EFFECTS OF HIGH VARIABILITY PHONETIC TRAINING ON MONOSYLLABIC AND DISYLLABIC MANDARIN CHINESE TONES FOR L2 CHINESE LEARNERS

By

Yingjie Li

Submitted to the graduate degree program in the Department of Curriculum and Teaching and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chairperson Dr. Manuela Gonzalez-Bueno

Co-Chairperson Dr. Joan Sereno

Dr. Marc Mahlios

Dr. Paul Markham

Dr. Jie Zhang

Date Defended: April 27th, 2016

The Dissertation Committee for Yingjie Li

certifies that this is the approved version of the following dissertation:

EFFECTS OF HIGH VARIABILITY PHONETIC TRAINING ON MONOSYLLABIC AND DISYLLABIC MANDARIN CHINESE TONES FOR L2 CHINESE LEARNERS

Chairperson Dr. Manuela Gonzalez-Bueno

Co-Chairperson Dr. Joan Sereno

Date approved: April 27th, 2016

Abstract

Although computer-assisted auditory perceptual training has been shown to be effective in learning Mandarin Chinese tones in monosyllabic words, tone learning has not been systematically investigated in disyllabic words. In the current study, seventeen native English-speaking beginning learners of Chinese were trained using high variability phonetic training paradigm. Two perceptual training groups, a monosyllabic training group and a disyllabic training group, were compared and accuracy in identifying the tonal contrasts in naturally produced monosyllabic and disyllabic words (produced by native Mandarin Chinese speakers) was evaluated. The learners' performance on tones in disyllabic words was also investigated in terms of syllable position (initial and final position), tonal context (compatible and conflicting context), and tonal sequence (same and different sequence). Results showed that after four training sessions in a two-week period, beginning learners of Chinese significantly increased their tonal identification accuracy from the pretest (60%) to posttest (65%) and this improvement in training generalized to new stimuli by a new speaker (12% increase). The current findings, however, did not show significant differences between the monosyllabic perceptual training group and disyllabic perceptual training group: both showed improvements from pretest to posttest. Although native English-speaking learners in both training groups made improvements in their tonal identification performance in general, when examining learning for the two types of stimuli (monosyllabic and disyllabic stimuli), the results showed distinct patterns in the learners' performance. While both training groups improved tonal perception in monosyllabic stimuli, training with disyllabic stimuli (disyllabic training group) was much more effective (especially for the disyllabic stimuli) and significantly helped native English-speaking participants to acquire the tones. These results illustrate the limitations of the current tone teaching based solely on monosyllabic words. Instead, the current results advocate for incorporating more common disyllabic words, which are highly variable, into tone learning routines in the classroom in order to achieve native-like tone acquisition.

Acknowledgments

The topic of my dissertation was first developed in my *Topics in Research in Experimental Linguistics* class (Ling850) in Spring 2013. I defended my proposal in April 2014, and successfully defended my dissertation in April 2016. Throughout this long journey toward my doctorate I have received immense support and love from my committee members, professors, friends and family.

First of all, I would like to express my sincere gratitude to all professors on my dissertation committee, who have made this graduation possible for me.

Dr. Joan Sereno, my advisor and co-chair, guided me consistently through my research. From the very first class I took from her—*Introduction to Psycholinguistics* –I knew that I could turn to her when I needed encouragement and help both professionally and personally. Dr. Sereno is not just an acknowledged researcher in her field, but also a great mentor and teacher for all her students. Her sense of humor and great knowledge of all the related fields made learning intriguing and inspiring. Over the past three years, she had numerous meetings with me, read countless drafts, and provided constructive feedback, from my proposal to the final version of my dissertation. She also made sure that I practiced many times to be ready for my final defense. I greatly appreciate her honest and straightforward comments about the quality of my work. This dissertation could not have been completed without her support and guidance.

Dr. Manuela Gonzalez-Bueno, my advisor and co-chair, provided continuous support and help throughout my graduate life here at the University of Kansas. The feedback and suggestions I received from her on my dissertation were invaluable. Dr. Gonzalez-Bueno is not just an advisor but also a great friend. I especially thank her for those coffee hours and tea talks that we shared on and off campus. Her advice and perspective saved me when I was at a low point in my life. She is and will always be a great friend of mine.

I would like to give special thanks also to Dr. Jie Zhang, who is the Graduate Studies Representative on my committee. Dr. Zhang is such a great teacher and so knowledgeable about Mandarin Chinese tones. In fact, it was the first class I took from him—*Structure of Chinese* – that triggered my interest in tones. His critical questions and suggestions on this topic prepared me for my final defense and I deeply appreciate his help along the way.

I am very grateful, also, to have had Dr. Marc Mahilos and Dr. Paul Markham as my committee members. They were willing to meet and discuss my dissertation with me at my request. A special thanks must go to Dr. Mahilos, who never hesitated to lend the entire contents of his bookshelf to me when I was working on my theoretical framework part.

In addition, I would like to thank all my peer colleagues in Ling 850 from Spring 2013 to Spring 2016, and to the professors of that class: Dr. Allard Jongman, Dr. Annie Tremblay, Dr. Jie Zhang and Dr. Joan Sereno. Their invaluable suggestions and provocative questions during my dissertation practice talks sharpened my presentation.

I would also like to thank students, colleagues and professors in the Chinese program at KU, especially Dr. Yan Li and Dr. Keith McMahon, who not only provided me with the opportunity to teach Chinese language in the program, but also helped me recruit participants from the program for my study. I am very grateful to my many great friends at KU who were there whenever I was in need of encouragement, a run-through or simply a hug. Goun Lee, Steve Politzer-Ahles, Maite Martínez-García, Hanbo Yan, Seulgi Shin, and Xiao Yang, thank you for listening to my practice talks multiple times and giving me feedback. And I would like to thank

Randi Hacker, my long-time friend at KU, and Philip Duncan, a friend in the Linguistic Department, both of whom proofread my dissertation and gave me valuable comments.

Last but not least, I would like to extend my deepest thanks to my dear parents and my wonderful family. My father, Xianming Li, and my mother, Yuefang Jiang, have given their unconditional love and trust to me all these years, and my mother, has never doubted that I would succeed. Whenever I encountered an obstacle, I heard my parents' words: "Only after you taste bitterness will you appreciate the sweetness of life". Finally, my most wholehearted gratitude goes to my husband, Tom, my daughter, Madison (Xuemeng Li), and my son, Raymond (Tingrui Li): You have been my backbones and cornerstones on this journey. You have given me the strength, love, care and courage I needed to keep going no matter what hurdles I might have had to jump along the way. Tom, thank you for being there for me whenever and wherever. I couldn't have done it without you.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

With China's long and rich history and quickly developing economy, more and more Americans are interested in learning Mandarin Chinese as a foreign language at the college level. Unlike English, Mandarin Chinese is a tonal language, and every Chinese character has a tone. That is to say, tone is a key component of the lexicon in the Chinese language. Chinese people use these phonemic tones to distinguish word meaning. Thus, perceiving and producing tones correctly is of critical importance for all Chinese language learners to communicate successfully in the language.

1.1 Statement of the problem

Native English-speaking learners of Chinese have difficulty perceiving and producing tones in Mandarin Chinese since the phonemic tone feature is not in part of their native language system (Miracle, 1989; Shen, 1989; Shen & Lin, 1991; Sun, 1998; Jongman, Wang, Moore, & Sereno, 2006; Lee, Tao, & Bond, 2010; He, 2010; He & Wayland, 2010, 2013; Chang, 2011; Hao, 2012). However, the majority of these studies focus solely on tones of monosyllabic words in an isolated environment instead of on tones in natural, connected speech. Moreover, few studies have examined tones in disyllabic words. Even when attention was given to the tones at the word and sentential level (Sun, 1998; He, 2010; He & Wayland, 2010, 2013; Guo & Tao, 2008), the final analyses still focus primarily on perception or production of the four basic tones in isolation. Many of these analyses additionally failed to examine the effect of adjacent tones. These coarticulated tones are a central part of real life conversation and contribute greatly to native-like speech.

At the same time, in the current Chinese language classroom in the United States, tones are introduced to the learners mainly in isolation within a short period of time at the very beginning of learning the target language. Xing (2006) investigated teaching and learning Mandarin Chinese as a foreign language in the United States across different levels from public schools to universities. She found that Chinese language teachers in the classroom usually focus on introducing Mandarin tones in isolation and focus on drill practice on perceiving tones in isolated monosyllabic words. Similarly, Orton (2013), after observing many college level Mandarin Chinese classrooms in the United States, found that “once the tone information is provided, at the beginning of the course or textbook … oral development work involves only short period of time in-class listening and repetition of tonal syllables, often monosyllables, with the occasional row of disyllables” (Orton, J., 2013, p.10).

