 11 right) echoed the results from the analysis of accuracy. RTs in the L2 group ( $M_{L2} = 2036$ ,  $SD_{L2} = 716$ ) were substantially longer overall than in the L1 group ( $M_{L1} = 1285$ ,  $SD_{L1} = 184$ ), indicating generally greater difficulty in recognizing Mandarin words among L2 learners. Statistical analysis was conducted using mixed-effect models of the inverse-gaussian family due to the skewed distribution of the RT data (Lo & Andrew, 2015). Otherwise the same modeling strategies were followed as in the analysis of accuracy. Results from the best-fitting model ( $RT \sim group * condition + (1 | participant) + (1 + condition | item)$ ) confirmed that the L2 group took longer than the L1 group in making correct choices ( $b = 700.69$ ,  $p < .001$ ), and that participants took longer in the SC than in the VC condition ( $b = 111.47$ ,  $p = .02$ ). The interaction between Group and Condition (SC vs. VC) was significant ( $b = 181.19$ ,  $p < .001$ ), prompting follow-up analyses within each group. In the L1 group, there were no differences in RT by condition (SC vs. VC:  $b = 20.78$ ,  $p = .29$ ; RC vs. VC:  $b = 2.44$ ,  $p = .90$ ; SC vs. RC:  $b = -18.35$ ,  $p = .37$ ). The L2 group, by contrast, took substantially longer to make correct choices in the SC condition compared to the VC ( $b = 194.13$ ,  $p < .001$ ) and the RC ( $b = 216.17$ ,  $p < .001$ ) condition, with no difference between the latter two ( $b = -21.93$ ,  $p = .46$ ).

In sum, the L2 group achieved accuracy comparable to the L1 group in the RC and VC conditions, and within the L2 group, learners were equally fast on correct target selections in these two conditions. In the SC condition, on the other hand, where tone was the only cue distinguishing the target from the competitor, L2 participants were significantly less accurate than L1 participants, and took longer on correct selections than in the other two conditions. L2 participants also showed substantially more variability on both accuracy and RT in the SC condition than L1 participants (Figure 11). It is possible that this variability stems from the inclusion of L2 participants who were unable to distinguish words by tone alone, and were thus simply guessing in the SC condition.

In order to identify such participants, I examined the probability of a participant guessing in the SC condition based on a binomial distribution. Assuming that the critical choice was between the target and the competitor (even though there was a third, phonologically unrelated distractor in the scene), chance was assumed to be at 0.5. Adopting an alpha level of 0.05, the binomial distribution indicates that correct responses on at least 9 out of 12 items represents performance significantly above chance. All participants in the L1 group met this criterion, as did 15 out of the 29 L2 learners. I will refer to this subgroup as the ‘L2-above-chance learners’. The remaining 14 L2 participants were at chance (‘L2-at-chance learners’). Proficiency measured on the listening task was higher in the L2-above-chance ( $M = 0.86$ ,  $SD = 0.13$ ) than the L2-at-chance ( $M = 0.67$ ,  $SD = 0.17$ ) subgroup ( $b = -0.19$ ,  $p = .002$ ).

In order to examine whether L1-L2 differences in the SC condition persist when comparing only L2 learners with statistically significant sensitivity to tones (the above-chance-subgroup) with L1 speakers, I reran the analysis of accuracy reported above with Group treated as a 3- rather than a 2-level factor (L1, L2-above-chance, L2-at-chance; simple coded with L1 as reference level). Results showed no significant difference in overall accuracy between the L2-above-chance and the L1 group ( $b = -0.005$ ,  $p = .99$ ), while the L2-at-chance group performed significantly below both ( $bs > |1.20|$ ,  $ps < .001$ ). Interactions between Group and Condition (SC-VC) remained significant for both the L2-at-chance vs. L1 ( $b = -2.17$ ,  $p < .001$ ) and the L2-above-chance vs. L1 ( $b = -1.80$ ,  $p = .001$ ) comparisons, but were non-significant for the L2-at-chance vs. L2-above-chance comparison ( $b = -0.37$ ,  $p = .5$ ). Within-group analyses showed no significant differences

between the RC and VC conditions in either L2 subgroup. In the L2-at-chance group, accuracy in SC was significantly worse than in the VC condition ( $b = -2.06, p < .001$ ); in the L2-above-chance group, this difference was only marginally significant ( $b = -1.58, p = .053$ ).

I thus find the pattern of results from the initial comparison between the L1 and L2 groups repeated in the comparison between the L1 and the L2-at-chance subgroup. This is unsurprising given that this L2 subgroup was defined by chance performance when the recognition of the target critically required reliance on tone. More importantly, I also find the pattern largely repeated, though somewhat weaker, in the comparison between the L1 and the L2-above-chance group. Notably, the interaction between group and the SC-VC comparison remained significant, and follow-up analysis within the L2-above-chance group still showed a marginal trend towards lower accuracy in the SC than the VC condition. Analogous analyses of RT on correct responses further showed that, unlike in the L1 group (see above), RTs in the SC vs. the VC condition were longer in both the L2-at-chance ( $b = 258.35, p = .002$ ) as well as the L2-above-chance ( $b = 179.05, p < .001$ ) subgroups. These findings suggest that even for L2 learners with demonstrated above-chance ability to recognize target nouns by tone alone, performance does not fully mirror that of L1 speakers.

#### 4.3.1.2 Eye-movement data

In order to further explore these differences between the L2-above-chance and the L1 groups, I investigated the time course of participants' looks to targets and competitors in the visual scene as they were listening to the noun in real time. Figure 12 illustrates L1 and L2-above-chance participants' looking patterns in the SC, VC and RC conditions on trials in which they selected the correct target. Visual inspection of fixation patterns in the L1 group shows little evidence of competition in any condition, with looks to competitors decreasing sharply, along with looks to phonologically unrelated distractors, about 200ms after the onset of the noun. In the L2-above-chance group, looks to the target in the VC and RC conditions increase on a similar timescale as in the L1 group, but

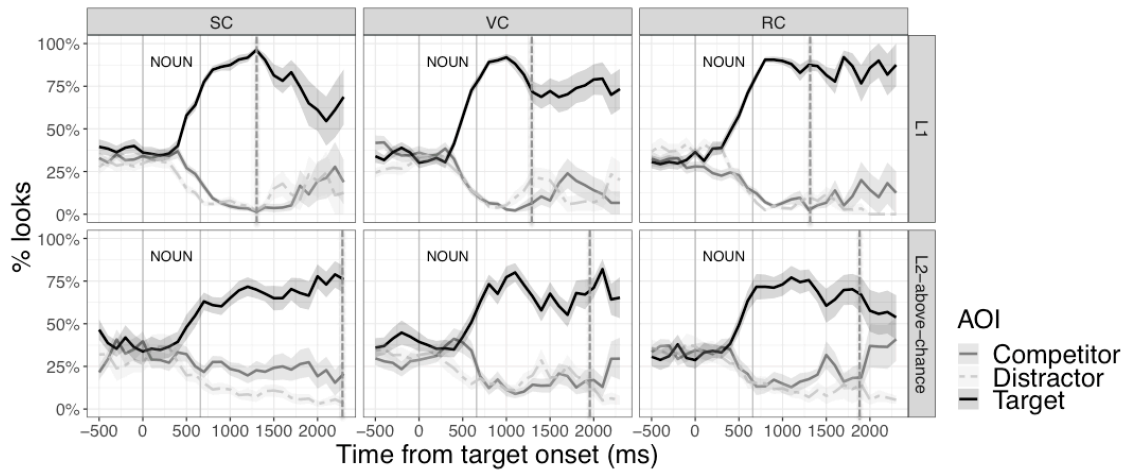
asymptote at a lower level. At the same time, looks to competitors remain more persistent, a pattern that appears particularly evident in the SC condition.

Statistical analysis was conducted to address the research question on L1 and L2 listeners' relative weighting of tonal and segmental cues (RQ1). Specifically, my goal was to assess whether competition from a competitor differing only in tone would be stronger than from a competitor differing in both tone and segmental content, and whether this effect would be more pronounced in the L2 than in the L1 group. To this end, I compared the proportion of looks to the competitor (versus the target) in the SC versus the VC and RC conditions in both groups. The large difference in RT between the two groups (see above), however, raised the difficult question of the appropriate time period within which to analyze these looking patterns. Choosing a specific window based on any previous studies could be a form of selection bias that will systematically distort the result. Thus, I decided to honor the variability in the timing of participants' decisions, as captured already by RT, and focus on a participant-driven time window, extending from 200 ms after noun onset (Matin et al., 1993) until mouse click, i.e., until the participant selected the (correct) target in a given trial. Within this period, which varied by trial, I calculated the proportion of frames with fixations to the competitor out of fixations to target and competitor combined. This measure captures the proportion of time the participant spent looking at the competitor before making a final decision.



**Figure 12**

*Proportion of Looks to Different AOIs over Total Looks to All Three AOIs by Condition and Group on Correct Trials*



*Note.* 0 on y-axis along with the first gray line represents noun onset; the second gray line represents mean noun offset; the third, dashed line represents average mouse-click reaction time by group and condition. Ribbon shows one standard error.

To this end, a linear mixed-effect model with Group (L1, L2-above-chance; contrast-coded and centered) and Condition (simple-coded, VC as reference) as fixed effects was fitted to these data. Given the highly non-normal distribution of the outcome measure at the trial level, models at the trial level including both random effects for subjects and items proved to be a poor fit. I therefore decided to aggregate data over subjects and over items, and fit two separate models to each (Barr, 2008). Table 10 presents the output from the best-fitting models, which showed similar patterns in the by-subject and by-item aggregations. The main effect of Group was significant ( $b_1 = 0.09, p_1 < .001; b_2 = 0.11, p_2 < .001$ ), indicating that the L2 learners were overall more likely than native speakers to look at competitors. An overall trend for more looks to competitors in the SC (vs. VC) condition also emerged ( $b_1 = 0.04, p_1 = .02; b_2 = 0.03, p_2 = .10$ ). This trend did not interact with Group, yet in light of the research question, I decided to explore its nature further through models fit to the data from each group separately. In the L1 group, no differences between SC vs. VC condition ( $b_1 = -0.02, p_1 = .16; b_2 = -0.03, p_2 = .35$ ) or

RC vs. VC condition ( $b_1 = 0.02, p_1 = .21; b_2 = 0.02, p_2 = .48$ ) emerged. Within the L2 above-chance group, a significant difference was found between the SC and VC condition in the by-participant ( $b_1 = 0.06, p_1 = .04$ ) but not in the by-item data ( $b_2 = 0.05, p_2 = .18$ ); no significant differences were observed between the RC and VC conditions ( $b_1 = 0.005, p_1 = .86; b_2 = 0.007, p_2 = .83$ ). In sum, even in trials with correct mouse click, L2 listeners with the ability to discriminate words by tone showed more consideration of competitors overall, and tended to look at competitors more when tone was the only differing cue between targets and competitors than when they differed in both tone and segmental content; native speakers, by contrast, did not show any differences between conditions.

**Table 10**

*Results of Linear Mixed-effect Model for Proportion of Looks to Competitor, Aggregated over Participants (top) and Items (bottom)*

Formula (lmer): PropCompetitor ~ group * condition + (1   <b>participant</b> )				
Fixed Effects				
	b	SE	t	p
Intercept	0.20	0.01	21.82	< .001
Group	0.09	0.02	4.51	< .001
RC	-0.01	0.02	-0.96	.34
SC	0.04	0.02	2.39	.02
Group x RC	0.03	0.03	0.93	.35
Group x SC	0.04	0.03	1.34	.18

Formula (lmer): PropCompetitor ~ group * condition + (1   <b>Item</b> )				
Fixed Effects				
	b	SE	t	p
Intercept	0.21	0.02	11.44	< .001
Group	0.11	0.02	6.45	< .001
RC	-0.01	0.02	-0.44	.66
SC	0.03	0.02	1.65	.10
Group x RC	0.03	0.04	0.81	.42
Group x SC	0.03	0.04	0.73	.47

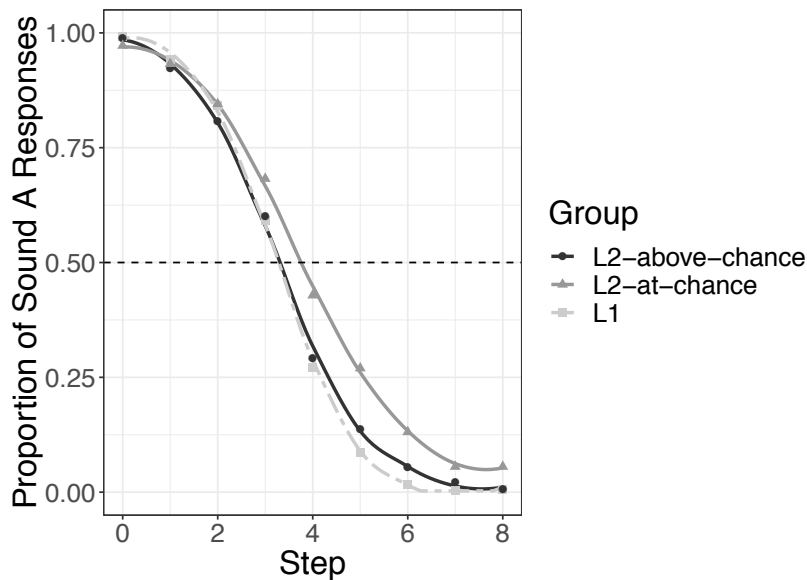
### **4.3.2 Identification task**

As in Experiment 1, I calculated the proportion of participants' Sound-A vs. Sound-B responses (e.g. Sound A = T1, Sound B = T2, in T1-T2 pair) for each tone pair and step. The results, collapsed over the six different tone pairs, are illustrated in Figure 13. At both endpoints, participants in all groups were highly accurate at identifying tone on tokens with natural pitch and intensity, indicating no general difficulty in perceiving

standard tones on isolated syllables. Visual inspection of Figure 13 shows that the L2-at-chance group had the shallowest slopes and the L1 group had the steepest slopes, while the L2-above-chance group patterned between the two. Slope values for each participant and tone pair were submitted to a linear mixed effect model with Group as a fixed effect and participants and tone pairs as random effects. Results showed that the slope for the L2-above-chance group ( $M = -1.82$ ,  $SD = 0.57$ ) was significantly steeper than for the L2-at-chance group ( $M = -1.49$ ,  $SD = 0.60$ ;  $b = 0.34$ ,  $p = .002$ ), but also significantly shallower than for the L1 group ( $M = -2.57$ ,  $SD = 0.61$ ;  $b = -0.75$ ,  $p < .001$ ).

**Figure 13**

*Identification Curve Averaged across Participants and Tone Pairs by Group*

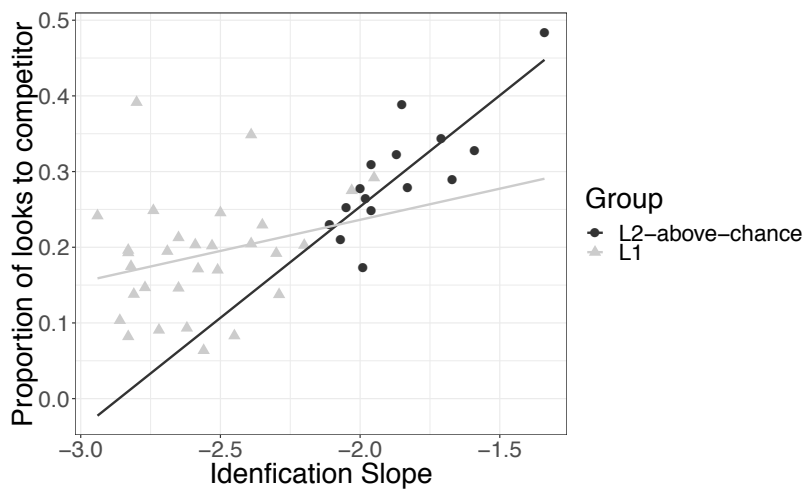


Finally, in order to investigate whether listeners' ability to perceive tone categorically was related to their use of tonal cues in spoken word recognition (RQ2), I conducted correlation tests between participants' identification slopes and their proportion of looks to the competitor in the SC condition in the visual world task. Recall that the latter was calculated over trials in which participants correctly selected the target. Since the number of such trials was low and variable in the L2-at-chance group, which was likely guessing in the SC condition, I confined this analysis to the L2-above-chance and the L1 groups. As illustrated in Figure 14, whereas no significant relation was found in the L1 group ( $\tau$

= 0.17,  $p = .19$ ), a strong positive correlation emerged in the L2-above-chance group ( $\tau = 0.63$ ,  $p = .001$ ), showing that learners with steeper identification slopes were less likely to look at competitors. These findings suggest that difficulty with using tonal cues in L2 spoken word recognition is related to the ability to perceive tone categorically at a phonological level.

**Figure 14**

*Scatterplot of Identification Slopes (x-axis) and Proportion of Looks to Competitor (y-axis) by Group*



*Note.* More negative slope values indicate steeper slopes and are indicative of more categorical perception.

#### 4.4 Discussion

The goal of this study was to investigate how L2 learners of Mandarin make use of tone during real-time word recognition (RQ1), and whether their use of tonal cues in word recognition relates to their ability to perceive tone categorically on isolated syllables (RQ2). To address RQ1, I used the visual world paradigm to assess L1 and L2 listeners' ability to use tone as a distinguishing cue in real-time word recognition. For this purpose, the analyses focused on the comparison between the SC condition, where target (e.g. *gou3* 'dog') and competitor (SC: *gou1* 'hook') overlapped completely in segmental

content but differed by tone, and the VC condition, where target and competitor (VC: *dou4* ‘bean’) differed in segmental content as well. A third condition in which target and competitor (RC: *shou3* ‘hand’) differed in onset but not rhyme was included to examine potential co-activation of rhyme competitors as found in previous work on English (Alloppenna et al., 1998). Notably, Malins and Joanisse (2010) reported no significant rhyme effects for L1 Mandarin speakers in their study. I obtained the same outcome in this study for both L1 and L2 speakers: no significant differences emerged between the RC and VC conditions for accuracy, RT, or looks to competitors, in either group. As recent work by Teruya and Kapatsinski (2019) has demonstrated, however, rhyme effects appear to be confined to disyllabic words, where the overlap between target and competitor is more extensive (e.g., *speaker-beaker* in Alloppenna et al., 1998). In both Malins and Joanisse’s (2010) and current study, all stimuli were monosyllabic. The absence of rhyme competition is thus consistent with Teruya and Kapatsinski’s (2019) observation, and will need further investigation in Mandarin in studies including disyllabic words. I will therefore confine the remainder of the discussion to the comparison between the SC and VC conditions most directly relevant to RQ1.

The overall comparison between the L1 and L2 groups showed that the latter were slower and less accurate in spoken word recognition. However, these main effects of Group interacted significantly with Condition. With regard to accuracy, the L2 group differed from the L1 group only in the SC condition, indicating that L2 learners were able to recognize words based on segmental differences as well native speakers (albeit still more slowly), yet they had considerably more difficulty in using tone alone to distinguish between words. This is consistent with Pelzl et al.’s (2019) observation that L2 learners were less accurate than native speakers at rejecting tonally, but not segmentally, mismatching non-words in a lexical decision task. Findings from this experiment add further evidence from a more ecologically valid listening task with natural stimuli showing that L2 learners allocate less weight to tonal cues than native listeners during language processing.

Further analyses within the L2 group indicated that even though I had admitted to the study only learners who were at least at the 3<sup>rd</sup>-year-Chinese level or equivalent by standards of U.S. college-level instruction, almost half of these learners (14/29) were not

significantly above chance at distinguishing words by tone alone. This is striking testimony to the observation by the Foreign Service Institute (U.S. Department of State, 2016) that Mandarin is “exceptionally difficult for native English speakers”, with the acquisition of lexical tones known to be one of the most challenging aspects for adult L2 learners of Mandarin (Wang et al., 2006). When comparing only the L2-above-chance ( $N = 15$ ) subgroup with the L1 group, the main effect of Group on accuracy disappeared; however, the interaction with Condition remained significant. Follow-up analyses showed a remaining marginal trend in the L2-above-chance subgroup towards lower performance in the SC compared to the VC condition. In order to further explore these remaining differences between native speakers and more advanced L2 learners who *are* able to differentiate words by tone alone above chance levels, I compared looking patterns to targets and competitors during real-time listening in the L1 and the L2-above-chance groups.

Overall, the L2 learners spent proportionally more time looking at the competitor before clicking on the (correct) target than native speakers, indicating greater uncertainty in all conditions. A main effect of Condition also emerged, indicating more consideration of competitors in the SC than in the VC condition; interestingly, this effect was not qualified by an interaction with Group, suggesting greater competition when words differed only by tone in both groups. Further exploratory inspection of this effect within each group, however, showed no difference between the SC and the VC condition for the L1 group, suggesting that native speakers might automatically process tone and segment as one unit for word recognition, regardless of the number of individual phonological differences between target and competing words. The effect thus appears to be driven predominantly by the performance of the L2 learners, and it is possible that the reduced number of participants in this L2 subgroup did not afford sufficient power to detect an interaction.

Returning to RQ1, my findings from real-time word recognition provide strong evidence that distinguishing words by tone alone remains difficult even for advanced learners of Mandarin. I have shown that even learners who are able to accomplish this task with above-chance accuracy take substantially more time to do so than native speakers, and show more uncertainty in the process, as indicated by proportionally more

looks to competitors minimally differing by tone only. This persistent between-group difference is consistent with Strange's (2011) Automatic Selective Perception (ASP) model: While the learners in the L2-above-chance group have clearly acquired enough *knowledge* of tone to distinguish between words in most cases – as indicated by their above-chance accuracy in the SC condition – they still appear to rely predominantly on their L1 selective perception routines (SPRs), with focus predominantly on segmental contrasts, during L2 lexical processing. In other words, they do not (yet) appear to have developed sufficiently automatized selective perception routines that allocate tone the weight it has in L1 processing.

Whether native-like SPRs are ever attained in L2 development is a critical question that I must leave for future research with long-term immersed and highly proficient L2 learners to further explore. The present study does, however, allow us to consider the issue of L2 development to some extent, namely by looking at the relationship between learners' use of tonal cues in word recognition and their ability to perceive tone categorically—my second research question. In order to address RQ2, a 9-step tone identification task was included, modeled after standard procedures in the categorical perception literature (e.g., Hallé et al., 2004). As expected, native listeners showed steeper identification slopes, indicating more categorical perception, than L2 learners. When I divided the L2 group into above-chance and at-chance subgroups as defined by performance on the word recognition task, I found significantly steeper slopes in the L2-above-chance than in the L2-at-chance group, although the identification slopes of both were significantly shallower than those in the L1 group. The L2-above-chance group also performed significantly better than the L2-at-chance group on the independent listening comprehension task. This is consistent with Experiment 1, which showed a correlation between proficiency and the degree of categorical perception of tone among L2 learners of Mandarin. These findings indicate that increasing language experience can lead to more categorical perception of tone, even among learners whose L1 does not instantiate this phonological contrast. At the same time, the finding that the more advanced L2 subgroup still differed significantly from the L1 group contrasts with the results obtained by Shen and Froud (2016), who found no significant differences between advanced learners and native speakers. However, several factors could have contributed to this



inconsistency. In particular, Shen and Froud (2016) included only two tone pairs (T1-T4 and T2-T3), and the number of participants in their study was more limited (10 L1 and 10 L2 speakers), thus the statistical power to detect any between-group effects may have been more limited. Importantly, the two studies converge in the observation that L2 listeners tend to perceive tone more categorically with increasing learning experience.

Returning to RQ2, my critical analysis consisted of assessing potential correlations between participants' performance on the identification and the word recognition tasks. More specifically, I assessed correlations between participants' slope parameters on the identification task with their proportional looking to segmental competitors in the visual world experiment. Only trials with correct final selections in the visual-world task were included. No significant correlations were obtained within the L1 group, potentially due to limited variance on both tasks—as expected in performance relying on highly automatized routines. For the L2 group, I confined the analyses to the L2-above-chance subgroup, i.e., the learners who demonstrated that they had the *ability* to reliably distinguish words by tone alone. Within this group, I found a strong and highly significant correlation between performance on the two tasks: Learners who perceived tone more categorically were also less likely to look at segmental competitors. This observation constitutes the first evidence that I am aware of that the ability to perceive tone categorically and the use of tonal cues in lexical processing are directly related at the level of individual learners in the L2 acquisition and processing of Mandarin. As such, these findings provide support for Wong and Perrachione's (2007) claims about the continuity between phonetic, phonological and lexical skills in L2 tone learning. I caution, however, that the findings show correlation, not causation. Future work, including longitudinal data and training studies, will be needed to identify the causal direction of these effects. Yet the strong contingency between the two that I have observed here provides a promising starting point for such further investigations, leading to potentially important insights for L2 training and curriculum design in the future.

Experiment 1 and Experiment 2 both provide evidence that L2 learners of Mandarin process tone differently from native Mandarin speakers at the phonological and lexical levels. The next question that needs to be answered is how to improve L2 learning of

lexical tone. In Experiment 3, I addressed this question by examining the effectiveness of the cue-focus training method in learning novel words with different tones.

## Chapter 5. Experiment 3: Learning words with tone in different cue-focus training conditions<sup>5</sup>

### 5.1 Research questions

As discussed in the literature review in Chapter 2, teachers and educators have used cue-focus training in an attempt to draw learners' attentional focus to tone with minimally contrastive features related to pitch (e.g. level, rising, falling) to improve the learning of tone (Godfroid et al., 2017; Lin, 1985; Liu et al., 2011; Tsai, 2011), and thus to improve vocabulary learning in tonal languages. However, the effectiveness of cue-focus training in vocabulary learning is mostly taken for granted based on existing theoretical models—the Noticing Hypothesis (Schmidt, 1990), the Automatic Selective Perception Model (Strange, 2011) and the Competition Model (MacWhinney, 2005, 2012)—but has never been tested directly.

Realizing the necessity of examining the assumption that cue-focus training is beneficial, and recognizing the gap between vocabulary teaching practices and theoretical models of learning, I developed the current experiment to investigate the effectiveness of cue-focus training in word learning in a controlled laboratory setting. I drew on the methodology and design of previous laboratory-based studies on word learning in tonal languages. First, to avoid the influence of statistical regularities associated with the distribution of tone in natural language (Wiener et al., 2019), I used novel words in an artificial language (Hayakawa et al., 2020; Wong & Perrachione, 2007). Second, a single-session laboratory-based auditory novel word-learning study design (Quam & Creel, 2017) enabled me to test the immediate learning outcome after short-term training. Third, other potentially confounding factors at the participant level were controlled by only including English speakers without previous experience of any tonal languages, randomly assigning them to different training conditions, and independently assessing their pitch perception ability prior to the experiment (Wong & Perrachione, 2007). Within this

---

<sup>5</sup> A short report of this study appeared in the BUCLD proceedings: Ling, W., & Grüter, T. (2020). Learning words with lexical tone: Is manipulation of attentional focus beneficial? In M. M. Brown & A. Kohut (Eds.) *Proceedings of the 44th Annual Boston University Conference on Language Development* (pp. 308-321). Cascadilla Press.

overall study design, any differences in learning outcomes between different training groups are likely to be attributable to the experimental manipulation.

This study is designed to directly examine the effectiveness of cue-focus training, assuming that manipulation of attentional focus by presenting contrastive tonal cues during training benefits the learning of words with lexical tone. More specifically, I will address the following two questions:

- (1) Does focusing the contrastiveness of a cue (e.g., tone) in training increase learners' use of that cue in subsequent lexical processing?
- (2) Does focusing the contrastiveness of a cue in training improve learners' overall word learning outcomes?

If the cue-focus training is effective, participants trained in a cue-focus condition will outperform participants trained in a control condition. Specifically, I expect that participants trained in a cue-focus (e.g., tone) condition will have higher accuracy, faster reaction times and more looks to the target referent when its label (*/pa/-falling*) differs from that of the competitor (*/pa/-rising*) by that cue only. I also predicted that participants trained in a cue-focus condition will have overall better performance on learning words than participants trained in a control condition, where no single cue is focused (e.g., */pa/-falling & /si/-level*).

## **5.2 Methods**

### **5.2.1 Participants**

Data from a total of 90 self-identified English native speakers (male = 26, other = 1; mean age: 22, range: 18–47) recruited from the University of Hawai'i community and randomly assigned to one of the three training groups (see Section 5.2.3.1) were included for analysis. An additional nine participants took part, but their data were excluded because they knew a tonal language ( $n = 3$ ; Cantonese, Thai and Vietnamese) or due to technical problems ( $n = 6$ ). All participants reported having normal hearing and normal or corrected-to-normal vision. Information on professional music experience and language

experience was collected as part of a background questionnaire. All except 11 of the remaining 90 participants reported having some experience with learning a second language. None of those languages was a tonal language and none of the participants had professional music experience. The study protocol was approved by the Institutional Review Board at the University of Hawai‘i, and participants were compensated with extra course credit or a small amount of money.

### 5.2.2 Materials

Two consonants (/p/ and /s/), three vowels (/a/, /u/, /i/) and three tones (*rising*, *level*, *falling*) were used to create 18 novel words, simultaneously comprising six triplets minimally contrastive by vowel (e.g. /pa/-*rising*, /pu/-*rising*, and /pi/-*rising*) and six triplets minimally contrastive by tone (e.g. /pu/-*rising*, /pu/-*level*, and /pu/-*falling*). Table 11 shows all 18 words and their associated meanings.

**Table 11**

*The Artificial Vocabulary*

/pa/-rising (flower)	/pu/-rising (house)	/pi/-rising (knife)	/sa/-rising (nose)	/su/-rising (pants)	/si/-rising (fork)
/pa/-level (cup)	/pu/-level (shoe)	/pi/-level (tree)	/sa/-level (book)	/su/-level (hand)	/si/-level (plate)
/pa/-falling (fire)	/pu/-falling (bag)	/pi/- falling (hat)	/sa/-falling (ball)	/su/-falling (pen)	/si/-falling (melon)

*Note.* Each novel word, written in the International Phonetic Alphabet, is followed by its tone with the associated meaning in English shown in parentheses.

The choice of consonants, vowels and tones was based on distinctiveness and familiarity. The bilabial stop /p/ and the alveolar sibilant /s/ are different in both manner and place of articulation. Low front /a/, high front /i/ and high back /u/ are the three most

distinctive vowels and are common across languages. All these consonants and vowels are familiar to English speakers. The three tones were generated based on the pitch patterns of the Mandarin rising (T2), level (T1) and falling (T4) tones. The Mandarin dipping tone (T3) was not included because it has been reported that it is the most confusing tone for both L1 and L2 speakers of Mandarin (e.g. Hao, 2012; Pelzl et al., 2019). Since the purpose of this study is to test the effectiveness of cue-focus training, I tried to match the perceptual difficulty between segments and tones. Though English does not have lexical tone, English speakers are generally familiar with rising, falling and level pitch as intonation markers (e.g. So & Best, 2010).

The meaning of each word was chosen to be easily imageable and common. To ensure participants would associate acoustic word forms with concepts rather than specific visual images, a set of clip-art images and a set of photos depicting the same concepts were selected from online resources and used as visual stimuli for the training session and the test session, respectively (see Figures 15 and 16).

Two speakers, a male and a female, were asked to pronounce the six words with level tone in isolation slowly and clearly with normal volume in a sound-proof booth three times, and were recorded via a built-in microphone in a Mac Pro computer at 44.1kHz using Praat (Boersma & Weenink, 2016). One of the three tokens for each word was selected from each talker according to the sound quality. The selected sound files for each word and talker were normalized by intensity at 80 dB using Praat. Pitch patterns were interpolated linearly through each stimulus using the PSOLA method implemented in Praat. Following the auditory manipulation in Wong and Perrachine (2007), the pitch contours in this study were also modeled on the values obtained by Shih (1988). Each pitch pattern changed linearly from the beginning point to the end point. For each minimal tone triplet from each talker, the average fundamental frequency (F0) of the level tone was used as the starting and ending points for level tone. Based on the word with level tone, the pitch contours of rising and falling tone in each triplet were generated according to the scheme in Table 12.