These studies reveal that the current tone teaching in the United States is problematic in two ways: On the one hand, considering its important role in communication, there is simply not enough attention given to tone teaching and learning. On the other hand, most current tone teaching concentrates mainly on perceiving tones in isolated monosyllabic words, when, in real conversation, monosyllable words are rarely used in authentic communication. As noted by Zhou, Marslen-Wilson, Taft and Shu (1999, p. 526), “compound words, which are all disyllable words in Chinese, compose 70% of all words used in Chinese”. Likewise, Duanmu (1999) also found that the disyllabic words are dominant in the vocabulary of modern Mandarin Chinese, rather than the monosyllabic words. Moreover, a statistical analysis was conducted for 31,159 Mandarin words used in public media, including newspapers, magazines and TV (as cited by Duanmu, 1999), and found that 22,941 (74%) of these words were disyllables, and only 12%

were monosyllables. The remaining 14% of the words have more than two syllables. It can be concluded from this data that disyllabic words and their connected tones are used most often in people's daily life rather than monosyllabic words with their isolated tones. Disyllabic tones mirror the tones perceived and produced at the sentence level more than isolated tones do.

From the above information, it can be seen that studying tones in monosyllable words alone will not be sufficient or, indeed, efficient, for learners of Mandarin Chinese. When teaching Mandarin Chinese pronunciation to native English-speaking learners, understanding how to improve their tonal perception is paramount if they are to succeed in communicating naturally and intelligently. As Orton (2013) strongly suggested, that the phonological challenges of Chinese for English language-speaking learners, tone specifically, must be tackled from the start, and constantly attended to thereafter. In light of this need, the current study investigates disyllabic tones in learners' perception as the first step to understanding their processing of the target language.

1.2 Pedagogical perspectives

Computer-assisted language learning has long been an effective pedagogical approach since it was integrated into foreign language pronunciation teaching in the 1980's. For instance, Molholt (1988) used a computer software program named Speech Spectrographic Display (SSD), which provided instant visual displays of the target sound, word or even sentences in English to Chinese learners, so that these learners were able to compare their production to the native speaker's production in order to overcome their pronunciation problems in English. Hiller, Rooney, and Jack (1993) examined a computer based project named Interactive System for Spoken European Language Training, which concentrated on teaching pronunciation of individual words or short phrases plus additional exercises for intonation, stress and rhythm to non-native speakers of English, French and Italian. Similarly, Quintana Lara (2009) also implemented Acoustic Visual Feedback Instruction into her traditional teaching classroom for pre-service English language teachers, who were native Spanish speakers. The teachers who trained in this instruction significantly improved their English high-front vowel production. These studies demonstrated how incorporating computer-assisted learning into the foreign language class does, indeed, help non-native learners to learn the target language's pronunciation.

However, current in-class pedagogical approaches to teach Mandarin Chinese tones are still using traditional methods that lack computer-mediated assistance. Some traditional approaches to teaching tones that are still utilized in classrooms include listen-and-repeat, minimal-pair drills, and reading aloud. All these practices require guidance by language teachers. In some recently published textbooks, the articulatory descriptions (mainly for the vowels) are added to give the learners a direct and visual description of the target vowel sound (Orton, 2013). Computer-assisted language learning has not been widely incorporated into the teaching and learning process as seen in ESL classrooms. As Philip Hubbard pointed out, computer assisted learning provides many advantages to modern foreign language teaching classes, such as learning efficiency and effectiveness, easy access, great convenience, strong motivation, and institutional efficiency (Hubbard, 2009).

Short term auditory training on computers has proved to be effective in assisting learners to acquire new phonetic contrasts that do not exist in their native phonological system in various languages (Logan, Lively & Pisoni, 1991; Lively, Logan, & Pisoni, 1993; Wang, Spence, Jongman & Sereno, 1999; Wang, Jongman & Sereno, 2003; Kingston, 2003; Francis, Ciocca, Ma, & Fenn, 2008; Herd, et al. 2013). In such cases, through carefully designed perceptual training procedures, learners listen to a large variety of stimuli produced naturally by multiple native speakers of the target language. Even in a short period of time, the learners' perception of the target sound (that originally is not in their native language system) is improved through the exposure. The results from these previous studies show that this type of training helps improve not only learners' perception, but also even pronunciation in the target languages, such as English, Chinese, German, Cantonese and Spanish. Furthermore, this perceptual improvement was successfully extended to the learners' production, as shown by Japanese learners of English learning /r/ and /l/ (Logan, et al., 1991; Lively, et al., 1993; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997, 1999), as well as by American learners of Mandarin Chinese learning monosyllabic tones (Wang et al. 1999, 2003).

1.3 Purpose of the study

Previous research by Wang et al. (1999) has found that through a short high variability phonetic training using monosyllabic tones in Mandarin Chinese, American beginning learners of Mandarin Chinese all improved significantly in their tonal perception and production of the four Mandarin Chinese tones in monosyllable words. But their study did not address whether the monosyllabic tone training and learning would help learners identify tones in disyllabic words, which more accurately reflect tones as they are used in sentences. This raises the question: Would learners' tonal perception improve through training on disyllabic words just as they did through training on monosyllabic ones?

The purpose of the current study is to examine learners' tonal behavior through perceptual training in order to find an effective teaching method for teaching Mandarin Chinese tones to native speakers of English. The goal is to determine which tones and tonal combinations are difficult for English-speaking learners to acquire as beginner foreign language learners. Moreover, this study also proposes possible pedagogical methods for learning tones to ultimately help learners gain greater proficiency in Mandarin Chinese. It is not only important to learn Mandarin tones correctly but also necessary for learners to perceive them accurately in order to achieve intelligibility in communication.

1.4 Research Questions

This study aims to find out if beginning English-speaking learners' perception of Chinese Mandarin tones in both monosyllabic words and disyllabic words will be improved after perceptual training involving either monosyllabic training or disyllable training. Towards this end, the following questions are investigated:

Research Question 1. After perceptual training, will native English-speaking learners improve their perception of tones generally in both monosyllabic words and disyllabic words in Mandarin Chinese?

Research Question 2. Compared to monosyllabic perceptual training, will disyllabic perceptual training be more effective in helping English-speaking learners shape their tonal categories and improve their tone perception of Mandarin Chinese?

Research Question 3. Contrasting two types of training materials, monosyllabic stimuli and disyllabic stimuli, which will be more effective in helping to learn monosyllabic tones? And which will be more effective in helping to learn disyllabic tones?

Research Question 4. Will training using monosyllabic material transfer to disyllabic tone identification? And will training using disyllabic material transfer to monosyllabic tone identification?

Research Question 5. Will factors, specifically syllable position, tonal context, and tonal sequence, affect native English-speaking learners' tone perception in disyllabic words?

1.5 Significance of the study

Learning Mandarin Chinese tones correctly is critical for achieving successful communication. Of particular importance is understanding how disyllabic tones are perceived and processed by learners, given that disyllabic words occur with greater frequency in real-world conversation. Conducting a perceptual training study for native English-speaking learners to train them in the learning of tones, especially disyllabic tones, thus, has great potential as a tool for facilitating tone learning.

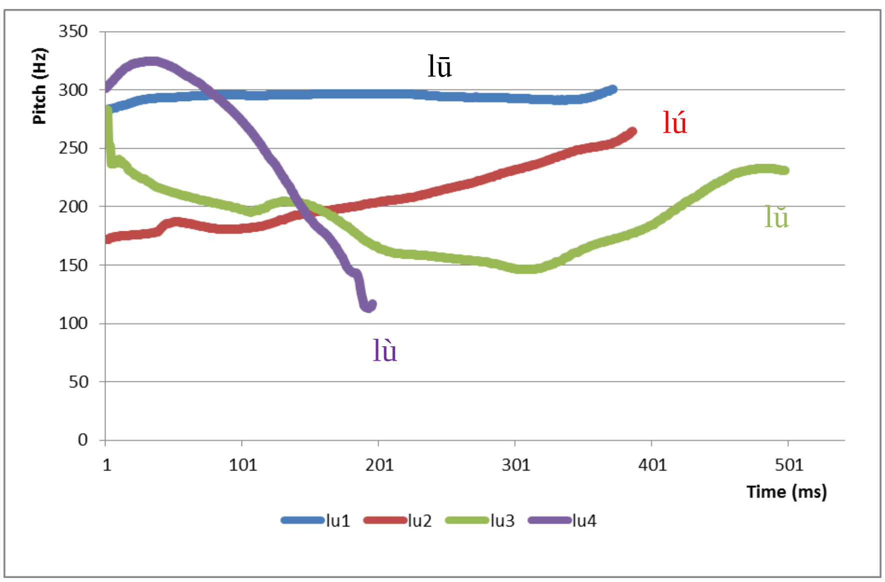
This is the first study to examine the effect of high variability phonetic training to native English-speaking learners of Mandarin Chinese by using disyllabic training stimuli. Previous studies investigated the tonal training effect only using monosyllable training stimuli. Moreover, it is the first study to observe the transfer of the training effect in perception. Specifically, the present study examines learners' tonal identification of monosyllabic tones while they are trained using disyllable stimuli, and the tonal identification of the disyllabic tones while trained using monosyllable stimuli. Additionally, the current study will provide evidence for the effectiveness of incorporating computer-assisted teaching into traditional Mandarin Chinese language teaching and learning classes if the native English-speaking learners' tonal perception is significantly improved within a short training period of time. This improvement could help the learners to achieve more native-like proficiency in Mandarin Chinese.