**Table 12***Scheme for Generating the Pitch Contour*

Tone	Beginning point	End point
Level	$F0_L$	$F0_L$
Rising	$0.74 * F0_L$	$F0_L$
Falling	$1.1 * F0_L$	$0.385 * F0_L$

Thus, except for  $F0$ , all other acoustic parameters (including duration and voice quality) were consistent in each minimal tone triplet. Stimuli produced by the female speaker were used for the training session and stimuli produced by the male speaker were used for the test session to ensure listeners would associate the phonological representation of each word with its concept instead of relying on specific acoustic details.

**5.2.3 Procedure**

After completing a web-based questionnaire about basic demographic information and language experience, all qualified participants came to the lab for a single experimental session that lasted approximately one hour.

**5.2.3.1 Pitch contour perception test (PCPT)**

First, participants completed a PCPT (Wong & Perrachione, 2007; Perrachione et al., 2011) presented in PsychoPy2 (Peirce, 2007). The task used in this experiment is on OpenBU (<https://open.bu.edu/handle/2144/16461>). This task consisted of 120 trials (4 talkers \* 5 vowels \* 3 pitch contours \* 2 repetitions) in which a sound file was played while two arrows depicting possible pitch contours (level → rising ↗, or falling ↘) were presented on the screen. Participants were instructed to press the corresponding button on the keyboard (“1” for the left arrow, “2” for the right arrow). No feedback was provided. The task was self-paced (see Wong & Perrachione, 2007; Perrachione et al., 2011 for further details). The PCPT was included as a control measure to make sure that the three training groups with randomly assigned participants had similar auditory pitch ability,

one of the most relevant factors for predicting success in tone learning (Wong & Perrachione, 2007). After completing this task, the word learning experiment started.

### 5.2.3.2 Training session

Participants were randomly assigned to one of three training groups: Tone-focus, Vowel-focus and Control group. All participants were told that they were going to see images and hear them named; they needed to repeat the words and try to learn them because they would be tested later. In the Tone-focus group, three words with the same segments but different tones (e.g. /pu/-*rising*, /pu/-*level* and /pu/-*falling*) were presented on one slide (Figure 15a). In the Vowel-focus group, three words with the same consonants and tones, but different vowels (e.g. /pa/-*rising*, /pi/-*rising* and /pu/-*rising*) were presented on one slide (Figure 15b). In the Control group, three words that differed in more than one cue (e.g. /pu/-*rising*, /sa/-*falling* and /si/-*level*) were presented on one slide (Figure 15c). Thus all participants learned the same 18 novel words in the training session, but they were exposed to them in different combinations depending on what training group they were assigned to.

Each time a new slide appeared on the screen, an arrow would point at the first image while it was being named. Participants were instructed to press the space bar to move the arrow to the second and third images so that they would be named in turn. Each slide was presented six times in six rounds, and within each round, the order of presentation was randomized, for a total of 36 training trials. The six-round-presentation was determined by a pilot test with nine participants with no tonal language experience. The nine participants were trained in the Control group condition and tested in a testing session similar to that described below to determine how many presentations of each slide would be necessary for participants to reach approximately 75% accuracy (to ensure learning accuracy would be above-chance level, but not reach ceiling), thus leaving room for meaningful variation in accuracy between training groups. I started with six presentations based on Quam and Creel (2017) and Wong and Perrachione's (2007) selection of number of presentation for single training session. Quam and Creel (2017) presented each of 18 words eight times in one training session, while Wong and Perrachione (2007)



presented each of 16 words four times. After six presentations, results showed that the nine pilot participants achieved a mean accuracy of 76% ( $SD = 11.7\%$ ), which met my requirements.

### Figure 15

#### *Examples of Triplets in the Three Training Groups*



/pu/-rising /pu/-level /pu/-falling /pa/-rising /pi/-rising /pu/-rising /pu/-rising /sa/-falling /si/-level

(a) Tone-focus group

(b) Vowel-focus group

(c) Control group

*Note.* No labels were presented during the experiment.

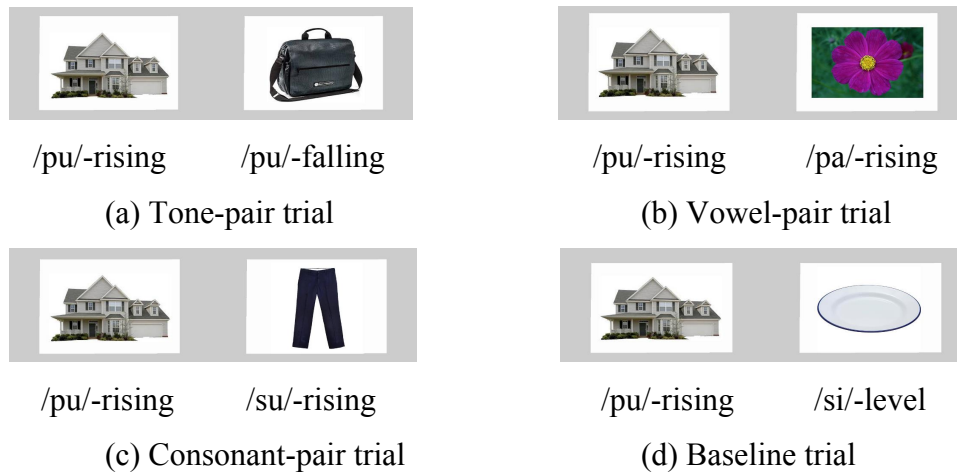
#### 5.2.3.3 Test session

Following the training session, all participants were tested in the same 2-alternative forced choice task in the visual world paradigm (VWP). This session was conducted on an SMI RED250 eye-tracker sampling at 250 Hz. Visual scenes contained two areas of interest (AOIs): the target and the competitor. On each trial, participants were presented with images of two objects while hearing one of them named and were asked to click on the named object. For each target word (e.g. /pu/-rising), there was one trial with a tone competitor differing from the target only by tone (Tone-pair trial: /pu/-falling, Figure 16a), one with a vowel competitor differing from the target by only a vowel (Vowel-pair trial: /pa/-rising, Figure 16b), one with a consonant competitor differing from the target by only a consonant (Consonant-pair trial: /su/-rising, Figure 16c), and two trials with words sharing no phonological similarities with the target (Baseline trial: e.g. /si/-level, Figure 16d). Overall, the test session consisted of a total of 90 trials (18 words \* 5 tokens) with each object appearing as a target and as a competitor five times. The location of different AOIs was rotated across trials. The order of trials was pseudo-randomized with no consecutive trials having the same targets or the same types of trials. Mouse-click location and reaction time (RT) as well as eye fixations were recorded through SMI

ExperimentSuite software. Fixation data were binned into 20 ms samples for further analysis.

## Figure 16

### *Examples of Different Trial Types*



*Note.* No labels were presented during the experiment.

## 5.3 Results

Data from 31 participants in the Tone-focus group, 31 in the Vowel-focus group and 28 in the Control group were included in the final analysis.

### *5.3.1 Pitch contour perception test (PCPT)*

As discussed above, the PCPT was included as a control measure to ensure that any differences that might be found between training groups on the experimental task were not due to pre-existing differences between groups in pitch contour awareness, one of the most relevant factors for predicting success in tone learning (Wong & Perrachione, 2007). Results from this task showed similar accuracy rates across the three groups (Tone-focus:  $M = 0.75$ ,  $SD = 0.13$ ; Vowel-focus:  $M = 0.72$ ,  $SD = 0.14$ ; and Control:  $M = 0.78$ ,  $SD = 0.14$ ). I conducted statistical analysis of these data using generalized linear mixed effect modeling with training group as a fixed effect and with intercepts for participants and

items as random effects. This and all subsequent statistical analyses were conducted in R (version, 3.6.0, R Core Team, 2019), using the lme4 package (version 1.1-21, Bates et al., 2015). I first dummy-coded training group with Control group as the reference level. I then reran the same model with Tone-focus group as the reference level. These models indicated no significant differences between the three training groups (all  $p > .11$ , see Table 13). Thus, I confirmed that participants randomly assigned to the three training groups had similar degrees of overall pitch contour awareness before training.

**Table 13**

*Results of Generalized Linear Mixed-effects Model for PCPT Accuracy*

---

Formula (glmer): PCPT\_accuracy ~ training group + (1 | participant) + (1 | item),  
family = binomial(link="logit")

---

	Fixed Effects			
	b	SE	z	p
Tone-focus vs. Control	-0.19	0.27	-0.73	0.47
Vowel-focus vs. Control	-0.43	0.27	-1.62	0.11
Vowel-focus:Tone-focus	-0.24	0.26	-0.91	0.36

---

### 5.3.2 Forced choice task

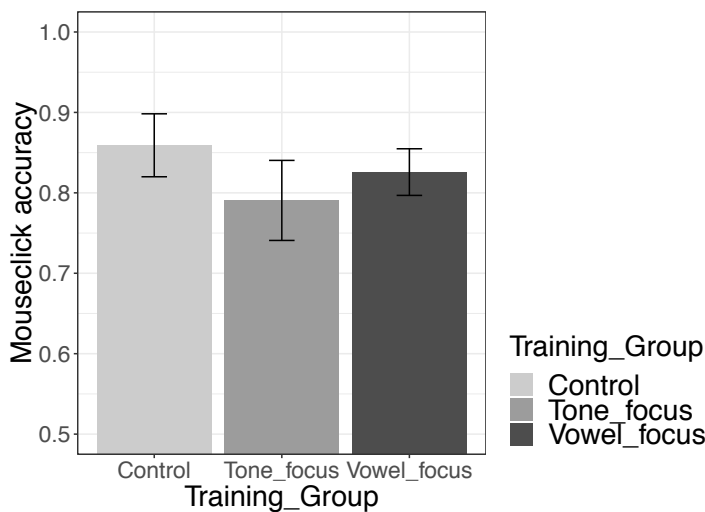
Participants in the Tone-focus group saw triplets minimally contrastive by tone. If the assumption of cue-focus training is true, participants in this group should be more likely to allocate their attentional focus to tonal cues. Similarly, participants in the Vowel-focus group should be more likely to allocate their attentional focus to vowel cues. The Control group, in which participants were not exposed to minimal triplets, served as a baseline for comparison with the two experimental groups. More specifically, I will compare the difference between the experimental groups (Tone-focus and Vowel-focus groups) and the Control group to assess the effectiveness of cue-focus training in vocabulary learning.

### 5.3.2.1 Mouse-click accuracy

After excluding one trial with a missing mouse-click response, data from the remaining 8,099 trials (90 participants, 90 test trials) were entered into the analysis. Figure 17 presents participants' overall accuracy in selecting the named target by group, collapsing over different trial types. Unexpectedly, the Control group showed the highest overall accuracy rate, while the overall accuracy in the Tone-focus group appeared to be numerically the lowest.

**Figure 17**

*Overall Accuracy by Training Group*



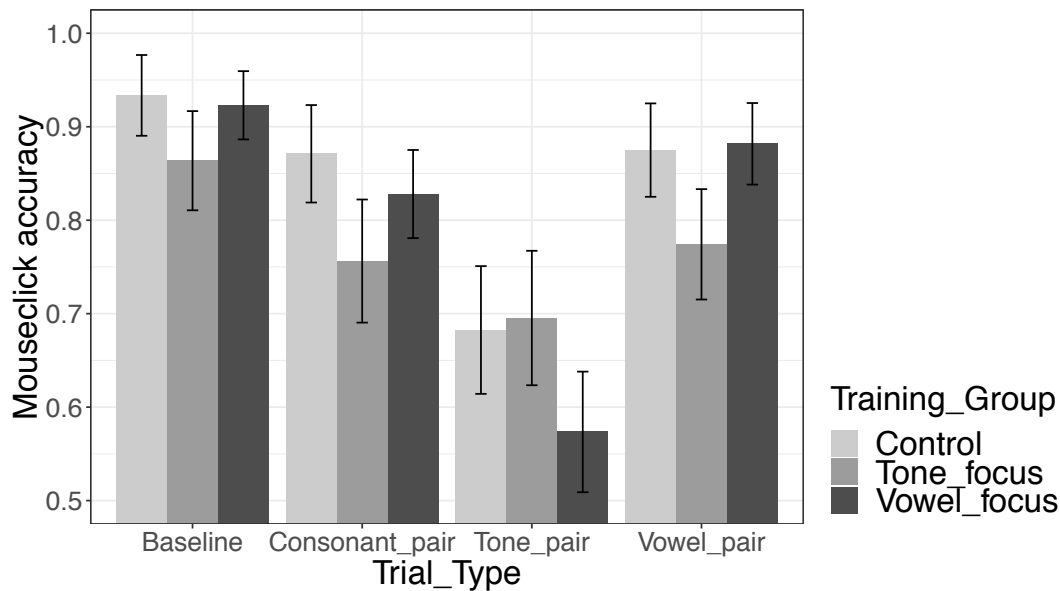
*Note.* Error bars = 95% confidence intervals based on participant means

Of critical interest, however, was whether the different training groups would perform differently on different trial types, and more specifically, whether participants trained in a given cue-focus condition would achieve higher accuracy with the trials differing only on the focused cue. Figure 18 illustrates accuracy by trial type for each group. I am especially interested in the three groups' performance on Tone-pair and Vowel-pair trials, where only a single cue differentiated the two words, and participants in the two experimental groups were trained to allocate their attentional focus to one of those two cues respectively. For Tone-pair trials, the Tone-focus group performed better than the Vowel-focus group, but did not differ from the Control group. For Vowel-pair trials, I see

the opposite pattern, with the Vowel-focus group showing higher accuracy than the Tone-focus group, but no difference from the Control group.

**Figure 18**

*Overall Accuracy by Training Group and Trial Type*



*Note.* Error bars = 95% confidence intervals calculated over individual trials

For statistical analysis, the binomial accuracy data were submitted to a series of generalized linear mixed effect models with training group, trial type and their interaction as fixed effects added one by one. The maximal random effect structures justified by the design were also attempted, and reduced if convergence problems arose (Barr et al., 2013). Model comparisons were carried out using the *anova()* function to identify the best-fitting model. In order to assess main effects, both training group and trial type were simple-coded, with Control group and Baseline trial as the reference level. Table 14 presents the output of this model. With regard to main effects of training group, the significant negative estimate for Tone-focus ( $b = -0.62, p = .005$ ) indicates overall lower accuracy in the Tone-focus compared to the Control group, while the difference between the Vowel-focus and the Control group was not significant ( $b = -0.34, p = .12$ ). Turning to main effects of trial type, all three minimal-pair trials showed significantly lower overall accuracy than baseline trials ( $p < .001$ ). The largest negative estimate for Tone-

pair ( $b = -1.89$ ) confirms that words differing by tone alone present the overall greatest difficulty for word recognition. Only one interaction effect was significant in this model.<sup>6</sup> To further explore this interaction effect as well as our predictions regarding the effects of cue-focus training, I conducted follow-up analyses on the data from each trial type separately.

**Table 14**

*Results of Generalized Linear Mixed-effects Model for Accuracy*

Formula (glmer): accuracy ~ training group * trial type + (1   participant) + (1   item), family = binomial(link="logit")					
	Fixed Effects				
	b	SE	z	p	
Intercept	1.77	0.14	12.88	< .001	***
Tone-focus	-0.62	0.22	-2.89	.005	**
Vowel-focus	-0.34	0.22	-1.57	.12	
Consonant-pair	-0.86	0.10	-9.00	< .001	***
Tone-pair	-1.89	0.09	-21.53	< .001	***
Vowel-pair	-0.65	0.10	-6.60	< .001	***
Tone-focus:Consonant-pair	-0.01	0.24	-0.05	.96	
Tone-focus:Tone-pair	-0.18	0.25	-0.70	0.48	
Tone-focus:Vowel-pair	0.94	0.22	4.36	< .001	***
Vowel-focus:Consonant-pair	-0.28	0.22	-1.24	.21	
Vowel-focus:Tone-pair	0.06	0.24	0.26	0.79	
Vowel-focus:Vowel-pair	0.26	0.26	0.99	0.32	

<sup>6</sup> Choosing Control group and Baseline trial as the reference level enables me to answer the research questions about the effectiveness of cue-focus training by comparing the experimental groups (Tone-focus and Vowel-focus) and minimal-pair trials to baselines. However, the interactions in the current model are not the only possible interactions. Visual inspection of Figure 18 indicates that significant interactions might also be found with other reference levels.

For Tone-pair trials, no significant differences were found between the Tone-focus group and the Control group ( $b = 0.08, p = .75$ ), while the Vowel-focus group showed significantly lower accuracy than the Control group ( $b = -0.55, p = .03$ ; see Table 15). For Vowel-pair trials, no significant differences were found between the Vowel-focus group and the Control group ( $b = 0.06, p = .86$ ), while the Tone-focus group showed significantly lower accuracy than the Control group ( $b = -0.90, p = .006$ ; see Table 16). These results suggest, contrary to the predictions of the assumption under investigation, that cue-focus training did not lead to more accurate learning of the focused cue, but to less accurate learning of non-focused cues. Furthermore, analyses of Baseline and Consonant-pair trials showed no significant differences between the Vowel-focus and Control groups (Baseline:  $b = -0.50, p = 0.23$ ; Consonant-pair:  $b = -0.46, p = 0.14$ ), but the Tone-focus group had significantly lower accuracy than the Control group (Baseline:  $b = -1.24, p = 0.002$ ; Consonant-pair:  $b = -0.95, p = 0.002$ ; see Tables 17 and 18). This further indicates that focusing on tone, the novel and more difficult cue for these learners, may have drawn their attention away from other, non-focused cues, with the consequence of negatively impacting their overall word learning.

**Table 15**

*Results of Generalized Linear Mixed-effects Model for Accuracy of Tone-pair Trials*

Formula (glmer): accuracy ~ training group + (1   participant) + (1   item)					
	Fixed Effects				
	b	SE	z	p	
Intercept	0.91	0.21	4.23	< .001	***
Tone-focus	0.08	0.25	0.32	.75	
Vowel-focus	-0.55	0.25	-2.23	.03	*

**Table 16***Results of Generalized Linear Mixed-effects Model for Accuracy of Vowel-pair Trials*

Formula (glmer): accuracy ~ training group + (1   participant) + (1   item)					
	Fixed Effects				
	b	SE	z	p	
Intercept	2.45	0.30	8.18	< .001	***
Tone-focus	-0.90	0.33	-2.76	.006	**
Vowel-focus	0.06	0.34	0.17	.86	

**Table 17***Results of Generalized Linear Mixed-effects Model for Accuracy of Consonant-pair Trials*

Formula (glmer): accuracy ~ training group + (1   participant) + (1   item)					
	Fixed Effects				
	b	SE	z	p	
Intercept	2.40	0.30	8.12	< .001	***
Tone-focus	-0.95	0.31	-3.09	.002	**
Vowel-focus	-0.46	0.31	-1.49	.14	



**Table 18***Results of Generalized Linear Mixed-effects Model for Accuracy of Baseline Trials*

Formula (glmer): accuracy ~ training group + (1   participant) + (1   item)					
	Fixed Effects				
	b	SE	z	p	
Intercept	3.68	0.36	10.37	< .001	***
Tone-focus	-1.24	0.41	-3.04	.002	**
Vowel-focus	-0.50	0.41	-1.20	.23	

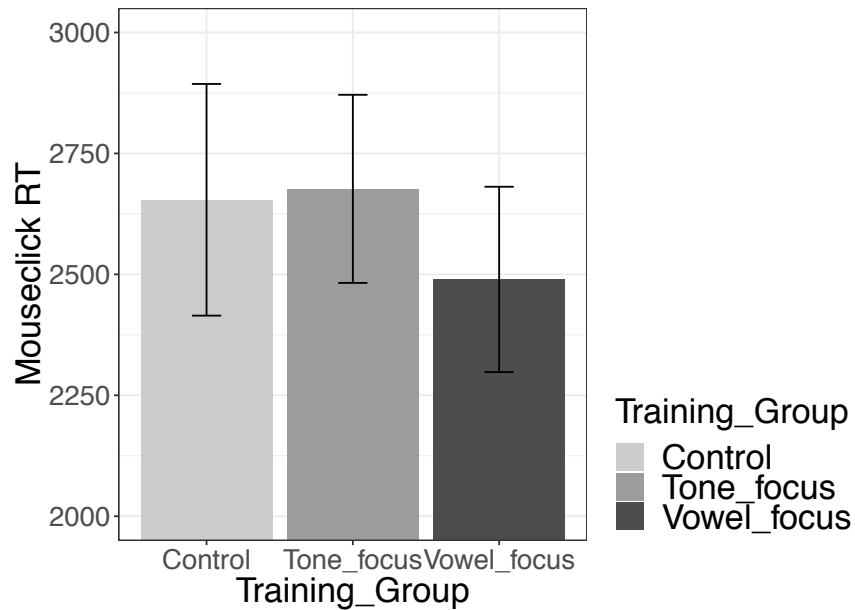
### 5.3.2.2 Mouse-click reaction time

Reaction times (RTs) were also collected to test how fast participants made their choices. The mouse-click data included information from both correct and incorrect trials, but I only included data from correct trials for RT analysis to examine how quickly participants were able to identify the correct picture after hearing the word named. 6674 out of 8099 trials (82.41%) were included. Then I excluded outlier RTs that differed from the group mean by more than three standard deviations. Data from the remaining 6550 out of 6674 trials (98.14%) were entered into the analysis.

Figure 19 presents the mean RTs by group, collapsing over different trial types. The Control group ( $M = 2654$ ,  $SD = 618$ ) and the Tone-focus group ( $M = 2677$ ,  $SD = 530$ ) had similar RTs, while the Vowel group ( $M = 2490$ ,  $SD = 522$ ) had somewhat shorter RTs.

**Figure 19**

*Overall Reaction Times for Correct Trials by Training Group*

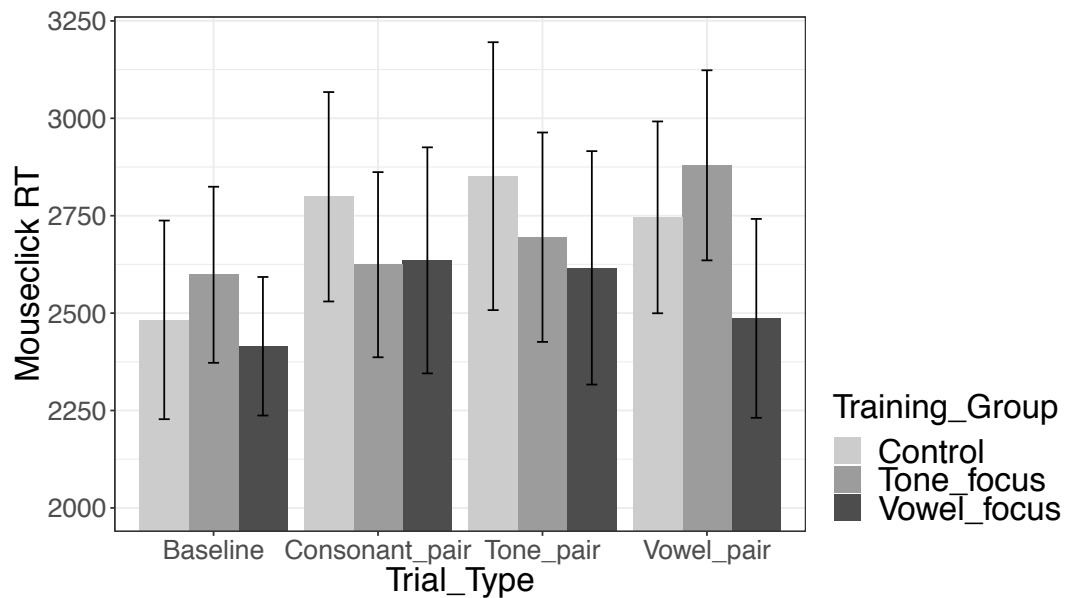


*Note.* Error bars = 95% confidence intervals by participant

I was also interested in how different training groups would perform differently on different trial types. In the case of RTs, would participants trained in a given cue-focus condition have shorter RTs than the Control group on trials differing only on the focused cue? Figure 20 illustrates the mean RTs by trial type for each group.

**Figure 20**

*Mean Reaction Times for Correct Trials by Training Group and Trial Type*



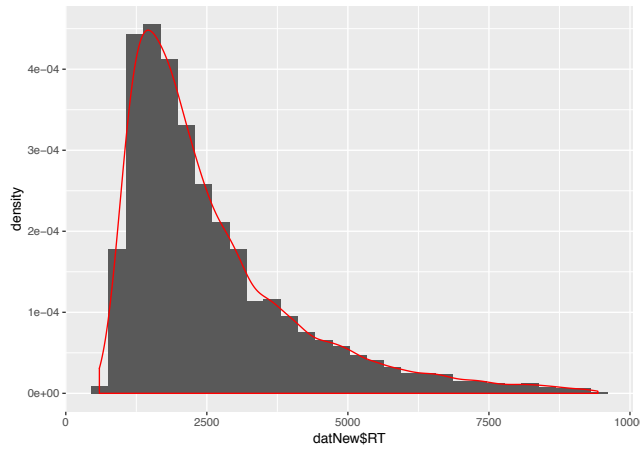
*Note.* Error bars = 95% confidence intervals by trial

I was especially interested in the participants' RTs on Tone-pair and Vowel-pair trials, where only a single cue differentiated the two words. Visual inspection indicated that for Tone-pair trials, the Tone-focus group had similar RTs to the Vowel-focus group, but shorter RTs than the Control group. For Vowel-pair trials, the Vowel-focus group had shorter RTs than the other two groups and the Tone-focus group had longer RTs than the Control group.

To investigate the training effect on different trial types, I performed a statistical analysis to check the significance of different predictors. Since the raw RT data were not normally distributed (Figure 21), I first performed a log transformation (Lo & Andrews, 2015) and checked its distribution, which was roughly normally distributed (Figure 22).

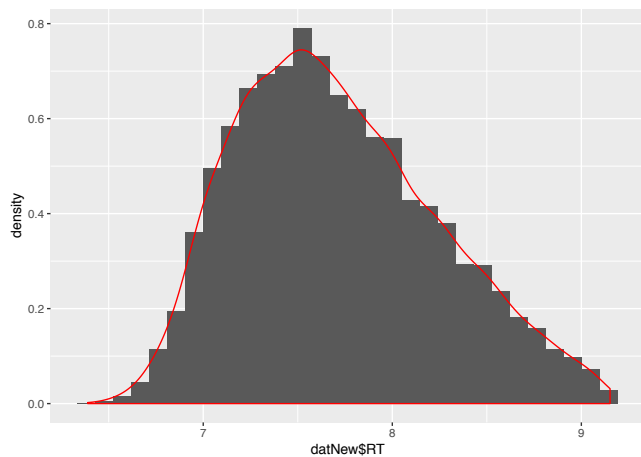
**Figure 21**

*The Frequency of Raw Reaction Time Data*



**Figure 22**

*The Frequency of Log Transformed Reaction Time Data*



Statistical analysis was conducted using linear mixed-effect models. The same modeling strategies (e.g. predictors, random effects and coding of predictors) were followed as in the analysis of accuracy. Table 19 presents the output of the final model. With regard to main effects of training group, neither the Tone-focus group ( $b = .004$ ,  $p = .95$ ) nor the Vowel-focus group ( $b = -.06$ ,  $p = .28$ ) was significantly different from the Control group in terms of RTs. Turning to main effects of trial type, all three minimal-

pair trials showed significantly longer RTs than the Baseline trials ( $p < .001$ ). Three of the interaction effects were either significant or marginally significant. To further explore these interaction effects as well as our predictions regarding the effects of cue-focus training, I conducted separate follow-up analyses on the data within each trial type.

**Table 19**

*Results of Linear Mixed-effects Model for Reaction Times on All Trials*

Formula (lmer): logRT ~ training group * trial type + (1   participant) + (1   item)					
	Fixed Effects				
	b	SE	t	p	
Intercept	7.74	.04	210.90	< .001	***
Tone-focus	.004	.06	.07	.95	
Vowel-focus	-.06	.06	-1.09	.28	
Consonant-pair	.06	.06	3.77	< .001	***
Tone-pair	.08	.02	4.50	< .001	***
Vowel-pair	.08	.02	5.24	< .001	***
Tone-focus:Consonant-pair	-.07	.04	-1.66	.10	.
Tone-focus:Tone-pair	-.02	.04	-.59	.55	
Tone-focus:Vowel-pair	-.07	.04	-1.79	.07	.
Vowel-focus:Consonant-pair	-.03	.04	-.67	.50	
Vowel-focus:Tone-pair	.01	.04	.28	.78	
Vowel-focus:Vowel-pair	-.08	.04	-2.08	.04	*

For Tone-pair trials, no significant differences were found between the Tone-focus group and the Control group ( $b = -.04, p = .63$ ) or between the Vowel-focus group and the Control group ( $b = -.04, p = .60$ ; Table 20), showing that there was no observable training effect in the RT data. For Vowel-pair trials, marginally significant differences were found between the Vowel-focus group and the Control group ( $b = -.11, p = .09$ ), but no difference was found between the Tone-focus group and the Control group ( $b = .05, p = .48$ ; Table 21). This indicated that participants trained in the Vowel-focus

group tended to make correct choices a little bit quicker when the targets differed from competitors by only one vowel. Similar analysis were done with Consonant-pair trials (Table 22) and Baseline trials (Table 23), where no significant differences were observed between the Vowel-focus group and the Control group (Consonant-pair:  $b = -.04, p = .50$ ; Baseline:  $b = -.03, p = .64$ ) or between the Tone-focus group and the Control group (Consonant-pair:  $b = -.03, p = .65$ ; Baseline:  $b = .04, p = .48$ ). In sum, the analyses within each trial type showed no significant effect of training group on RTs, though Vowel-focus training did appear to be associated with a slight improvement in performance on vowel-pair trials.

**Table 20**

*Results of Linear Mixed-effects Model for Reaction Times on Tone-pair Trials*

Formula (lmer): logRT ~ training group + (1   participant) + (1   item)					
	Fixed Effects				
	b	SE	t	p	
Intercept	7.79	.07	118.68	< .001	***
Tone-focus	-.04	.07	-.49	.63	
Vowel-focus	-.04	.08	-.53	.60	

**Table 21**

*Results of Linear Mixed-effects Model for Reaction Times on Vowel-pair Trials*

Formula (lmer): logRT ~ training group + (1   participant) + (1   item)					
	Fixed Effects				
	b	SE	t	p	
Intercept	7.79	.07	115.66	< .001	***
Tone-focus	.05	.06	.71	.48	
Vowel-focus	-.11	.06	-1.69	.09	.