Chapter 2: Literature Review

2.1 Tones in Mandarin Chinese

Each Mandarin Chinese character has a tone. Tone in Mandarin Chinese is a suprasegmental feature, and it differentiates lexical meaning of a syllable. There are four phonemic tones in Mandarin Chinese, which can be perceptually distributed on a five point pitch scale that provides a direct visual representation of the pitch contours. Figure 1 below shows the pitch contour based on fundamental frequency for the four phonemic tones. In monosyllabic words, Tone 1 (T1) is high and flat with a pitch value of 55; Tone 2 (T2) is a high-rising tone with a pitch value of 35; Tone 3 (T3) is a low-dipping tone with a pitch value of 213; and Tone 4 (T4) is a high-falling tone with a pitch value of 51 (Chao & Pian, 1955).

Figure 1: Fundamental frequency contours (Hz) of four phonemic tones of /lu/ as spoken by a female native Chinese speaker



The pitch value of each tone affects the lexical meaning of its Chinese word. Consider the syllable “lu”: when the pitch value is 55 (T1), the syllable means “sound of grumbling or chattering"; if the pitch value is 35 (T2), it means "stove"; when the pitch value is 213 (T3), the syllable means "brine"; and if "lu" has a pitch value of 51 (T4), it means "road". The four tones are usually indicated by four diacritic marks in Pinyin, as illustrated in the examples in Table 1. In terms of tonal classification, T1 is a level tone due to its relatively consistent high pitch level 55, and T2, T3 and T4 are contour tones that contain the pitch rising and falling changes within a syllable.

Therefore, it is not only important to learn Mandarin tones correctly but also necessary for learners to perceive and produce them accurately in order to achieve intelligibility in communication.

Table 1: Descriptions of four Chinese phonemic tones, pitch values and examples.

Tone	Description	Pitch Value	Example
1	high level	55	lū "sound of grumbling or chattering"
2	high rising	35	lú "stove"
3	dipping/falling-rising	213	lŭ "to brine"
4	high falling	51	lù "road"

The descriptions given above are for each tone's canonical form—the contour with which that tone is pronounced in isolation which is quite stable in pronunciation. Mandarin tones often undergo alternation when produced in connected speech. In disyllable words, for example, when T3 is followed by another T3, the first T3 will change to a rising T2. In other non-final positions, when preceding any tone other than T3, T3 is pronounced as a low tone with pitch value of 21—

without the final rise that occurs when the tone is produced at the end of a prosodic phrase or in isolation. Also, T4 changes to a high-mid tone with pitch value of 53 in connected speech (Lin, 2007). Mandarin Chinese tone coarticulation was investigated systematically by Xu (1994, 1997). Xu (1994) examined tonal variation in naturally produced tri-syllable Chinese words by native Chinese speakers. He proposed the concept of “compatible” versus “conflicting” tonal contexts, in which the pitch value of one tone was affected by the adjacent tone. In compatible contexts, adjacent tones share identical or similar pitch values at the syllable boundary. Thus, little or no compromise of the temporal overlap is necessary in production, and the pitch value shared by both tones is realized to the fullest possible extent. However, in conflicting contexts, temporal overlap is a compromise between adjacent phonetic units that differs substantially in their pitch value. As a consequence, this compromise results in variations in the onset and offset and even overall height of the tone. In perception tasks, through phonetic manipulation, Xu (1994) found that native speakers use information from coarticulation of adjacent tones to help identify the target tones correctly. Moreover, fundamental frequency (f0) analyses suggest that there is greater carryover (from preceding to following) than anticipatory (from following to preceding) tonal coarticulation in tri-syllabic words and phrases in Mandarin Chinese. This carryover effect is supported by Xu’s (1997) study on disyllable words. Sixteen possible tonal combinations of four Mandarin tones were investigated in a CV syllable sequence /mama/. He pointed out that the offset’s pitch value on the preceding tone affects the onset of the following tone greater than vice versa. That is to say, the carry-over effect is larger than the anticipatory effect at the disyllable level. But this finding disagrees with Shen (1990), who also studied Mandarin tri-syllables and found symmetrical bi-directional effects. This suggests that the carry-over effect between adjacent tones is equal to the anticipatory effect.

These studies about tones show that the nature of tones differs depending on context. Learning monosyllabic tones can therefore only provide a partial picture of tone learning in Mandarin Chinese. This suggests that learning should be extended to disyllabic tones in order to accurately simulate the variability of natural speech. .

2.2 Native English-speaking learners' perception of Mandarin Chinese monosyllabic tones

Many studies have analyzed native English-speaking learners' perception of Mandarin tone in isolation (Sun, 1998; Gottfried & Suiter, 1997; Wang, et al., 1999; Wang, et al., 2003; Jongman, et al., 2006; Guo & Tao, 2008; Lee, Tao, & Bond, 2010, Hao, 2012). When Mandarin tones are in isolation, it is found that American listeners have particular difficulty differentiating T2 and T3. For example, Sun (1998) compared American learners' identification of tones on three word types in monosyllabic words in Mandarin Chinese: common and uncommon real monosyllable words, as well as nonsense monosyllable Chinese words. She found that learners' identification accuracy between common and uncommon real words was not significantly different, although they had a higher accuracy perceiving T1 and T4 than T2 and T3 when these tones were in isolation. Also, the learners identified tones better in real monosyllable words than in nonsense monosyllable words. Her results showed T3 posed the most difficulty to identify for the learners across groups. The next difficult tone was T2, and then followed by T1 and T4. The American learners in Sun's study were all recruited from an intensive Chinese language immersion summer program in China. These learners were immersed in a Chinese-speaking environment, hearing and using Chinese words regularly. Not to mention that all participants already had more than one year of Chinese language learning experience by the time of the experiment. Their knowledge of lexical items (both common and uncommon words) resulted in learners' better performance on real words than nonsense stimuli in their perception tasks. Wang et al. (1999, 2003) had similar findings with beginning learners for monosyllable words in isolation before and after their perceptual training, in which the T2 and T3 confusion was greater than other tones in American learners' perception, and T3 was the worst in learners' tone production. In Wang et al.'s study, all American participants were just beginning learners with one or two semesters of Chinese language courses at the college level. None of the participants had ever lived in a Mandarin-speaking environment. These findings demonstrate that despite the length of language learning experience, in monosyllable words, T3 and T2 are hardest for English-speaking learners to perceive.

Gottfried and Suiter (1997) also anlayzed American listerners' tonal error patterns in an identfication task on monosyllable Chinese words, but they manipulated the extracted monosyllable stimuli from a sentence carrier, and had American listeners percieve target tones of intact syllables, syllables with the initial and final protions removed, syllables with the centers removed, and syllables with only the intial transition presented. Tone identification results show that T2 and T3 are still the most challenging ones to differentiate. When analyzing tonal error patterns, Gottfried and Suiter (1997) pointed out that confusion between T2 and T3 in perception is due to the fact that American listeners paid primary attention to the pitch height of these two tones, which share a relatively low f0 pitch value at onset. One interesting tonal error in American listeners was the confusion between T3 (relatively low f0) and T4 (relatively high f0), which are distinctive at their onset f0 value. Gottfried and Suiter explained that this type of error was related to the phonological change in the stimuli, since T3 was produced in the middle of a sentence, where it has a low-falling tone instead of the dipping-rising pattern in isolation. Therefore, when American listeners paid more attention to the movement/direction, they would confuse these two tones. In this study, Gottfried and Suiter also compared American listeners to native Chinese speakers. They stated that American listeners are less able to use acoustic information such as tone coarticulation context (f0 contour) to help identify target tones. Using similar manipulated stimuli, Lee, Tao, & Bond (2009) likewise investigated American listeners' perception of monosyllabic Mandarin minimal pairs contrasting in tone in intact, center-only, silent-center and onset-only syllables in isolation or with a precursor carrier phrase. Lee et al. also found T2 and T3 confusion as previous studies showed in isolation. They attributed the confusion to American listeners assigning more weight to f0 height than f0 direction when perceiving Mandarin T2 and T3 in isolation, which is consistent with Gottfried and Suiter (1997). Moreover, Lee et al. (2009) found that American listeners are less effective in making use of the extrinsic information (context) to help identify target tones when syllable-intrinsic information (f0) is absent or compromised in stimuli as compared to native Mandarin speakers.