**Table 22***Results of Linear Mixed-effects Model for Reaction Times on Consonant-pair Trials*

Formula (lmer): logRT ~ training group + (1   participant) + (1   item)					
	Fixed Effects				
	b	SE	t	p	
Intercept	7.78	.07	109.81	< .001	***
Tone-focus	-.03	.06	-.45	.65	
Vowel-focus	-.04	.06	-.68	.50	

**Table 23***Results of Linear Mixed-effects Model for Reaction Times on Baseline Trials*

Formula (lmer): logRT ~ training group + (1   participant) + (1   item)					
	Fixed Effects				
	b	SE	t	p	
Intercept	7.68	.06	138.76	< .001	***
Tone-focus	.04	.06	.72	.48	
Vowel-focus	-.03	.06	-.47	.64	

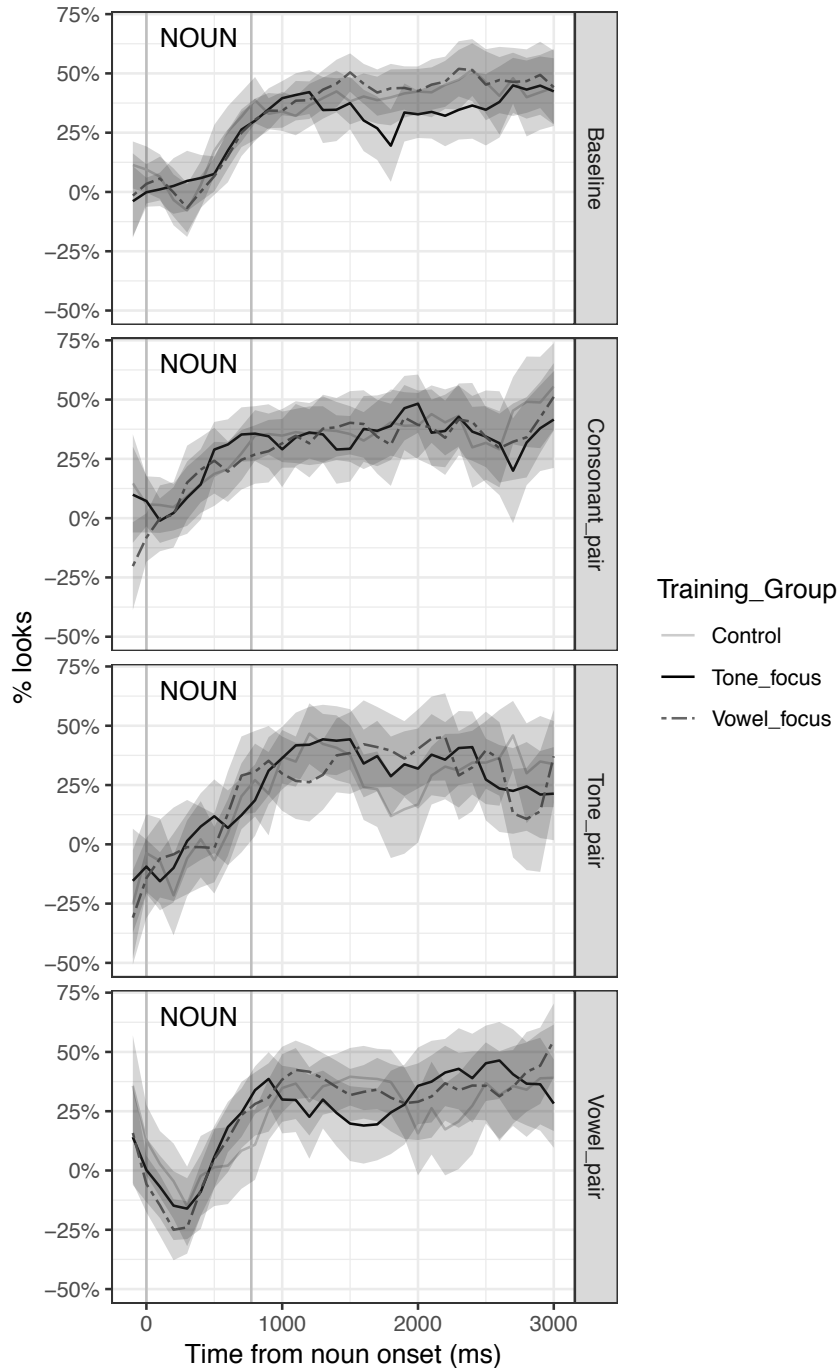
### 5.3.2.3 Eye-movement data

Because Quam and Creel's (2017) study and the findings from Experiment 2 in this dissertation both suggested that eye movements can provide additional information about participants' fine-grained temporal processing, I also investigated the time course of participants' looks to targets and competitors in the visual scene as they were listening to the noun unfold in real time. First, similar to the analysis of the RT data, I excluded trials with RTs in excess of three standard deviations above the group mean, and I only included data with correct mouse clicks. Next, I excluded all trials with a tracking ratio lower than 75%, which led to the exclusion of 4.43% (290/6550) of the remaining trials. All in all, data from 77.30% (6260/8099) trials were included for the following analysis.

Figure 23 illustrates the mean proportion of looks to the target minus the mean proportion of looks to the competitor by training group and trial type in the window of 100 ms before noun onset to 3000 ms after noun onset. Positive values thus indicate that there were more looks to the target than to the competitor, while negative values indicate that there were more looks to the competitor than to the target. Visual inspection of Figure 23 indicates no obvious differences between the training groups for any trial type.



**Figure 23.** Differences between Mean Proportion of Looks to Target and Mean Proportion of Looks to Competitor (y-axis) by Training Group and Trial Type on Correct Trials



*Note.* Window 100ms before noun onset to 3000ms after noun onset; First solid vertical line represents the noun onset and the second solid vertical line represents the noun offset. Shaded region shows 95% confidence intervals.

In order to further explore potential differences between the training groups, two sets of exploratory statistical analyses were conducted. First, I used a data-driven window from 200 ms after the noun onset to the mouse click, i.e., until the participant clicked the (correct) target in a given trial. Two hundred milliseconds (ms) after the noun onset was selected because previous studies have shown that it takes about 200 ms for participants to plan and execute their eye movements (Matin et al., 1993). This time window varied by trial. Second, I used the window from 200 ms after the noun onset to 1,100 ms after the noun onset based on Quam and Creel's (2017) study. This window was consistent throughout all trials and was expected to reveal more details about participants' eye movements at the beginning stages of processing. Within both windows, the dependent variable was TargetAdvantage, which was defined as the difference between the number of 20 ms bins spent looking at the target vs. the competitor for each trial and participant. I chose TargetAdvantage as the dependent measure because it takes looks to both targets and competitors into consideration and is more normally distributed than other measures, such as proportion of looks to the target or proportion of looks to the competitor.

#### *5.3.2.3.1 Analysis of the window from 200 ms after noun onset to mouse click*

A linear mixed-effect model with training group (Control, Tone-focus and Vowel-focus), trial type (Baseline, Consonant-pair, Tone-pair and Vowel-pair) and their interaction as fixed effects was fitted to the TargetAdvantage data. Intercepts for participant and item were added into the model as random effects. The same method of coding predictors was applied as in the analysis of the mouse-click accuracy and RT data. Table 24 presents the output from the model. The nonsignificant main effects of Tone-focus group ( $b = 2.35, p = .73$ ) and Vowel-focus ( $b = 4.66, p = .50$ ) indicate that cue-contrastive training did not influence participants' looks to targets vs. looks to competitors. However, the significant negative main effects of Consonant-pair trial ( $b = -32.80, p < .001$ ), Tone-pair trial ( $b = -38.58, p < .001$ ) and Vowel-pair trial ( $b = -37.68, p < .001$ ) show that overall participants were more likely to look at the competitor in an experimental trial than in a baseline trial. No interactions in this model reached significance, although the interaction between Tone-focus and Vowel-pair was

marginally significant, indicating that the main effect of trial type did not differ consistently by group.

**Table 24**

*Results of Linear Mixed-effects Model for Eye-movement Data between 200 ms after Noun Onset to Mouse Click*

Formula (lmer): RT ~ training group * trial type + (1   participant) + (1   item)					
	Fixed Effects				
	b	SE	t	p	
Intercept	49.35	3.65	13.54	< .001	***
Tone-focus	2.35	6.84	0.34	.73	
Vowel-focus	4.66	6.84	0.68	.50	
Consonant-pair	-32.80	2.01	-16.29	< .001	***
Tone-pair	-38.58	2.16	-17.91	< .001	***
Vowel-pair	-37.68	2.00	-18.82	< .001	***
Tone-focus:Consonant-pair	2.45	5.01	0.49	.62	
Tone-focus:Tone-pair	7.22	5.24	1.38	.17	
Tone-focus:Vowel-pair	9.71	5.01	1.94	.05	.
Vowel-focus:Consonant-pair	-5.08	4.92	-1.03	.30	
Vowel-focus:Tone-pair	4.10	5.39	0.76	.45	
Vowel-focus:Vowel-pair	0.71	4.89	0.14	.89	

### 5.3.2.3.2 Analysis between 200 ms to 1100 ms after noun onset

The same linear mixed-effect model was used to analyze eye-movement data between 200 ms to 1100 ms after noun onset, a window used in Quam and Creel’s (2017) analysis. Table 25 presents the output from the model. All the results are similar to the analysis of the window from 200 ms after the noun onset to the mouse click. The only difference is that the main effect of the Consonant-pair trial became nonsignificant with this smaller window size ( $b = -1.00$ ,  $p = .28$ ), showing no significant difference between Baseline trials and Consonant-pair trials. No interaction effects in this model were significant.

**Table 25**

*Results of Linear Mixed-effects Model for Eye-movement Data between 200 ms and 1100 ms after Noun Onset*

Formula (lmer): RT ~ training group * trial type + (1   participant) + (1   item)						
	Fixed Effects					
	b	SE	t	p		
Intercept	2.47	2.10	1.17	.26		
Tone-focus	-0.12	1.02	-0.12	.91		
Vowel-focus	-1.24	1.02	-1.21	.23		
Consonant-pair	-1.00	.93	-1.08	.28		
Tone-pair	-5.67	.99	-5.72	< .001	***	
Vowel-pair	-4.10	0.92	-4.46	< .001	***	
Tone-focus:Consonant-pair	1.63	2.30	0.71	.48		
Tone-focus:Tone-pair	1.07	2.41	0.44	.66		
Tone-focus:Vowel-pair	2.40	2.30	1.04	.30		
Vowel-focus:Consonant-pair	1.17	2.27	0.52	.61		
Vowel-focus:Tone-pair	2.80	2.49	1.12	.26		
Vowel-focus:Vowel-pair	1.77	2.53	0.78	.43		

In all, both analyses show that cue-contrastive training did not influence participants' looks to the target vs. the competitor and that participants tended to look more at competitors when the word pairs differed only by tone or vowel.

#### **5.4 Discussion**

The goal of this study was to examine the effectiveness of cue-focus training in word learning. More specifically, I wanted to test an assumption implicit in previous research and pedagogical practice on the learning of lexical tone, namely that manipulation of learners' attentional focus to the contrast of tonal cues would benefit their learning of words with lexical tones. Since many factors are likely to influence learning outcomes in more naturalistic learning settings, I aimed to minimize the role of previous language experience by training English speakers in three different training groups to learn novel words in a laboratory setting. Learning outcomes were measured by accuracy rates, RTs and eye-fixations in a forced-choice spoken word recognition task following training. If tone-focus training is effective, I expected to see higher accuracy, shorter RTs and more looks to targets vs. competitors in the Tone-focus group than in the Control group, especially on trials where tone was the only cue differentiating the two words (Tone-pair trials). If cue-focus training is effective more generally, I would expect to see an analogous pattern with participants trained in the Vowel-focus group outperforming the Control group, especially in trials where vowel quality was the only cue differentiating the two words (Vowel-pair trials).

Unexpectedly, I found that participants trained in the Tone-focus group showed overall lower accuracy in word recognition than those in the Control and Vowel-focus groups. More importantly, even in Tone-pair trials, where tone was the only cue differentiating the two words, the Tone-focus group did not achieve higher accuracy than the Control group, indicating tone-focus training was not effective. Critically, the Vowel-focus group performed significantly lower than the Control group on Tone-pair trials, suggesting Vowel-focus training hurt learning of tone. This was further supported by the fact that this pattern was mirrored with Vowel-pair trials, where no difference was

observed between the Control and Vowel-focus groups, but the Tone-focus group showed significantly lower accuracy than the Control group.

No significant main effect of training group was found with RT or eye-movement data, suggesting that there was no training group effect on those two measurements. However, I found that participants took significantly longer to make correct choices in all three minimal cue contrastive trials (Tone-pair, Vowel-pair and Consonant pair) than in the Baseline trials. This finding was understandable since participants were able to use more than one cue to differentiate the target from the competitor in Baseline trials, while in minimal cue-contrastive trials, only one cue was available.

In sum, and contrary to my initial predictions, the findings from this study suggested that cue-focus training did not improve learning of the focused cue, but instead hurt learning of non-focused cues. Since tonal words are composed of both tones and segments, training learners to focus on a single cue such as tone might make it harder for them to learn a word as a unit. These results are consistent with Zhao et al.'s (2011) hypothesis that “the recognition of Chinese monosyllabic words might rely more on global similarity of the whole syllable structure or syllable-based holistic processing rather than phonemic segment-based processing” (p. 1761). If learners’ attention is fragmented by considering each cue separately during learning, this might prevent them from processing words holistically during later word recognition.

I chose to focus on spoken word recognition because tone is as a necessary component for distinguishing between words in Mandarin. Bearing this in mind, any teaching or learning methods that do not work towards the successful use of lexical tone in word learning would be misguided, and methods that are not able to increase the accurate processing of tone in the context word recognition would ultimately be ineffective. The current study shows that cue-focus training is not always effective. In this laboratory-based auditory novel word learning experiment, neither the Tone-focus group nor the Vowel-focus group showed higher accuracy than the Control group in overall word recognition. However, cue-focus training did reallocate participants’ attentional focus, albeit to the detriment of non-focused cues rather than to the benefit of the focused cue.

The results from this study, however, do not necessarily conflict with findings from previous research indicating that certain versions of tone-focus training (e.g. the contour + pinyin condition in Liu et al., 2011) might be more effective than others (e.g. the contour only condition from the same study) at reducing errors in tone identification. Yet my study highlights the importance of considering the difference between tone identification in isolated syllables and in meaning-bearing words, and my findings indicate that it is helpful to include a control group receiving no cue-focus training for comparison.

The current findings also do not directly conflict with existing learning theories. As the Noticing hypothesis (Schmidt, 1990) states, noticing is necessary, but it is not necessarily sufficient for intake. Consistent with the Automatic Selective Perception model (Strange, 2011), my L1-English participants paid less attention to tone, having their lowest accuracy rates in the Tone-pair trials across the three training groups. The Competition Model (MacWhinney, 2005, 2012) suggests that presenting the contrastive form can increase the relative strength of the cue in acquisition. However, I did not observe a benefit for word learning by presenting the contrastiveness of a cue in the current experiment. This might be owing to how I measured the learning outcome. Instead of focusing solely on attention to form, as in previous studies, the word learning and recognition tasks in this experiment required use of both form and meaning. The findings suggest that in the context of word learning, where both tones and segments must be taken into consideration, trying to increase the relative cue-strength of tone by highlighting the contrastiveness of this cue in training is not effective, and may ultimately hurt syllable-based holistic processing and learning of words in tonal languages.

## **Chapter 6. General discussion, implications and limitations**

### **6.1 Summary of findings**

This dissertation investigates L2 learners' processing and use of tones in order to gain a broader perspective and understanding of the challenges in the L2 acquisition of Mandarin tones by native speakers of English. To this end, three experiments were conducted to address three broad research questions about the L2 acquisition of lexical tone at the levels of speech perception (Experiment 1), lexical processing (Experiment 2), and word learning (Experiment 3).

Experiment 1 (Chapter 3) investigates how L2 learners with various levels of Mandarin proficiency perceive Mandarin tone-pair continua compared to native listeners and naïve listeners with all six possible tone pairs in Mandarin. The results corroborate earlier studies in finding that native listeners perceive tones more categorically than naïve listeners do. The current study also extends these findings to all tone pairs in Mandarin. Participants perceived tone pairs with different degrees of pitch similarity in different ways; across all three groups, tone pairs with similar pitch contours (e.g., T2-T3) were perceived less categorically than tone pairs with more distinct pitch contours (e.g., T1-T3). I also found that the L2 learners had a pattern of results that was intermediate between the other two groups on the identification task, which suggests that they were in the process of developing mental representations of tone categories. In addition, I found that L2 proficiency was positively correlated with identification slopes, which provides further, novel evidence that CP of tone increases with higher L2 proficiency.

I hypothesized that if the L2 learners really did perceive tone similarly to the native listeners, they might also pattern between the native listeners and the naïve listeners on the discrimination task. However, that was not the pattern of results I observed for that task. Overall, native listeners had lower accuracy across all tone pairs than naïve listeners. This was due to native listeners' decreasing accuracy close to the two endpoints, whereas both the L2 learners and the naïve listeners consistently showed high discrimination accuracy across all tone pairs, regardless of whether the steps crossed potential category boundaries or not. This indicates that the L2 learners and naïve listeners had similar underlying processing mechanisms. Furthermore, L2 proficiency was positively



correlated with discrimination accuracy, suggesting that the L2 learners with higher sensitivity to pitch differences also tended to have higher L2 proficiency.

However, the results from the discrimination task cannot be used as evidence that L2 learners perceived lexical tone categorically because the L2 learners performed similarly to the naïve listeners. Following Liberman et al.'s (1957) design, Experiment 1—like most previous studies on the CP of phonemes—tested participants on both identification and discrimination. This design is originally devised with a stringent criterion for CP in mind, which states that a listener only discriminate stimuli to the extent that he identify them as different phonemes (Liberman et al., 1957). In other words, the same processing mechanism is behind identification and discrimination. However, Gerrits and Schouten (2004) have presented evidence against this assumption in a series of experiments. They found that discrimination tasks are more influenced by bottom-up processing, such as perceptual sensitivity to acoustic signals, than identification tasks are. Since access to mental representations of categories is necessary to demonstrate the presence of CP, the results from Experiment 1 suggest that identification tasks might be a better means of examining this phenomenon than discrimination tasks, at least for L2 learners. Besides this methodological finding, Experiment 1 also provides the first direct evidence that L2 Mandarin proficiency correlates with the robustness of the developing mental tone categories in an L2 as measured by the steepness of the learners' identification slopes. However, access to these categories during tone processing was still not highly automatic for the L2 learners tested in this study, as suggested by the similar discrimination patterns observed for the L2 learners and the naïve listeners. This low level of automaticity might also explain why the L2 learners showed evidence of difficult processing tone at the lexical level in Experiment 2.

Experiment 2 (Chapter 4) examines the relation between learners' difficulties with acquiring lexical tones at the phonological level in L2 speech perception and the challenges they encounter with using lexical tones at the level of lexical access during L2 word recognition. Drawing on evidence from CP and real-time spoken word recognition, I found that English-speaking L2 learners of Mandarin, even those with considerable L2 experience, differed from native Mandarin speakers in both the extent to which they perceived tone categorically as well as in their ability to use tonal cues to distinguish

between words in real-time listening comprehension. At the same time, I observed substantial variability among L2 learners' performance on both tasks, with at least some of this variability being due to more experienced learners showing patterns of performance more similar to those observed in the L1 group. Critically, Experiment 2 provides the first direct evidence showing that the ability to perceive tone categorically is related to the weighting of tonal cues during spoken word recognition among adult L2 learners. This finding supports Wong and Perrachione's (2007) conclusion regarding the importance of the continuity between phonetic, phonological, and lexical abilities in the learning of tone. A better understanding of the links between learners' developing abilities across linguistic domains, often studied in isolation from each other in different research subfields, is essential not only for theoretical models of L2 development and processing, but also for more applied purposes such as curriculum design and instruction in Mandarin-as-a-foreign-language contexts, where the acquisition of tone remains a topic of central concern.

The first two experiments in this dissertation show that there are differences between L2 learners and native speakers of Mandarin in the processing of tone at both the phonological and lexical levels, which might be related to the persistent difficulties that L2 learners have with lexical tone learning. Experiment 3 (Chapter 5) was designed to investigate how we can improve L2 learning of tones by testing the effectiveness of cue-focus training, a technique commonly used by educators for teaching phonological contrasts such as tone. Cue-focus training is used to draw learners' attentional focus to tone with minimally contrastive pitch features (e.g. level, rising, falling) presented to improve learning of tone (Godfroid et al., 2017; Lin, 1985; Liu et al, 2011; Tsai, 2011). In Experiment 3, native English speakers who had no prior experience with tonal languages were trained to learn artificial words with tone in a controlled laboratory setting. Results show that cue-focus training did not facilitate the learning of the focused cue but instead inhibited the learning of non-focused cues, which in turn led to a decrease in overall word learning success. Thus, this study has provided experimental evidence indicating that the commonly held assumption that cue-focus training is effective in vocabulary learning does not always hold. The results also suggest that vocabulary

learning in a tonal languages may be better supported through syllable-based holistic training than through directing attentional focus to a specific phonemic cue.

The combined findings from the three experiments presented in Chapters 3, 4, and 5 show that L2 learners process tone differently from native listeners at both the phonological and lexical levels. The difference between these two groups is closely related to the degree of automaticity in (a) their access to mental representations of lexical tone and (b) their processing of tonal information along with segmental information. This difference can be explained well by the ASP model (Strange, 2011), which characterizes how native listeners and L2 learners identify reliable cues for language processing. According to the ASP model, one of the key differences between native listeners and L2 learners is their language experience, which influences how they form mental representations of different phonemes in their long-term memory and shapes their selective perception routines (SPRs, Strange, 2011) in procedural memory.

Unlike native speakers of Mandarin, who have lifelong experience with lexical tone, most L2 learners with non-tonal L1s are still in the process of developing their mental representations of tone categories. These developing mental representations of tone categories stored in their long-term memory enable L2 learners to perceive tone more categorically than naïve listeners. In a non-tonal language, such as English, fluctuations in pitch do not change the lexical meanings of words. Thus, listeners do not need to register those pitch differences in their long-term memory for word recognition. In order to learn a tone language like Mandarin, though, L2 learners first have to realize that lexical tone exists; then they need to gradually build mental representations of tones in their long-term memory for quicker and more effective processing. As they progress from naïve listeners to advanced L2 learners, their mental representations of lexical tone are under constant development, and with increasing exposure to the target language they may eventually be able to approximate the mental representations that native listeners have.

However, the access that L2 learners have to their mental representations of lexical tone is not as automatic as it is for native listeners. In order to process tonal information quickly and effectively in daily life, native listeners have developed highly automatic routines for processing pitch information during tone identification. Even in

discrimination tasks, native listeners cannot inhibit their access to mental representations of lexical tone and therefore show reduced sensitivity to within-category differences. In the words of the ASP model (Strange, 2011), between-category pitch differences are more perceptually salient than within-category pitch differences for native listeners, and selective attention is therefore automatically allocated to between-category differences.

In contrast, tone perception is not as automatic in L2 learners. Though they have the ability of perceive pitch more categorically than naïve listeners in the phonological mode, L2 learners can still perform like naïve listeners when it comes to processing pitch in the phonetic mode, at which times access to mental representations of lexical tones is not necessary, such as when it is necessary to discriminate between several consecutive tokens with different pitch contours. The fundamental difference between the phonological mode and the phonetic mode is whether listeners need to optimize their processing strategy to deal with phonological contrasts (between-category pitch differences) or within-language phonetic variations (within-category pitch differences). Though naïve listeners have often been used as a control group in L2 studies, they can potentially be regarded as learners at the initial stage of L2 acquisition. The ASP model (Strange, 2011) hypothesizes that a phonetic mode of speech perception is used at the beginning stage of L2 learning to process unfamiliar sounds. In this mode, listeners use a fine-grained analysis of acoustic details to process context-dependent phonetic variants, which might have counterparts in their native languages. Thus, the phonetic mode is generally less robust and automatic than the phonological mode and requires a great deal of cognitive resources, especially memory load. In contrast, native listeners are more likely to process in a phonological mode, where processing is more automatic and attuned to the acoustic-phonetic information that is needed for CP.

The lower degree of automaticity in L2 learners compared to native listeners could also explain differences observed between these two groups in tone processing at the lexical level. The L2 learners' longer RTs, lower accuracy rates and higher rates of eye fixations to competitors that differ from targets by tone in Experiment 2 were all indicative of a lower degree of automaticity in their use of tonal information at the lexical level. These results are consistent with those of Pelzl et al. (2019, 2020), who found that L2 learners have more difficulty using tonal information than segmental information.

According to Wong and Perrachione (2007) and Cooper and Wang (2013)'s claims, there are phonetic-phonological-lexical continuities for adult non-native learning and more basic, low-level speech perception abilities mediate more complex, high-level word processing. Thus, the lower degree of automaticity in using tonal information at the lexical level that the L2 learners displayed in the current study might also be related to the lower levels of automaticity they displayed at the phonological level. As discussed above, access to mental representations of lexical tone is not as automatic for L2 learners as it is for native listeners. Though recognizing Mandarin words requires listeners to use both segmental and tonal information, tones might not be as readily available as segments at the lexical level due to the fact that L2 learners process tones in a less categorical and less automatic fashion.

Also, L1-English L2 learners of Mandarin have already developed highly automatic selective perception routines in English for processing segmental information during word recognition. Processing tonal information along with segmental information at the lexical level thus requires learners to change their perception routines for processing words in Mandarin, which could be a difficult task. As discussed in the literature review, a variety of teaching methods, such as visualization of tone contours (Liu et al., 2011), using color or number coding (Godfroid et al., 2017), music (Lin, 1985), hand gestures or other body movements (Tsai, 2011), and instruction focusing on pitch direction and height (He et al., 2016) have all been used to draw learners' attentional focus to tone as a minimally contrastive feature. In the ASP model (Strange, 2011), attentional focus is closely related to the notion of "noticing" in the Noticing Hypothesis (Schmidt, 1990), which refers to listeners' conscious attunement to information with clear goals and purposes. Because Schmidt (1990) claims that "noticing is necessary for intake" (p. 141), the goal of allocating attentional focus is to intake the contrastive feature that is critical for learning words in the L2. However, focus on tonal contrasts alone can draw learners' attentional focus away from segmental contrasts, which is actually detrimental to their ability to learn tones and segments as a single entity, as shown in Experiment 3. Thus, it appears that a more ideal way to teach lexical tone may not be to teach tones in isolation, but instead to teach them in the context of words, which will be discussed more in the following part on the implications for teaching (6.2).

All in all, this project has shown that L2 learners with various levels of Mandarin listening proficiency process lexical tone differently from native listeners at both the phonological and lexical levels. This difference is partly due to L2 learners' lower automaticity in using mental representations of lexical tone for perceiving tones categorically and processing tonal information along with segmental information for lexical access.

## **6.2 Implications for teaching**

This dissertation advances our understanding of how L2 learners perceive, process, and learn lexical tone, an indispensable dimension of tonal languages, through a set of experiments targeting multiple steps in tone acquisition, thus contributing to a better understanding of phonetic-phonological-lexical continuities in L2 word learning. The three experiments were designed with potential contributions to L2 instruction and vocabulary learning in mind to provide first-hand evidence to inform evidence-based L2 instruction and curricular materials.

The results from Experiment 1 suggest that L2 learners are still developing their mental representations of tone categories and that they do not access that information automatically, regardless of L2 proficiency. Crucially, Experiment 2 shows that this ability of tone perception at the phonological level correlates with the difficulty of using tonal information along with segmental information by L2 learners during lexical access, and Experiment 3 suggests that the popular tone-focus training method might not always be effective in vocabulary learning. The teaching implications from these three experiments are twofold: (a) L2 instructors should help learners to continue developing their mental representations of tone categories at all stages of learning, not just at the beginning level, and (b) they should focus not only on explicit instruction on the shapes of the different tonal contours, but also on improving learners' automatic processing of tones along with segments. In the following paragraphs, I will elaborate on this point further by providing an example.

Though the primary goal of this dissertation was to provide research-based evidence of L2 learning difficulties rather than to develop new teaching methods, it might still be

helpful to provide an example of how existing methods could be modified in light of the current findings. One of the most widely studied and used training methods for developing a capacity for efficient discrimination of different L2 phonemes is High Variability Phonetic Training (HVPT; Li et al., 2016; Lively et al., 1993; Logan et al., 1991; Perrachione et al., 2011; Wang et al., 1999; Wiener et al. (2020). In a groundbreaking study, Logan et al. (1991) tested whether training involving speech sounds with high variability (e.g. pronounced by different talkers) might help learners form robust categories of L2 phonemes by ruling out irrelevant acoustic differences. L1-Japanese speakers learning English were trained on English /l/ and /r/ minimal pairs using recorded natural production. All participants took the same identification task (e.g., Is this word *rock* or *lock*?) as a pretest and half of them were trained with recordings from five talkers, while the other half were trained with recordings from a single talker. In the posttest, all participants took a similar identification task, but with recordings from a new talker. Results showed greater improvement on identifying English /l/ and /r/ in the multi-talker group than in the control group, indicating that there is a benefit of using HVPT in L2 learning.

A similar research design was adopted by Wang et al. (1999) to test the efficacy of using HVPT to learn non-native suprasegmental contrasts, i.e., Mandarin lexical tone. L1-English speakers learning Mandarin were trained in eight sessions during a two-week course. The results also showed large improvements in the identification of tones for the multi-talker group, and these improvements were retained six months later. Moreover, the improvements gained from HVPT were generalized to new stimuli and new talkers.

Since Logan et al.'s (1991) research, a series of studies have been conducted to investigate how different factors can influence the effectiveness of HVPT. Perrachione et al. (2011) pointed out that the one-size-fits-all approach to HVPT might not be appropriate, since they found an interaction between individual perceptual abilities and the training method used. Their study (Perrachione et al., 2011) found that only learners with high perceptual abilities actually benefit from HVPT and that it hurts speakers with low perceptual abilities. Other factors that might influence the efficacy of HVPT include the level of variability (e.g., number of talkers), training set size (e.g., number of items), and procedure (e.g., number of repetitions) and explicit instruction (Wiener et al., 2020).

In light of the current study and previous studies, a computer-mediated and student-oriented HVPT approach to tone learning with real words and explicit instruction might help learners of Mandarin build more robust mental representations of tone categories as well as more automatic processing of tonal cues during lexical access. The training phase could use naturally produced real words from different talkers along with pictures. It might be helpful to give students the opportunity to choose how many items, talkers, and repetitions they want to hear in one session. A testing phase might help students to evaluate their learning outcomes, and feedback and explanation would be useful to further enhance learning. Recordings from new talkers could be added to test the extent to which students are able to generalize what they have learned to other voices, reminders compatible with Google Calendar could be sent to remind learners to practice, and a delayed posttest could be given to examine students' retention of the things they had learned. Such training could be assigned to students outside of the classroom at all learning stages, and it would be helpful if the program could send information about the students' progress to their teachers. These are just a few possibilities for how the current findings could be applied to actual teaching. Of course, many additional details need to be discussed before such a system could be implemented, which are beyond the scope of this dissertation.

### **6.3 Limitations and Directions for Future Work**

Though all three experiments in this dissertation were carefully designed, there are still some limitations associated with them. In Experiment 1, I used identification and discrimination tasks to test how listeners with different language experience perceived pitch variations. Though this is standard in experimental research, the tasks themselves are artificial and listeners are barely ever asked to identify pitch contours or discriminate between several consecutive pitch tokens in daily communication. Researchers who are interested in how L2 learners perceive lexical tone categorically at the lexical level could adapt those two tasks by asking L2 learners and native listeners to identify and discriminate words minimally differing in tone, rather than isolated syllables. However, we should note that asking listeners to identify or discriminate words rather than isolated



syllables might require them to process in a phonological (as opposed to a phonetic) mode, where listeners have to access their mental tone categories for recognizing words. Researchers could also modify the identification and discrimination tasks to test how different factors, such as number of speakers, contexts or background noise, might influence the categorical perception of lexical tone. A similar limitation related to the artificiality of the task also exists in Experiment 2, which required listeners to choose one of three pictures according to the word they heard.

In Experiment 3, I used an artificial language learning (ALL) task in a controlled laboratory setting to study the effectiveness of cue-focus teaching of lexical tone in words. Despite the benefits ALL tasks (see Section 2.2.3.3), they also have several limitations. First, the research findings cannot be directly generalized to the processing of natural languages, which are far more complex than any artificial languages or stimuli used in a laboratory. Second, due to time restrictions, the experiment was only conducted in one session and could not examine long-term learning outcomes. Third, even with the most carefully created artificial languages, not all confounds can be eliminated. Researchers who are interested in the long-term effect of cue-focus training could adapt the research design of Experiment 3 to multiple-session learning studies. Experiment 3 could be considered as a simple beginning point for a series of follow-up studies on cue-focus training by manipulating various factors, such as using real languages (e.g. Mandarin, Thai), testing participants with various language backgrounds (with or without tonal language background) or testing in real classrooms.