Taken together, these studies show that when tones are in isolation in monosyllabic words, T2 and T3 are confusable and challenging for native English-speaking learners to perceive.

2.3 Native English-speaking learners' perception of Mandarin Chinese disyllabic tones

Understanding native English-speaking learners' perception of monosyllable tones is necessary and important since it is the very first, basic step of acquiring Chinese phonemic tones.

However, the majority of words in Mandarin are disyllabic (Zhou et al., 1999; Duanmu, 1999). Therefore, investigating how learners acquire disyllable Chinese words is critical.

Only a limited numbers of studies have investigated native English-speaking learners' perception of disyllable words (Sun, 1998; He, 2010; Hao, 2012, He & Wayland, 2013). He (2010), He and Wayland (2013) and Sun (1998) investigated the relationship between linguistic experience/proficiency levels and tonal perception of both monosyllabic and disyllabic words in Mandarin Chinese by native English-speaking learners, and their final results echo each other. These researchers found that across learning experience and proficiency level, native English-speaking learners did significantly better at identifying tones in monosyllabic words than in disyllabic words. Moreover, native English-speaking learners' accuracy rate of tonal perception was systematically improved according to their learning experience: the higher the proficiency level or the longer they studied Mandarin Chinese, the better their accuracy was.

When examining learners' identification performance of four phonemic tones across both monosyllabic and disyllabic words, Sun (1998) found that T2 and T3 across proficiency levels were identified significantly poorer than T1 and T4 across all four proficiency level groups. Sun (1998) also tested adult American listeners' perception of tones in disyllabic Chinese words in three word types: common, uncommon and nonsense words, and she found that American learners' tonal perception of nonsense disyllabic words was significantly worse than common and uncommon real disyllable words. This indicated that the familiarity with the disyllabic words helped learners better identify tones in the words that they knew rather than the words they did not hear before. Similarly, He (2010) found that, of all four tones, T3 was most difficult to identify, then T1, T2 and T4 by inexperienced learners while T2 was the most difficult to identify among the four tones by experienced learners across both monosyllabic and disyllabic tonal contexts.

At the disyllable word level, Sun (1998) also analyzed tones at two syllable positions, initial and final positions. She found that the accuracy rate of tone identification at the final position was better than at the initial position in all disyllabic words. In addition, T1 and T4 were identified with higher accuracy at both initial- and final- position than T2 and T3 in disyllable words. According to Sun, American listeners' perception was significantly better on final syllable due to word stress in disyllabic words that Sun chose in the study. In other words, final-syllable stress cues, which are more salient to perceive than the unstressed initial-syllable, helped learners identified tones in final position more accurately. This finding echoes those of He (2010) and He and Wayland (2013), who also found that in disyllabic tone perception, all four tones were identified with a higher accuracy in final syllable position than in initial syllable position by native English-speaking learners. He (2010) explained that the better identification of final syllable tone was probably due to the longer duration at the final syllable in natural productions.

He (2010) and He and Wayland (2013) also examined disyllable words in compatible and conflicting context environments (Xu, 1994) to see the effect of tonal coarticulation on native English-speaking learner's tonal identification task. She found that learners' tonal perception of disyllabic words was significantly better in compatible contexts than conflicting contexts, and T3 was still the worst among both tonal environments across four tones in identification. He (2010) analyzed two types of errors that affected learners' perception—tonal direction misperceptions and tonal height misperceptions. According to He (2010), inexperienced learners tended to make more tonal directional errors due to their little experience with tonal coarticulation in disyllables.

For example, the T4 + T2 tonal combination, in which the offset of preceding tone (T4) and onset of the following tone (T2) differ greatly, exhibits a big change in the direction of f0 contour. Bi-directional T2-T3 confusion was also observed in American learners' identification tasks on both mono- and di-syllable words of Mandarin by Hao (2012). According to Hao, the major difficulty shown in perception and production tasks appeared to be caused by American learners' lacking the association between the pitch of a tone and its corresponding tonal category. Thus, building up native English-speaking learners' phonetic tonal categories of Mandarin Chinese might be the first step towards achieving native-like pronunciation in the target language.

From the above four studies, it seems that English-speaking learners' identification performance of tone perception on both monosyllabic and disyllabic words can only be improved with an increase in linguistic experience. Learners struggle with poor pronunciation at the beginning stage of the learning. In current college level Mandarin Chinese classes in the United States, this stage is usually defined as the first year of learning. Meanwhile, tone pronunciation is often introduced to native English-speaking learners only for a few weeks (Xing, 2006; Orton, 2013) at this beginning stage. These beginning learners may habitually and repeatedly make the same pronunciation errors without much training and feedback due to the lack of emphasis on tone learning in general. For this reason, a very harmful consequence—fossilization (Selinker, 1972) of the incorrect tone pronunciation could potentially develop. For learners who have reached fossilization, their tone pronunciation will be very difficult to correct in the future because of the habitual and repeated incorrect tone pronunciation that they perceived and produced at the beginning of learning. In fact, Orton (2013) witnessed such learners in her study, who even at the fourth or fifth year of their language learning still felt incompetent to communicate in Mandarin Chinese due to poor pronunciation. From the anecdotal experiences of many leaners in Orton's study, she found that inability to perceive or produce tones correctly often leaves learners feeling miserable. Such a feeling could possibly lead to frustration and helplessness in learning. What is worse is that some learners will give up learning Chinese, which is the last thing any language teacher or language program would like to see.

With this in mind, again the importance and urgency of building up native English-speaking learners' tonal categories in Mandarin Chinese from the very beginning of learning the language is evident. Current in-classroom tone teaching should not only pay attention to monosyllabic tone practice but also give more attention to disyllabic tone practice, including tone alternation and coarticulation among the two adjacent tones. These high variability and coarticulated tones regularly occur in Mandarin Chinese natural speech, and by focusing on disyllabic words, English speakers may be able to improve their perception of tones.

2.4 High variability phonetic training

Research has shown that Mandarin monosyllabic tones can be improved through a short perceptual training in a computer lab at learners' convenience (Wang et al. 1999, 2003). High variability phonetic training has proven an effective method for improving learners' perception and production of both segmental and suprasegmental properties in the target language. Significant improvement has been reported cross-linguistically in many studies (Logan, Lively, & Pisoni, 1991; Lively, Logan, & Pisoni, 1993; Yamada, Yamada, & Strange, 1996; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Wang et al. 1999, 2003; Iverson, Hazan & Bannister, 2005; Herd, Jongman & Sereno, 2013).

High variability phonetic training was first proposed by Logan et al. (1991) in training Japanese learners to differentiate between /r/ and /l/ in English. This type of training includes the following: stimuli are presented in a variety of phonetic environments; natural speech tokens are used instead of synthesized ones; and multiple speakers are used. These characteristics converge to enable listeners to form robust phonetic categories by increasing stimulus variability (Logan et al., 1991; Lively et al., 1993). Logan et al. found that Japanese learners learned to differentiate English contrast /l/ and /r/ perceptually after a short 3-week high variability training (1991). The result of this study also suggested that using natural speech tokens as stimuli instead of synthetic stimuli (Strange and Dittmann, 1984) helped learners not only learn the new contrast, but also generalize to new talkers and new stimuli. To tease apart the effect of training, Lively et al. (1993) conducted two types of training with emphasis on two different procedures: one group was trained with a single-talker and five different phonetic environments while the other group was trained with multiple talkers and only three phonetic environments (1993). Comparing these two training results, the multi-talker group performed better than the single-talker group despite being exposing to fewer phonetic environments of the target contrast. These results suggested that talker variability plays an important role in perceptual learning and formation of a robust target category.

These previous studies showed a significant improvement on leaners' perception in the identification and discrimination of target phonetic contrasts. Some studies even further extended learners' perceptual improvement to their production ability. At the segmental level, Bradlow et al. (1997) conducted a perceptual training of the English /r/-/l/ contrast for Japanese listeners by using a high-variability phonetic training technique. This training technique involved natural recording of minimal pairs in the target contrast by multiple native English speakers, at multiple syllable-positions and various linguistic environments (such as word-initial, word-medium, word-final, initial cluster, and final cluster). The results showed that within 3-4 weeks of training, Japanese listeners showed substantial improvement in identification accuracy of /r/-/l/ contrast. Furthermore, this progress in perceptual abilities transferred to their production. Bradlow et al. (1997) concluded that their findings supported the hypothesis that language learning in perception and production are closely linked, since perceptual learning of the /r/-/l/ contrast transferred to the production domain.