There were also limitations associated with the participant groups selected for this study. For example, the participants in all three experiments were college students, who might not provide an accurate representation of the general population. I recruited my participants from university communities for reasons of accessibility and convenience. L2 learners of Mandarin, for example, were very difficult to recruit outside of universities.

With limited resources of space, time and participants, the laboratory experiments used in this dissertation were carefully designed to investigate how L2 learners perceive, process and learn Mandarin lexical tone. However, when one wants to apply those findings to real classroom teaching, limitations related to stimuli, participants, procedures,

and design should all be taken into consideration and the findings of the study should be carefully interpreted within that experimental context.

For future research, I plan to explore how different training methods could help L2 learners build more categorical perception of lexical tone and better learning of words with tones. For example, instead of training listeners with words produced by one speaker in Experiment 3, I plan to use words produced by multiple speakers and test their learning outcome. I also plan to use identification and discrimination tasks as follow-up study to test how naïve listeners learn lexical tone with high variability phonetic training (HVPT).

All three experiments in this dissertation focus on the comprehension of lexical tone. However, the production of tone is also critical and could be a more challenging part of vocabulary learning in tonal languages. More specifically, future studies could investigate whether and how categorical perception is related to learners' production of tone, and how different training methods could influence learners' production of words with tone.

#### **6.4 Concluding remarks**

This dissertation is designed to broaden the investigation of SLA by looking at an L2 that is not an Indo-European language, a domain still underrepresented in SLA research but critical in light of the large numbers of people worldwide who learn and use Chinese as an L2. This dissertation simultaneously contributes to (a) research in SLA by bringing conceptual insights from psycholinguistic research on tones in native processing to the study of lexical tone in an L2 and (b) research in the fields of speech perception and psycholinguistics by probing the generalizability of findings about human language processing to L2 learners. Methodologically, three paradigms that are well-established in psycholinguistic research—identification and discrimination tasks to assess CP, visual-world eye-tracking to measure incremental use of information during lexical processing, and novel word learning—were adopted to test L2 learners' performance in comparison to native listeners and naïve listeners, thus allowing for comparisons to be made with previous work that has used the same methods with different populations. Future studies should test how well the current findings generalize to the acquisition of other tonal

languages (e.g., Cantonese, Thai) by learners with different L1s (e.g., French speakers, Thai speakers); future investigators could either use the same research design or broaden the investigation of L2 lexical tone learning by employing different research methods (e.g., ERPs).

## APPENDICES





### Appendix A1

*Mandarin listening proficiency test with Chinese instruction for Exp.1*




#### 第一部分

听录音，如果图片符合你所听到的，请打勾。如果不符合，请打叉。

#### 第 1-10 题

例如：		✓
		✗
1.		
2.		



3.		
4.		
5.		
6.		
7.		





8.		
9.		
10.		

第二部

听录音, 填写你所听到的符合对话内容的图片编码。

第 11-15 题

A		B	
---	---	---	---

C		D	
E		F	

例如: 男: 你喜欢什么运动?






女: 我最喜欢踢足球。

(D)

- 11.
- 12.
- 13.
- 14.
- 15.

- ( )
- ( )
- ( )
- ( )
- ( )

第 16-20 题

A		B	
C		D	
E			

16.

( )

17.

( )

18.

( )

19.

( )

20.

( )







## Appendix A2







*Mandarin listening proficiency test with English instruction for Exp.1*

### Part 1

**Instruction: Listen to the recording. As in the example, put a check mark next to the picture if it is consistent with what you heard. Put a cross next to the picture if it is inconsistent with what you heard.**

**No. 1-10**

Example:		✓
		✗
1.		
2.		

3.		
4.		
5.		
6.		
7.		
8.		



9.		
10.		

**Part two**

**Instruction:** Listen to the recording of ten conversations. As illustrated in the example, write down the letter of the picture that best matches the conversation you heard for each item.

**No. 11-15**

A		B	
C		D	

E		F	
---	---	---	---

Example: Man: Which sports do you like?

Woman: I like football the most.

(D)

11.

( )

12.

( )

13.

( )





14.

( )

15.

( )

**No. 16-20**

A		B	
C		D	

E			
---	---	--	--

- 16. ( )
- 17. ( )
- 18. ( )
- 19. ( )
- 20. ( )

## Appendix B

### Mandarin listening proficiency test for Exp. 2

#### 第 1-10 题:判断对错。

例如:

我想去办个信用卡,今天下午你有时间吗?陪我去一趟银行?

★ 他打算下午去银行。(√)

现在我很少看电视,其中一个原因是,广告太多了,不管什么时间,也不管什么节目,只要你打开电视,总能看到那么多的广告,浪费我的时间。

★ 他喜欢看电视广告。(×)

1. ★ 飞机还没起飞。 ( )
2. ★ 不饿就不要吃早饭。 ( )
3. ★ 经理发现了小王的一些缺点。 ( )
4. ★ 女朋友听过这个笑话。 ( )
5. ★ 他没有翻译第二部分。 ( )
6. ★ 服务员的京剧唱得很好。 ( )
7. ★ 王老师现在是教授了。 ( )
8. ★ 他想买个大房子。 ( )

9.★ 他在理发店。 ( )

10.★ 这个咖啡馆儿很热闹。 ( )

**第 11-25 题:请选出正确答案。**

例如:

女:该加油了,去机场的路上有加油站吗?

男:有,你放心吧。

问:男的主要是什么意思?

A 去机场                  B 快到了                  C 油是满的                  D 有加油站 ✓

11. A 没纸了                  B 男的没发                  C 打印机坏了 D 传真机坏了

12. A 将来                  B 理想                  C 小说                  D 职业

13. A 办签证                  B 去学校                  C 打网球                  D 打羽毛球

14. A 不想出国                  B 换个箱子                  C 不符合规定 D 早点儿回来

15. A 变胖了                  B 很难受                  C 正在减肥                  D 工作很辛苦

16. A 是研究生                  B 参加工作了 C 已经毕业了 D 在准备考试

17. A 打扫                  B 等人                  C 爬山                  D 购物

18. A 幽默                  B 很难过                  C 很粗心                  D 没有耐心

19. A 很酸            B 很甜            C 很咸            D 很辣
20. A 他们输了            B 他们赢了    C 他们放弃了    D 他们很愉快
21. A 学钢琴            B 去旅游            C 做生意            D 锻炼身体
22. A 肚子疼            B 感冒了            C 觉得热            D 穿得太少
23. A 周末            B 下周            C 两周后            D 下个月
24. A 医生            B 导游            C 卖家具的    D 开出租车的
25. A 我不会            B 马上来            C 没法解释    D 解决不了



## Appendix C

**Table 26.** *Experimental Stimuli and English Gloss in Parentheses*

Stimuli set	Target	Segmental competitor	Rhyme competitor	Vowel competitor	Distractor
1	Cha1 (fork) [2.33; 5]	Cha2 (tea) [3.10; 1]	Sha1 (sand) [2.87; 5]	Fa4 (hair) [3.80; 3]	Bi3 (pen) [3.52; 3]
2	Cheng2 (orange) [1.99; 6]	Cheng4 (scale) [1.64; 6]	Sheng2 (rope) [2.56; 5]	Deng4 (stool) [1.79; 7]	Guo1 (pot) [2.54; 5]
3	Chi3 (ruler) [2.79; 5]	Chi4 (wing) [1.74; 5]	Zhi3 (finger) [3.76; 5]	Shi1 (lion) [2.22; 4]	Mao4 (hat) [2.66; 3]
4	Dao3 (island) [3.47; 5]	Dao1 (knife) [3.44; 4]	Cao3 (grass) [2.87; 3]	Pao4 (cannon) [2.43; 5]	Jian4 (arrow) [3.02; 7]
5	Di2 (flute) [2.07; 7]	Di4 (floor) [4.26; 3]	Bi2 (nose) [2.59; 3]	Ji1 (chicken) [3.47; 2]	Gua1 (melon) [2.48; 2]
6	Fang2 (house) [3.38; 2]	Fang1 (square) [3.23; 5]	Tang2 (candy) [3.18; 3]	Tang1 (soup) [3.00; 4]	Huo3 (fire) [3.55; 4]
7	Gou3 (dog) [4.07; 1]	Gou1 (hook) [2.43; 6]	Shou3 (hand) [4.18; 3]	Dou4 (bean) [2.61; 5]	Qiu2 (ball) [3.85; 2]
8	Gu3 (bone) [2.89; 5]	Gu1 (mushroom) [1.08; 7]	Shu3 (mouse) [2.74; 5]	Ku4 (pants) [2.76; 3]	Xiang4 (elephant) [3.70; 5]
9	Jing1	Jing3	Bing1	Ting2	He2

	(whale)	(well)	(ice)	(pavilion)	(river)
	[1.95; 7]	[2.68; 6]	[3.20; 6]	[1.94; 5]	[3.06; 3]
10	Shao2	Shao4	Tao2	Bao1	Xie2
	(spoon)	(whistle)	(peach)	(bag)	(shoes)
	[1.97; 5]	[1.91; 6]	[2.50; 5]	[3.57; 3]	[3.37; 3]
11	Tu4	Tu2	Shu4	Shu1	Nao3
	(rabbit)	(picture)	(tree)	(book)	(brain)
	[2.60; 5]	[3.06; 3]	[3.33; 3]	[3.85; 1]	[3.26; 5]
12	Zhu1	Zhu2	Shu1	Gu3	Niao3
	(pig)	(bamboo)	(comb)	(drum)	(bird)
	[3.30; 4]	[1.90; 6]	[2.11; 5]	[2.61; 4]	[3.34; 3]

---

*Note.* The first value in square brackets denotes word frequency (log10W) of each item from the SUBTLEX-CH corpus and the second value denotes HSK level.

## REFERENCES

- Abramson, A. S. (1977). Noncategorical perception of tone categories in Thai. *Journal of the Acoustical Society of America*, *61*, S66. <https://doi.org/10.1121/1.2015837>
- Alexander, J., Wong, P. C. M., & Bradlow, A. (2005). Lexical tone perception in musicians and nonmusicians. *Proceedings of 9th European conference on speech communication and technology*. [https://www.isca-speech.org/archive/interspeech\\_2005/i05\\_0397.html](https://www.isca-speech.org/archive/interspeech_2005/i05_0397.html)
- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. <https://psycnet.apa.org/doi/10.1006/jmla.1997.2558>
- Barr, D. J. (2008). Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. <https://doi.org/10.1016/j.jml.2007.09.002>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. <https://arxiv.org/abs/1406.5823>
- Best C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research* (pp. 167–200). New York Press.
- Best, C. T. (2019). The diversity of tone languages and the roles of pitch variation in non-tone languages: considerations for tone perception research. *Frontiers in Psychology*, *10*, 364. <https://doi.org/10.3389/fpsyg.2019.00364>
- Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition*, *77*(1), 1–10. <https://doi.org/10.1016/j.bandc.2011.07.006>

- Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PloS one*, 8(4), e60676. <https://doi.org/10.1371/journal.pone.0060676>
- Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer [Computer software]. Version 6.0.16, <http://www.praat.org/>
- Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, 66(4), 774–808. <https://doi.org/10.1111/lang.12159>
- Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes*, 27(7–8), 1205–1224. <https://doi.org/10.1080/01690965.2012.660170>
- Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System: An International Journal of Educational Technology and Applied Linguistics*, 36(1), 22–34. <https://doi.org/10.1016/j.system.2007.11.003>
- Broersma, M., & Cutler, A. (2011). Competition dynamics of second–language listening. *Quarterly Journal of Experimental Psychology*, 64(1), 74–95. <https://doi.org/10.1080/17470218.2010.499174>
- Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PloS one*, 5(6), e10729. <https://doi.org/10.1371/journal.pone.0010729>
- Camp, A. & Schafer, A. J. (2016, May). *Perception of Thai tones in sentence medial and final positions* [poster presentation]. The Workshop on Experimental Approaches to East Asian Linguistics. Honolulu, HI. <https://experimentalapproachestoeastasianlinguistics.wordpress.com>
- Cao, W. (2003). Xiandai hanyu kouyu ci he shumian yuci de chayi chutan [A preliminary analysis of the differences between spoken and written lexis in modern Chinese]. *Yuyan jiaoxue yu Yanjiu [Chinese language teaching and research]*, 6, 39–44.
- Chao, Y. R. (1930). A system of tone letters. *La Maître phonétique*, 45, 24–27. [http://en.cnki.com.cn/Article\\_en/CJFDTOTAL-FYZA198002000.htm](http://en.cnki.com.cn/Article_en/CJFDTOTAL-FYZA198002000.htm)

- Cheng, C., Chen, J. Y., & Xu, Y. (2014). An acoustic analysis of Mandarin tone 3 sandhi elicited from an implicit priming experiment. *4th International Symposium on Tonal Aspects of Languages*, 36–40.  
[https://www.iscaspeech.org/archive/tal\\_2014/papers/tl14\\_036.pdf](https://www.iscaspeech.org/archive/tal_2014/papers/tl14_036.pdf)
- Confucius Institute in Atlanta (2017). HSK vocabulary and sample tests. Retrieved January 6, 2017, from  
[http://confucius.emory.edu/hsk\\_and\\_resources/hsk/hsk\\_samples.html](http://confucius.emory.edu/hsk_and_resources/hsk/hsk_samples.html)
- 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.  
<https://doi.org/10.1121/1.4812435>
- Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, 29, 201–220.  
<https://doi.org/10.1177/002383098602900302>
- Cutler, A. (2008). The abstract representations in speech processing. *Quarterly Journal of Experimental Psychology*, 61(11), 1601–1619.  
<https://doi.org/10.1080/13803390802218542>
- Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.
- Cutler, A., & Chen, H. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics*, 59(2), 165–179.  
<https://doi.org/10.3758/BF03211886>
- Demir, S. (2017). An Evaluation of Oral Language: The Relationship between Listening, Speaking and Self-Efficacy. *Universal Journal of Educational Research*, 5(9), 1457–1467. <https://eric.ed.gov/?id=EJ1151845>
- Devine, T. G. (1968). Reading and listening: New research findings. *Elementary English*, 45(3), 346–348. <https://www.jstor.org/stable/41386320>
- Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language*, 42(4), 445–464.  
<https://doi.org/10.1006/jmla.1999.2697>

- Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed.). University Press.
- Duanmu, S. (2008). *Syllable structure: the limits of variation*. University Press.
- Esper, E. A. (1925). A technique for the experimental investigation of associative interference in artificial language material. *Language Monographs*, 1, 1–47.
- Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P. C. (2016). The relationship between artificial and second language learning. *Cognitive science*, 40(4), 822–847. <https://doi.org/10.1111/cogs.12257>
- Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics*, 65(7), 1029–1044. <https://doi.org/10.3758/BF03194832>
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 233–277). New York Press.
- Gandour, J. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A Linguistic survey* (pp. 41–76). Academic Press.
- Gandour, J. (1983). Tone perception in far eastern-languages. *Journal of Phonetics*, 11(2), 149–175.
- Garner, W. R. (1974). *The processing of information and structure*. Erlbaum.
- Garner, W. R. (1976). Integration of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, 8, 98–123.
- Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1, 225–241.
- Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, 66(3), 363–376. <https://doi.org/10.3758/BF03194885>
- Godfroid, A., Lin, C-H, & Ryu, C. (2017). Hearing and seeing tone through color: An efficacy study of web based, multimodal Chinese tone perception training. *Language Learning*, 67(4), 819–857. <https://doi.org/10.1111/lang.12246>

- Goldberg, D., Looney, D., & Lusin, N. (2013). *Enrollments in languages other than English in United States institutions of higher education*. Modern Language Association. [https://apps.mla.org/pdf/2013\\_enrollment\\_survey.pdf](https://apps.mla.org/pdf/2013_enrollment_survey.pdf)
- Goldstone, R. L., & Hendrickson, A. T. (2009). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(1), 69–78. <https://doi.org/10.1002/wcs.26>
- Hagtvet, B. E. (2003). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading and writing*, 16(6), 505–539. From <https://link.springer.com/article/10.1023/A:1025521722900>
- Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32(3), 395–421. [https://doi.org/10.1016/S0095-4470\(03\)00016-0](https://doi.org/10.1016/S0095-4470(03)00016-0)
- Hao, Y. C. (2014). The Application of the speech learning model to the L2 acquisition of Mandarin tones. *Proceedings of the 4th international symposium on tonal aspects of languages*. 67–70. [http://www.iscaspeech.org/archive/tal\\_2014](http://www.iscaspeech.org/archive/tal_2014)
- Harnad, S. (2003). Categorical perception. *Encyclopedia of cognitive science*. Nature Publishing Group/Macmillan.
- Hastuti, U. N., & Kalim, N. (2019). Correlation between reading comprehension and listening comprehension Skills in completing TOEFL-PBT. *Journal of English Education*, 1(2), 45–52. <https://link.springer.com/article/10.1023/A:1025521722900>
- Hayakawa, S., Ning, S., & Marian, V. (2020). From Klingon to Colbertian: Using artificial languages to study word learning. *Bilingualism: Language and Cognition*, 23(1), 75–80. <https://doi.org/10.1017/S1366728919000592>
- He, Y., Wang, Q., & Wayland, R. (2016). Effects of different teaching methods on the production of Mandarin tone 3 by English speaking learners. Chinese as a Second Language. *The journal of the Chinese Language Teachers Association*, 51(3), 252–265. <https://doi.org/10.1075/csl.51.3.02he>

- Hoffmann, C. W., Sadakata, M., Chen, A., Desain, P., & McQueen, J. M. (2014). Within-category variance and lexical tone discrimination in native and non-native speakers. *Proceedings of the 4th international symposium on tonal aspects of languages* (pp. 45–49). [http://www.isca-speech.org/archive/tal\\_2014](http://www.isca-speech.org/archive/tal_2014)
- Huang, Y. (2005). “Frequency of Mandarin words based on Google@,” [Data file]. Retrieved from <http://yong321.freeshell.org/misc/ChineseCharFrequencyG.txt>
- Huang, T. (2001). The interplay of perception and phonology in tone 3 sandhi in Chinese Putonghua. *OSU Working Papers in Linguistics*, 55, 23–42. [https://www.researchgate.net/publication/285822401\\_The\\_interplay\\_of\\_perception\\_and\\_phonology\\_in\\_tone\\_3\\_sandhi\\_in\\_Chinese\\_Putonghua](https://www.researchgate.net/publication/285822401_The_interplay_of_perception_and_phonology_in_tone_3_sandhi_in_Chinese_Putonghua)
- Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). Perception and production of Mandarin Chinese tones. In P. Li, L. H. Tan, E. Bates & O. J. L. Tzeng (Eds.), *Chinese [Handbook of East Asian Psycholinguistics, Vol. 1]* (pp. 209–217). Cambridge University Press.
- Jusczyk, P.W. (1997). *The discovery of spoken language*. Cambridge, MIT.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13). doi: 10.18637/jss.v082.i13
- Lee, C. Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech*, 50(1), 101–123. <https://doi.org/10.1177%2F00238309070500010501>
- Li, F., Xie, Y., Yu, X., & Zhang, J. (2016, October). A study on perceptual training of Japanese CSL learners to discriminate Mandarin lexical tones. In *2016 10th international symposium on Chinese spoken language processing (ISCSLP)* (pp. 1-5). IEEE. <https://ieeexplore.ieee.org/abstract/document/7918484>
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358–368. doi: 10.1037/h0044417



- 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), 2827–2838. <https://doi.org/10.1121/1.4898047>
- Lin, W. C. J. (1985). Teaching mandarin tones to adult English speakers: Analysis of difficulties with suggested remedies. *RELC Journal*, 16, 31–47. <https://doi.org/10.1177/003368828501600207>
- Liu, S., & Samuel, A. G. (2007). The role of Mandarin lexical tones in lexical access under different contextual conditions. *Language and Cognitive Processes*, 22(4), 566–594. <https://doi.org/10.1080/01690960600989600>
- Liu, Y., Wang, M., Perfetti, C. A., Brubaker, B., Wu, S., & MacWhinney, B. (2011). Learning a tonal language by attending to the tone: An in-vivo experiment. *Language Learning*, 61, 1119–1141. <https://doi.org/10.1111/j.1467-9922.2011.00673.x>
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, 94(3), 1242–1255. <https://doi.org/10.1121/1.408177>
- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, 6, Article 1171. <https://doi.org/10.3389/fpsyg.2015.01171>
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, 89(2), 874–886. <https://doi.org/10.1121/1.1894649>
- Loewen, S. (2020). *Introduction to instructed second language acquisition*. Routledge.
- Loewen, S., & Sato, M. (Eds.). (2017). *The Routledge handbook of instructed second language acquisition*. Taylor & Francis.
- Lu, S., Wayland, R., & Kaan, E. (2015). Effects of production training and perception training on lexical tone perception—A behavioral and ERP study. *Brain Research*, 1624, 28–44. <https://doi.org/10.1016/j.brainres.2015.07.014>

- MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M.B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). Oxford University Press.
- MacWhinney, B. (2012). The logic of the unified model. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 211–227). Routledge.
- Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, 62(4), 407–420. <https://doi.org/10.1016/j.jml.2010.02.004>
- Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, 50(8), 2032–2043. <https://doi.org/10.1016/j.neuropsychologia.2012.05.002>
- Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and Cognition*, 6(2), 97–115. <https://doi.org/10.1017/S1366728903001068>
- Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, 53(4), 372–380. <https://link.springer.com/article/10.3758/BF03206780>
- McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, 10(8), 363–369. <https://doi.org/10.1016/j.tics.2006.06.007>
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475–494. <https://doi.org/10.1016/j.jml.2007.11.006>
- Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America*, 102(3), 1864–1877. <https://doi.org/10.1121/1.420092>

- Moulines, E., & Laroche, J. (1995). Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Communication, 16*(2), 175–205.  
[https://doi.org/10.1016/0167-6393\(94\)00054-E](https://doi.org/10.1016/0167-6393(94)00054-E)
- Packard, J. L. (2015). Morphology: Morphemes in Chinese. In Wang, S-Y & Sun, C. (Eds.), *Oxford handbook of Chinese linguistics* (pp. 263–273). Oxford University Press.
- Peirce, J. W. (2007). PsychoPy-psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1), 8–13.  
<https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2019). Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition, 41*(1), 59–86. <https://doi.org/10.1017/S0272263117000444>
- Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2020). Even in the best-case scenario L2 learners have persistent difficulty perceiving and utilizing tones in Mandarin: Findings from behavioral and event-related potentials experiments. *Studies in Second Language Acquisition, 1*–29.  
<https://doi.org/10.1017/S027226312000039X>
- Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (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>
- Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., & Wang, W. S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics, 38*(4), 616–624. <https://doi.org/10.1016/j.wocn.2010.09.003>
- Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics, 2*, 33–52.  
<https://doi.org/10.1146/annurev-linguistics-030514-125050>
- Poltrock, S., Chen, H., Kwok, C., Cheung, H., & Nazzi, T. (2018). Adult learning of novel words in a non-native language: consonants, vowels, and tones. *Frontiers in Psychology, 9*, 1211. <https://doi.org/10.3389/fpsyg.2018.01211>

- Qin, Z. (2017). *How native Chinese listeners and second-language Chinese learners process tones in word recognition: An eye-tracking study* [Doctoral dissertation, University of Kansas]. Ku ScholarWork.  
<https://kuscholarworks.ku.edu/handle/1808/26474>
- Quam, C., & Creel, S. C. (2017). Mandarin-English bilinguals process lexical tones in newly learned words in accordance with the language context. *PloS one*, *12*, 1–27.  
<https://doi.org/10.1371/journal.pone.0169001>
- R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available at: <https://www.R-project.org/> (accessed April 2019).
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*(1), 71–80. [https://doi.org/10.1016/S0167-6393\(02\)00094-8](https://doi.org/10.1016/S0167-6393(02)00094-8)
- 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.  
<https://doi.org/10.1177/0023830914522956>
- Shao, J. (2001). *Xiandai hanyu tonglun [Modern Chinese linguistics]*. Shanghai Education Press.
- Shen, G., & Froud, K. (2016). Categorical perception of lexical tones by English learners of Mandarin Chinese. *The Journal of the Acoustical Society of America*, *140*(6), 4396–4396. <https://doi.org/10.1121/1.4971765>
- Shen, G., & Froud, K. (2019). Electrophysiological correlates of categorical perception of lexical tones by English learners of Mandarin Chinese: an ERP study. *Bilingualism*, *22*(2), 253–265.  
<https://doi.org/10.1017/S136672891800038X>
- Shen, X. S. (1989). Toward a register approach in teaching Mandarin tones. *Chinese Language Teachers Association*, *24*, 27–47.
- Shih, C. (1988). Tone and intonation in Mandarin. *Work Papers of the Cornell Phonetic Laboratory*, *3*, 83–109.

- So, C. K., & Best, C. T. (2008). Do English speakers assimilate Mandarin tones to English prosodic categories? *Interspeech*, Article 1120.  
<http://handle.uws.edu.au:8081/1959.7/45242>
- 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. <https://doi.org/10.1016/j.wocn.2007.06.005>
- Song, S., Georgiou, G. K., Su, M., & Hua, S. (2016). How well do phonological awareness and rapid automatized naming correlate with Chinese reading accuracy and fluency? A meta-analysis. *Scientific Studies of Reading*, 20(2), 99–123.  
<https://doi.org/10.1080/10888438.2015.1088543>
- 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>
- Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final-/t/. *Journal of Memory and Language*, 52, 322–338. doi: 10.1016/j.jml.2009.01.001
- Taft, M., & Chen, H. (1992). Judging homophony in Chinese: The influence of tones. In H. Chen & O. J. L. Tzeng (Eds.), *Language processing in Chinese* (pp.151–172). North-Holland.
- Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, 29(6), 557–580.  
<https://doi.org/10.1023/A:1026464108329>
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634.  
<https://doi.org/10.1126/science.7777863>
- Tao, H. (2015). Profiling the Mandarin spoken vocabulary based on corpora. In Wang, S-Y & Sun, C. (Eds.), *Oxford handbook of Chinese linguistics* (pp. 336–347). Oxford University Press.

- Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, 2(3), 339–363.  
<https://doi.org/10.3758/BF03210972>
- Teruya, H. & Kapatsinski, V. (2019). Deciding to look: revisiting the linking hypothesis for spoken word recognition in the visual world. *Language, Cognition and Neuroscience*, 34(7), 861–880. <https://doi.org/10.1080/23273798.2019.1588338>
- 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>
- Tsai, R. (2011). Teaching and learning the tones of Mandarin Chinese. *Scottish Languages Review & Digest*, 24, 43–50.  
[https://scilt.org.uk/Portals/24/Library/slr/issues/24/24\\_5\\_Tsai.pdf](https://scilt.org.uk/Portals/24/Library/slr/issues/24/24_5_Tsai.pdf)
- U.S. Department of State, Foreign Service Institute, School of Language Studies. (2019). Languages. Available at: <https://www.state.gov/key-topics-foreign-service-institute/foreign-language-training/> (accessed 4 June 2019)
- Van Hessen, A. J., & Schouten, M. E. H. (1992). Modeling phoneme perception. II: A model of stop consonant discrimination. *The Journal of the Acoustical Society of America*, 92(4), 1856–1868. <http://doi.org/10.1121/1.403842>
- Wang, S-Y (1967). Phonological features of tone. *International Journal of American Linguistics*, 33(2), 93–105.
- Wang, S-Y (1976). Language change. *Annals of the New York Academy of Sciences*, 280(1), 61–72. <https://doi.org/10.1111/j.1749-6632.1976.tb25472.x>
- Wang, S-Y, & Sun, C. (2015). Introduction. In Wang, S-Y & Sun, C. (Eds.), *Oxford handbook of Chinese linguistics* (pp. 3–18). Oxford University Press.
- Wang, X. (2013). Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal*, 97(1), 144–160.  
<https://doi.org/10.1111/j.1540-4781.2013.01386.x>
- Wang, Y., Behne, D. M., Jongman, A., & Sereno, J. A. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics*, 25(3), 449–466. <https://doi.org/10.1017/S0142716404001213>

- Wang, Y., Jongman, A., & Sereno, J. A. (2006). L2 acquisition and processing of Mandarin tone. In P. Li, L. H. Tan, E. Bates & O. J. L. Tzeng (Eds.), *Chinese handbook of east asian psycholinguistics, Vol. 1*, (pp. 250–256). Cambridge University Press.
- Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649–3658. <https://doi.org/10.1121/1.428217>
- Wayland, R., & Guion, S. (2003). Perceptual discrimination of Thai tones by naive and experienced learners of Thai. *Applied Psycholinguistics*, *24*(1), 113–129. <https://doi.org/10.1017/S0142716403000067>
- Webb, S. (Ed.). (2019). *The Routledge handbook of vocabulary studies*. Routledge.
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. [https://doi.org/10.1016/S0749-596X\(03\)00105-0](https://doi.org/10.1016/S0749-596X(03)00105-0)
- Wee, L-H, & Li, M. (2015). Modern Chinese phonology. In Wang, S-Y & Sun, C. (Eds.), *Oxford handbook of Chinese linguistics* (pp. 474–489). Oxford University Press.
- Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35–44. <https://doi.org/10.3758/BF03207136>
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- Wiener, S., Chan, M. K. M., & Ito, K. (2020). Do explicit instruction and high variability phonetic training improve non-native speakers' Mandarin tone productions? *The Modern Language Journal*, *104*(1), 152–168. <https://doi.org/10.1111/modl.12619>
- Wiener, S., & Ito, K. (2016). Impoverished acoustic input triggers probability-based tone processing in mono-dialectal Mandarin listeners. *Journal of Phonetics*, *56*, 38–51. <https://doi.org/10.1016/j.wocn.2016.02.001>
- Wiener, S., Lee, C-Y, & Tao, L. (2019). Statistical regularities affect the perception of second language speech: Evidence from adult classroom learners of Mandarin Chinese. *Language Learning*, *69*, 527–558. <https://doi.org/10.1111/lang.12342>

- Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and Speech*, 59(1), 59–82.  
<https://doi.org/10.1177/0023830915578000>
- Wong, P. C., & Perrachione, T. K. (2007.) Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, 28(4), 565–585. <https://doi.org/10.1017/S0142716407070312>
- Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, 170(1), 223–231.  
<https://doi.org/10.1016/j.neuroscience.2010.06.077>
- Xiao, R., Rayson, P., & McEnery, T. (2009). *A Frequency dictionary of Mandarin Chinese: Core vocabulary for learners*. Routledge.
- Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, 120(2), 1063–1074.  
<https://doi.org/10.1121/1.221357>
- Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes Special Issue: Processing East Asian Languages*, 14(5-6), 609–630. <https://doi.org/10.1080/016909699386202>
- Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68(2), 123–139.  
<https://doi.org/10.1016/j.jml.2012.09.004>
- Yip, M. (2002). *Tone*. Cambridge University Press
- Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, 49(7), 1761–1770. <https://doi.org/10.1016/j.neuropsychologia.2011.02.054>
- Zhao, T. C., & Kuhl, P. K. (2015). Effect of musical experience on learning lexical tone categories. *The Journal of the Acoustical Society of America*, 137(3), 1452–1463.  
<https://doi.org/10.1121/1.4913457>



Zou, T., Chen, Y., & Caspers, J. (2017). The developmental trajectories of attention distribution and segment-tone integration in Dutch learners of Mandarin tones. *Bilingualism: Language and Cognition*, 20(5), 1017–1029.  
<https://doi.org/10.1017/S1366728916000791>