Iverson, Hazan, and Bannister (2005) compared the effectiveness of four different training techniques for teaching English /r/ and /l/ contrast to Japanese adult learners. These training techniques included high variability phonetic training by using natural words and multiple talkers, and the other three techniques in which the natural production were altered by manipulating various acoustic cues, such as F2 (second formant frequency), F3 (third formant frequency) and duration. The training period was about 2-3 weeks. Results showed that all four training methods improved learners' perception of the target /l/-/r/ contrast, and there were no difference between these techniques. From the perspective of L2 phoneme learning, Iverson et al. suggested that high variability phonetic training with natural speech seems to be the best method among the four training techniques due to the minimal labor required when setting up an experiment. In addition, Herd et al. (2013) compared three training modalities within the high variability phonetic training method, including perception training only, production training only, and a combination of perception and production training, to see which modality would help

American learners to improve their Spanish intervocalic sound /d, r, ɾ/ in both perception and production. Herd and colleagues found that all three training modalities were effective, in which both perception-only and production-only trainees made primarily gains in perception, and the combination trainees made gains in production. This indicates that high variability phonetic training is the most effective method to help the learners improve their target language's segmental acquisition.

High variability phonetic training is not only proven to be effective at the segmental level but also at the suprasegmental level. It has been shown to improve perception of Mandarin monosyllable tonal categories and these gains are retained for at least 6 months (Wang et al. 1999, 2003). Through a three-step design (pretest, training and posttest), with eight 40-minute training sessions, Wang et al. (1999) successfully helped American learners of Mandarin Chinese improve their tone perception on monosyllabic words, from pretest accuracy rate of 69% to posttest's accuracy rate of 90%. This pre- to post-test improvement (21%) was significant. Furthermore, they then tested the trained American learners tone perception with new stimuli by a new speaker. The trainees performed significantly better on all tests than the control group who hadn't received any training, showing a generalization of the learning to new words and new speakers. This training effect was also retained after six month when trainees were tested again in an identification task on monosyllabic tones. In Wang et al. (2003), the researchers extended their training effect from American learners' tone perception to tone production. Before learners' perception pretest and after their post-test, trainees were recorded producing a list of Mandarin words. Their production performance was not only judged by native Mandarin speakers but also analyzed acoustically by comparing learners' pitch contours to native productions. The results showed that identification of trainees' post-test tone productions improved by 18% from pre-test productions and the learners' pitch contours approximated native norms. This indicated a significant tone improvement after the short perceptual training.

These studies demonstrate that high variability perceptual training is an effective training method to improve nonnative learners' perception and production in segmental (English /r/-/l/ contrast, Spanish intervocalic sound /d, r, ɾ/) and suprasegmental (Mandarin four monosyllabic tones) features in a target language. Therefore, my proposed perceptual training study is designed using this high-variability phonetic training paradigm for disyllabic Mandarin words. The current study aims to find out if the established perceptual training method will function effectively in training native English-speaking listeners to accurately perceive more naturalistic disyllabic words, which involve tone coarticulation. Monosyllabic and disyllabic training will be compared in order to determine the amount of improvement in tone identification. In addition, both monosyllabic and disyllabic stimuli will be examined to determine which type of training material is more effective at helping native English-speaking learners to shape tonal categories that do not exist in their phonological inventory.

2.5 Research Questions

Following the review of previous studies, the present study aims to answer the following research questions:

RQ1. After perceptual training, will native English-speaking learners improve their perception of tones generally in both monosyllabic words and disyllabic words in Mandarin Chinese?

RQ2. Compared to monosyllabic perceptual training, will disyllabic perceptual training be more or less effective in helping English-speaking learners shape their tonal categories in their tone perception of Mandarin Chinese?

RQ3. Contrasting two types of training materials, monosyllabic stimuli and disyllabic stimuli, which is more effective in helping learn monosyllabic tones? And which is more effective in helping learn disyllabic tones?

RQ4. Will training using monosyllabic material transfer to disyllabic tone identification? And will training using disyllabic material transfer to monosyllabic tone identification?

RQ5. Will factors, specifically syllable position, tonal context, and tonal sequence, affect native English-speaking learners' tone perception of disyllabic words?

2.6 Hypotheses

Hypothesis 1: It is hypothesized that both monosyllabic perceptual training and disyllabic perceptual training will help improve native English-speaking learners' tonal perception in Mandarin Chinese.

Hypothesis 2: When compared to monosyllabic perceptual training, disyllabic perceptual training is hypothesized to help native English-speaking learners more.

Hypothesis 3: When contrasting two types of training stimuli, it is hypothesized that monosyllable training stimuli may help improve learners' perception of monosyllabic tones more. On the other hand, it is hypothesized that the highly variable and coarticulated disyllable training stimuli may help improve learners' identifying disyllabic tones more.

Hypothesis 4: Examining the transfer effect of training, it is hypothesized that there may be a transfer of learning in both directions. That is to say, monosyllabic training may help identify tones in disyllable stimuli and disyllabic training may also help identify monosyllabic tones. However, the learning effect from disyllabic training stimuli to monosyllabic tone identification may be greater because of the beneficial effect of high variability and tone coarticulation present in disyllabic stimuli. Therefore, disyllabic training may be more effective than monosyllabic training in improving English speakers' tone perception.

Hypothesis 5: Regarding three linguistic factors, such as syllable position (tone on the initial syllable versus tone on the final syllable); tonal context (compatible tones versus conflicting tones), and tonal sequence (same versus different), it is hypothesized that tone on the initial syllable may be more difficult to identify correctly than tone on the final syllable within a disyllabic word. Also, disyllabic stimuli in compatible tonal contexts might be easier for English-speaking learners to perceive than in conflicting tonal contexts. Finally, contrast to He (2010), who claimed that there was no difference between tones in the same tonal sequence versus tones in the different tonal sequence, the current study hypothesizes that tonal sequences in which the same tone was repeated are predicted to be identified more accurately than sequences with different tones.

Chapter 3: Chapter Three: Methods and experimental design

The current perceptual training experiment was designed to be similar to the early perceptual training in Mandarin Chinese tones by Wang et al. (1999, 2003). In their studies, a high variability training procedure was used to achieve significant learning of four individual phonemic Mandarin tones by American learners of Chinese. In the present study, though, monosyllabic training was contrasted with disyllabic training to determine whether introducing different types of training and, more importantly, more variable training materials, would facilitate learning of Mandarin tones.

The goal of this experiment was to determine which perceptual training (monosyllabic or disyllabic) and which training material (monosyllable stimuli produced in isolation or disyllable stimuli produced in connected speech) would help native English-speaking learners of Chinese to improve their perception of Chinese words.

Beginning native English-speaking learners of Mandarin Chinese at the college level were recruited to participate voluntarily in the study. The perceptual training included three phases: pretest, training, and posttest. Both tests and the training were conducted at the Phonetics and Psycholinguistics Laboratory at the University of Kansas. First, all participants took a pretest. The duration of the training phase lasted two weeks. Afterwards, they all completed a posttest. The posttest also included a generalization test in order to investigate any perceptual improvements due to the training.

Two training groups were contrasted based on whether they were trained on monosyllable stimuli or disyllable stimuli. Both groups participated in identical pre- and post-tests, and the generalization test. The group with monosyllable training was trained only in naturally produced monosyllable words in isolation that covered all possible phonetic environments in Mandarin Chinese, which were adopted from the training stimuli in Wang et al. (1999). A second training group was trained only in naturally produced disyllable Chinese words. Disyllabic words have not been used before in any previous training studies. The motivation of using disyllabic Chinese words as stimuli was due to the following reasons. First, disyllabic words provide more tonal variation in the stimuli, similar to natural speech. In addition, such stimuli are embedded with information about tonal coarticulation, which are also present in tones that occur in natural connected speech.

During the training sessions, immediate feedback was given to the learners in order to help them focus their attention on the critical acoustic cues of the four tones either in monosyllable or disyllable words in a consistent manner from trial to trial.

For training, stimuli with the four Mandarin tones were presented in a variety of phonetic contexts in the experiment, and were produced naturally by native Chinese speakers of both genders.

A forced-choice identification (ID) task was used throughout the entire procedure, including pre- and post-tests, trainings, and the generalization test. Previous studies have shown that the nature of the ID task during testing and training helps language learners to maintain a consistent mapping between the stimuli and the target phonemic contrasts (Logan, et al., 1991; Bradlow et al., 1999).

The two different training groups' performance in pretest and post-test were compared to observe any improvement after the training. In addition, the performance for the two types of training material (monosyllable and disyllable training stimuli) were examined to determine which type of training material showed the most learning improvement. The generalization test contained new stimuli spoken by a new native Chinese speaker who was not recorded in the training sessions. This design choice helped determine whether learners' perception of the four phonemic Chinese tones can be generalized both to novel, as well as to speakers that not heard before.

3.1 Participants

Two groups of participants were recruited in this study.

1. Native English-speaking learners of Chinese

Seventeen native English speakers were participants in the perception training experiments. They were all beginning learners of the Chinese language with less than two semesters (less than 7 months) of learning Mandarin. Native English speakers were randomly assigned to one of the aforementioned groups: Nine in the Monosyllabic Training Group and eight in the Disyllabic Training Group. None of these seventeen learners had any history of hearing, speech, or language difficulties. All were college students and had studied at least one foreign language in high school (most often French or Spanish). Due to sickness, one subject in the Monosyllabic Training Group withdrew from the study after finishing pretest, training and posttest, not the generalization test. Therefore, this subject's performance was only reported in pretest and posttest results, but not in generalization test results.

Prior to any test or training sessions, all participants completed a human consent form. A background questionnaire was given to ascertain information about age, gender, and any knowledge of other languages.

2. Native Chinese participants (speakers)

Eight native speakers of Mandarin Chinese were recruited to produce the stimuli for the perception experiments.

Production of stimuli: Six native speakers of Mandarin Chinese were recorded for all stimuli used in the experiment, three males and three females. Native Chinese speaker One, a male, produced the pre- and post-test stimuli. Speakers Two (female), Three (male), Four (female), and Five (male) produced stimuli for the two different training sessions. Native Chinese speaker Six, a female, read the generalization test stimuli. To preserve the characteristics of disyllable words in connected speech, all six speakers were instructed to produce the stimuli as natural as possible, and to avoid producing any disyllable stimuli as two separate individual syllables (Xu, 1994). Prior to recording, the native Chinese speakers completed a human consent form. A background questionnaire was also given to obtain information about age, gender, and knowledge of other languages.

Perception of stimuli: Two additional native Chinese listeners (one male and one female) served as the judges for assessing the intelligibility of all the recorded stimuli used in perception study. They listened to each stimulus and determined whether the recorded stimuli were clear and intelligible productions of the Mandarin words. For the female listener, identification accuracy was 99% for all stimuli and all speakers; for the male listener, identification accuracy was 98% for all stimuli and all speakers. Prior to any evaluation of the stimuli, both participants also completed a human consent form, and a background questionnaire to acquire information about age, gender, and knowledge of other languages.

3.2 Stimuli

Two types of stimuli, monosyllabic stimuli and disyllabic stimuli, were used throughout the pretest, training, and posttest. All monosyllabic stimuli were adopted from Wang et al. (1999). These monosyllabic stimuli included all possible permissible combinations of various initial consonants and final vowels, and different syllabic structures in Mandarin Chinese (i.e. V, CV, CVNasal, VN, CGlideV, and CGVN). Contrastively, each disyllabic stimulus was composed of two randomly combined syllables from the monosyllabic stimuli. Thus, every individual syllable used for the disyllabic stimuli was identical to those used in the monosyllabic stimuli. For example, the monosyllabic stimuli "*mǎ*" (*"horse"*) and " *shāng*" (*"injury"*) were combined to form a two-syllable word that served as a disyllabic stimulus, "*mǎ shāng*" . All monosyllabic stimuli were real words in Mandarin Chinese; the randomly combined disyllabic stimuli were non-words with a decomposable meaning.

All the stimuli were recorded by six native Mandarin Chinese speakers, three males and three females, in order to ensure speaker variability.

3.2.1 **Pretest Stimuli**

a) Pretest monosyllabic stimuli. Stimuli in the monosyllable pretest were the same 96 monosyllabic stimuli used in the pretest by Wang et al. (1999) study. There were 24 monosyllable words for each of the four phonemic Mandarin tones.

b) Pretest disyllabic stimuli. The 48 disyllabic stimuli shared identical syllables as those in the monosyllabic pretest. There were 3 disyllable words for each of the 16 tone combination.

3.2.2 **Training Stimuli**

a) Training monosyllabic stimuli. There were 128 monosyllabic training stimuli, which consisted of 32 monosyllable words for each of the four tones. Since four native Chinese speakers (speaker Two, Three, Four and Five) produced these stimuli, there were 512 monosyllabic stimuli in the monosyllable training sessions.

b) Training disyllabic stimuli. 64 disyllabic stimuli were used in training, and these stimuli shared the same syllables as those in the monosyllabic training stimuli. The same four native Chinese speakers (speakers Two, Three, Four and Five) produced these 64 stimuli, thus, there were 256 disyllabic training stimuli.

3.2.3 **Posttest Stimuli (same as Pretest stimuli)**

a) Posttest monosyllabic stimuli. The posttest stimuli were identical to the 96 monosyllabic stimuli used in the pretest.

b) Posttest disyllabic stimuli. The posttest stimuli were identical to the pretest 48 disyllabic stimuli.

c) Generalization test (GT) monosyllabic stimuli. 64 new monosyllabic stimuli never appearing in the previous tasks were used in the monosyllable generalization test. These were produced by female native Chinese speaker Six.

d) Generalization test disyllabic stimuli. There were 32 new disyllabic stimuli that shared the same 64 syllables in the monosyllabic generalization test. These stimuli were also produced by speaker Six.

In total, there were 288 monosyllabic stimuli and 144 disyllabic stimuli in the current experiment.

3.3 Procedure

The present experiment consisted of three phases: pretest, training, and posttest (including the generalization test). Both the tests and training were conducted on computers in the KU Phonetics and Psycholinguistics Laboratory. Seventeen native English-speaking learners participated in the two week training program. Each learner participated for a total of six days for the entire experiment (Pretest; Training Day 1; Training Day 2; Training Day 3; Training Day 4; Posttest and Generalization test). Each training session was 30 minutes long. All stimuli were randomized using a forced-choice perceptual identification task presented in Paradigm (Tagliaferri, 2008).

3.3.1 **Pretest**

Learners in both training groups participated in the pretest. The pretest consisted of two parts, a monosyllable word identification task and a disyllable word identification task. During both tasks, all learners provided their best judgments indicating on a computer keyboard which Mandarin Chinese tone(s) they hear. The pretest lasted about 60 minutes, approximately 30 minutes for each task.

3.3.1.1 ***Monosyllabic Pretest***

In the monosyllable word identification task, the learners first heard a monosyllable stimulus from the computer through headphones, and were instructed to give their tone identification response by pushing the corresponding button that represented one of the four tones (1=T1, 2=T2, 3=T3, and 4=T4). All tonal diacritics and numbers were labeled on the buttons on the keyboard. There were 96 stimuli in the pretest for the monosyllable identification task. All monosyllabic stimuli were presented with a 3 second inter-trail interval (ITI). No feedback was given in the pretest. Learners' reaction time and accuracy during the identification task were recorded in Paradigm.

3.3.1.2 ***Disyllabic Pretest***

After a ten minute break, the learners participated in the disyllable word identification task. In this second task, participants heard a disyllable stimulus from the computer, and they were asked to give their tone identification response by pushing two corresponding buttons (one after another) on the computer keyboard that represented the tone of the first syllable followed by the tone of the second syllable. There were 48 disyllable stimuli in pretest for disyllable identification task, and the ITI was 3 seconds as well. All disyllable tonal diacritics and numbers were labeled on the buttons. No feedback was given in this pretest. Learners' reaction time and accuracy in the identification task were recorded in Paradigm.

3.3.2 **Training**

Seventeen native English-speaking learners of Mandarin Chinese participated in the two week training program. Nine learners participated in the monosyllable training group, and the other eight participated in the disyllable training group. Both Monosyllabic and Disyllabic training consisted of four perceptual training sessions that lasted 30 minutes each. Learners were then asked to participate in a forced-choice ID task. Immediate feedback was after each response for all training sessions (see details in feedback in two types of trainings below).

3.3.2.1 ***Monosyllabic training***

The monosyllabic training group was trained exclusively with monosyllabic stimuli. There were 512 stimuli in the monosyllable training produced by four native Chinese speakers. In each session, the trainees were trained only auditorily with the stimuli produced by one speaker. For instance, the participant heard a stimulus, *"má"*, which contained a target tone (T2) in a monosyllabic word, and he/she then made the best choice among four tones (1=T1, 2=T2, 3=T3, and 4=T4) by pushing the corresponding button (2 in this case) on the computer keyboard. If the choice was correct, the participant would hear, "Correct! That was Tone 2, it is *má*." The trial then proceeded to the next stimulus. If the response was incorrect, the participant would hear, "Uh-oh! That was má, Tone 2. Let's hear it again *má*". With incorrect responses, training proceeded only after feedback.

Each training session was followed by a test containing the re-randomized trained stimuli produced by the same speaker. No feedback was given. Four such training assessments were given to the learners.

3.3.2.2 ***Disyllabic training***

The disyllabic training group was trained auditorily only with disyllable stimuli. There were 256 disyllable stimuli in the four training sessions. In each session, the learners heard stimuli only produced by one speaker. For example, the learner heard a disyllabic stimulus, *"mǎ shāng"*, which was a Tone 3 + Tone 1 combination. The learner would then make two responses by pushing two buttons (here 3 and 1) on the computer keyboard. Immediate feedback was given just as in the monosyllabic training. For instance, if the choice was correct, the participant would hear, "Correct! That was Tone 3 and Tone 1, it is *mǎ shāng*." The trial then presented the next stimulus. If the response was incorrect, the participant would hear, "Uh-oh! That was *mǎ shāng*, Tone 3 and Tone 1. Let's hear it again *mǎ shāng*. " After feedback, the trial continued.

Similar to the monosyllabic training assessment, there was an assessment test at the end of each training session, consisting of re-randomized trained stimuli produced by the same speaker. Therefore, four disyllabic training assessment tests were given to the learners without feedback.

3.3.3 **Posttest**

After the training sessions, both groups took the posttest, which was identical to the pretest (with re-randomized stimulus presentation) for both the monosyllabic test and the disyllabic test. No feedback was given for the posttest and it took approximately one hour to complete.

3.3.4 **Generalization Test**

Immediately after the posttest, the learners took a generalization test that contained two parts: the monosyllabic test and the disyllabic test, in which new stimuli were produced by a female speaker who they had not heard before. A ten minutes' break was given between posttest and generalization test.

3.3.5 **Data analysis**

The statistical design of the present study included the dependent variable: tone identification accuracy which includes monosyllable stimuli tone accuracy (correct or incorrect) and disyllable stimuli tone accuracy (when both tones were correct, then considered as one correct item). There are four dependent variables: test (pretest, posttest, and generalization test), training group (monosyllabic training group and disyllabic training group), stimuli (monosyllable stimuli and disyllable stimuli), and tone (T1, T2, T3, and T4). Analyses of the dependent variables were conducted to determine if there were significant differences between the two training groups in identification of the two types of stimuli from pretest to posttest.

A repeated measures ANOVA and Paired Sample t-test were used in the study to compare accuracy of learners' responses in the tests. All statistical analyses were performed by using software SPSS. All the *p*-values and the *F*-values were adjusted by using the Greenhouse-Geisser correction (Greenhouse and Geisser, 1959), and post-hoc pairwise comparison and paired t-tests were adjusted by using the Bonferroni correction ($p<.05$). All significant results and results that are marginally significant $p < .10$ were reported.

Chapter 4: Chapter Four: Results and Findings

This chapter includes two main parts: results and findings from pretest to posttest, and results and findings from the generalization test. Due to the difference of the nature in monosyllable stimuli and disyllable stimuli, the learners' tonal performance in each stimuli type were analyzed in pretest, posttest and generalization test separately. The effect of three linguistic factors on the learners' tonal perception was investigated in the disyllable stimuli results. Moreover, the learners' tone confusion in both types of stimuli were also reported in order to examine the most and least confusable tones in their tonal perception, as well as the improvement of these tone pairs from pretest to posttest.

Repeated measures ANOVAs and Paired Sample t-test were conducted to analyze the results in all tests. Again, the p-values and the F-values were adjusted by using the Greenhouse-Geisser correction (Greenhouse and Geisser, 1959), and post-hoc pairwise comparison and paired t-tests were adjusted by using the Bonferroni correction ($p<.05$). All significant results and results that are marginally significant $p < .10$ were reported.

4.1 Overall improvement from pretest to posttest

Listeners' accuracy on monosyllable stimuli and disyllable stimuli from the two training groups at pretest and posttest are displayed in Figure 2.

The overall results were analyzed in a three-way repeated measures ANOVA, with Test (pretest, post-test) and Stimuli (monosyllable stimuli, disyllable stimuli) as within-subjects factors and Training Group (monosyllabic training group, disyllabic training group) as a between-subjects factor.

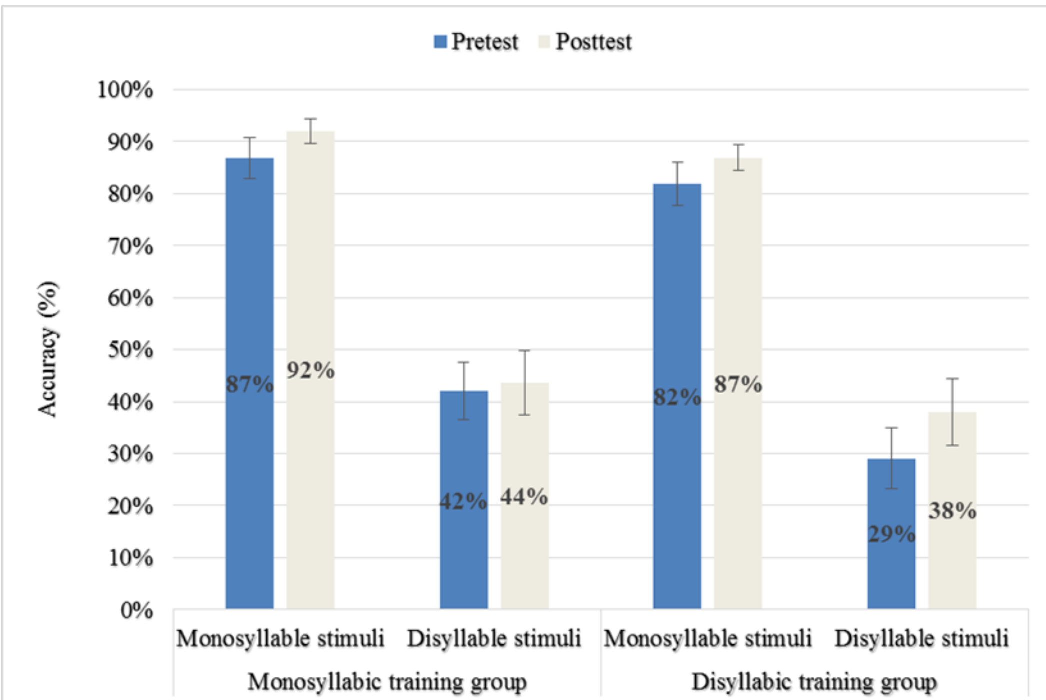


Figure 2: Accuracy rate and standard errors (SE) of monosyllable and disyllable stimuli by native English-speaking learners of Chinese in monosyllabic and disyllabic training groups in pretest and posttest

The analysis yielded a significant main effect of Test [$F(1, 15) = 16.225$, p=.001], which indicated that the native English-speaking learners' performance on tone identification, averaged across both groups and all stimuli, was significantly different from pretest to posttest. Learners did significantly better in their posttest at a 65% accuracy rate compared to a 60% accuracy rate in pretest. This significant 5% difference showed that there was training effect on the learners' tonal perception in the posttest as in Figure 3 .

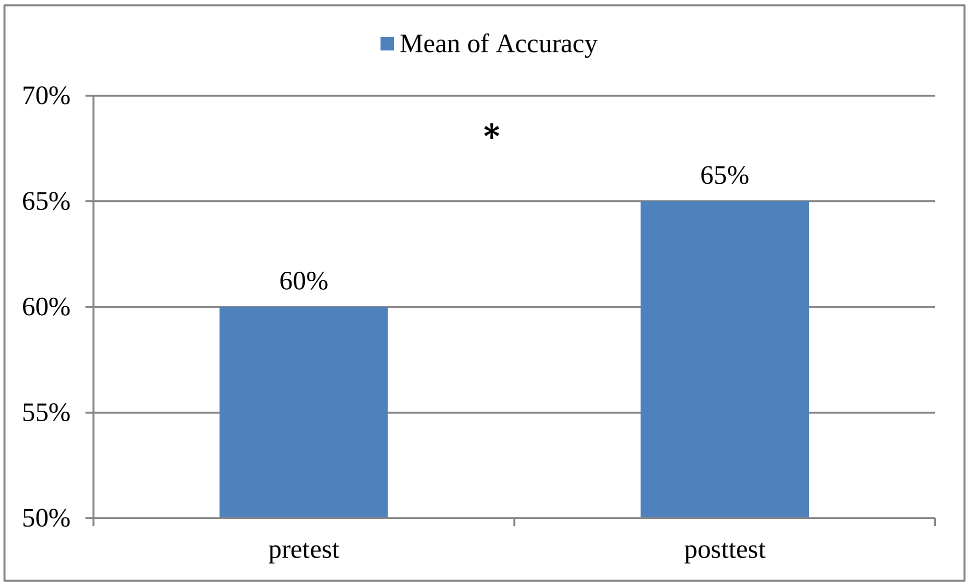


Figure 3: Mean of accuracy in the pretest and posttest by native English-speaking learners.

The main effect of Stimuli [$F(1, 15) = 354.929$, $p<.001$] showed that there was a significant difference between monosyllable stimuli and disyllable stimuli in learners' tone identification averaged across the two tests and the two training groups. The learners identified tones significantly better in monosyllable stimuli (87%) than tones in disyllable stimuli (38%).

There was no main effect for Training Group [$F(1, 15) = 1.270$, $p=.277$]. This suggested that the learners behaved similarly in the two training groups when they identified tones in monosyllabic and disyllabic stimuli across pre- and posttest.

There was no significant interaction of Stimuli X Training Group [$F(1, 15) = .512$, $p=.485$], Test X Stimuli [$F(1, 15) = .000$, $p=.989$], Test X Training Group [$F(1, 15) = 2.356$, p=.145], nor was there a three way interaction of Test X Stimuli X Training Group [F(1, 15)=1.682, p=.214].

Figure 2 does show that there was some numerical increase in accuracy rate in learners' tonal performance when examining the monosyllabic training group and the disyllabic training group from pretest to posttest across all stimuli.

The accuracy of the monosyllabic training group increased 4% from pretest 64% to posttest 68%. The disyllabic training group accuracy rate increased 8% from pretest 55% to posttest 63%. Two paired sample t-tests, one for the monosyllabic training group and the other for the disyllabic training group, were conducted to compare the learners' improvement on tonal perception from pretest to posttest. For the monosyllabic training group, there was a significant difference in accuracy from pretest to posttest, $t(8)= -3.83$, $p=.005$. There also was a significant difference from pretest to posttest For disyllabic training group, $t(7)= -2.86$, $p =.002$. These results indicated that both types of training were effective in helping English-speaking learners improve their tonal perception.

The two groups' performance on monosyllable stimuli and disyllable stimuli were analyzed separately in order to find out which training group, monosyllabic or disyllabic, helped learners more in tone identification of monosyllable words and disyllable words respectively in Mandarin Chinese.

4.1.1 **Monosyllable stimuli from pretest to posttest**

To determine which training group helped learners more in tone identification of monosyllable words in Mandarin Chinese, the two training groups' performance on monosyllable stimuli were analyzed using a repeated measures ANOVA.

A two-way repeated measures ANOVA, with Test (pretest and posttest) as a repeated measure, and Training Group (monosyllabic training group, disyllabic training group) as a between-subjects factor, showed a significant main effect of Test [$F(1,15)=13.166$, $p=.002$]. It demonstrated that there was a significant difference across groups from pretest to posttest. Averaged across two training groups, learners did significantly better in the posttest with an accuracy rate of 90% than in pretest with accuracy rate of 84% as shown in Figure 2. Such results suggested that perceptual training indeed improved learners' monosyllable tone identification from pretest to posttest after four short training sessions.

The two-way ANOVA revealed that there was no main effect of Training Group [$F(1,15)=.971$, $p=.340$], which is to say that there was no significant difference between the two training groups across tests. Learners trained on disyllabic stimuli did equally well to those trained on the monosyllabic stimuli when identifying monosyllable tones.

There was no interaction of Test X Training Group [$F(1,15)=.344$, $p=.566$], suggesting that learners in both training groups showed a similar pattern in their tonal identification in monosyllable stimuli from pretest to posttest.

4.1.1.1 ***Individual Tones in monosyllable stimuli***

The accuracy rates of the four individual tones and tone confusions within monosyllable stimuli were analyzed to see whether there was any difference in perception of the four phonemic tones by learners in the two training groups.

Native English-speaking learners' tone identification performance of four individual tones in monosyllable stimuli in pretest and posttest are presented in Figure 4 (monosyllabic training group) and Figure 5 (disyllabic training group) below.

A three-way ANOVA, with Test (pretest and posttest) and Tone (T1,T2,T3,T4) as repeated measures, and Training Group (monosyllabic training group, disyllabic training group) as a between-subjects factor, revealed a main effect of Test [$F(1,15)=12.653$, $p=.003$]. This suggested that across groups learners were significantly better in identifying all four tones in monosyllable stimuli in posttest (90%) than pretest (84%) after training.

A main effect of Tone [$F(3, 45)=8.221$, $p<.001$] was also found, indicating that there was a significant difference among the four tones in monosyllable stimuli. A post hoc pairwise comparison with Bonferroni correction revealed that, in monosyllable stimuli, T4 (96%) was significantly better than T1 (86%) ($p=.029$), and T2 (84%) ($p=.005$), and T3 (84%) ($p<.001$). Additionally, there were no significant differences among T1, T2 and T3 ($p>.999$).

No main effect of Training Group [$F(1, 15)=1.022$, $p=.328$] was found, neither were there any two-way interactions of Test X Training Group [$F(1, 15)=.110$, $p=.745$]; Tone X Training Group [$F(3, 45)=.763$, $p=.521$]; or Test X Tone [$F(3, 45)=2.062$, $p=.119$].

The results yielded a trend of a three-way interaction of Test X Tone X Training Group [$F(3, 45)=2.175$, $p=.104$]. Two separate 2-way repeated measures ANOVA by two different training groups were conducted to tease apart this interaction.

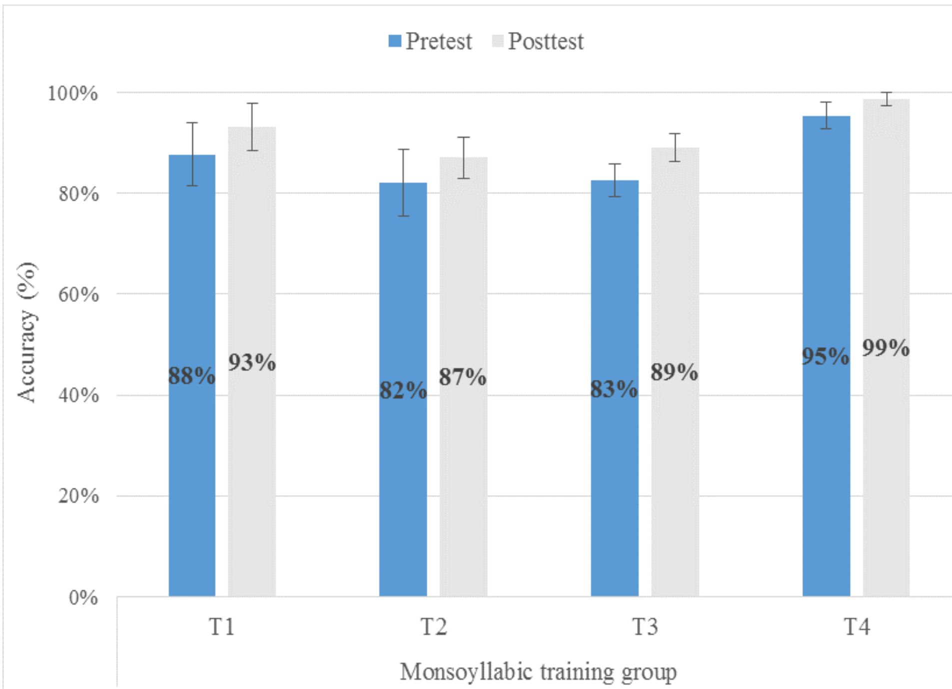


Figure 4: Accuracy rate and standard errors (SE) of monosyllable stimuli by native English-speaking learners of Chinese in monosyllabic training group in pretest and posttest.

The first 2-way ANOVA, as shown in Figure 4, used Test (pretest and posttest) and Tone (T1, T2, T3, T4) as repeated measures by the monosyllabic training group to examine the source of the interaction. The analyses showed a main effect of Test [$F(1, 15)=14.791$, $p=.005$]. Learners in the monosyllabic training group, across all four tones, did significantly better in their posttest (92%) than pretest (87%) on monosyllable stimuli after training.

A main effect of Tone [$F(3, 24)=5.106$, $p=.007$] was also found, which revealed that across both tests, there was a significant difference among these four tones. Pairwise comparison with Bonferroni correction demonstrated that monosyllabic learners identified T4 (97%) better than T3 (86%) (*p*=.035), and marginally better than T2 (85%) (*p*=.065) in monosyllable stimuli. However, there was no interaction of Test X Tone [F(3, 24)=.219, p=.882]. Comparison showed that learners in the monosyllabic training group made equal amount of improvement on all four tones from pretest to posttest.

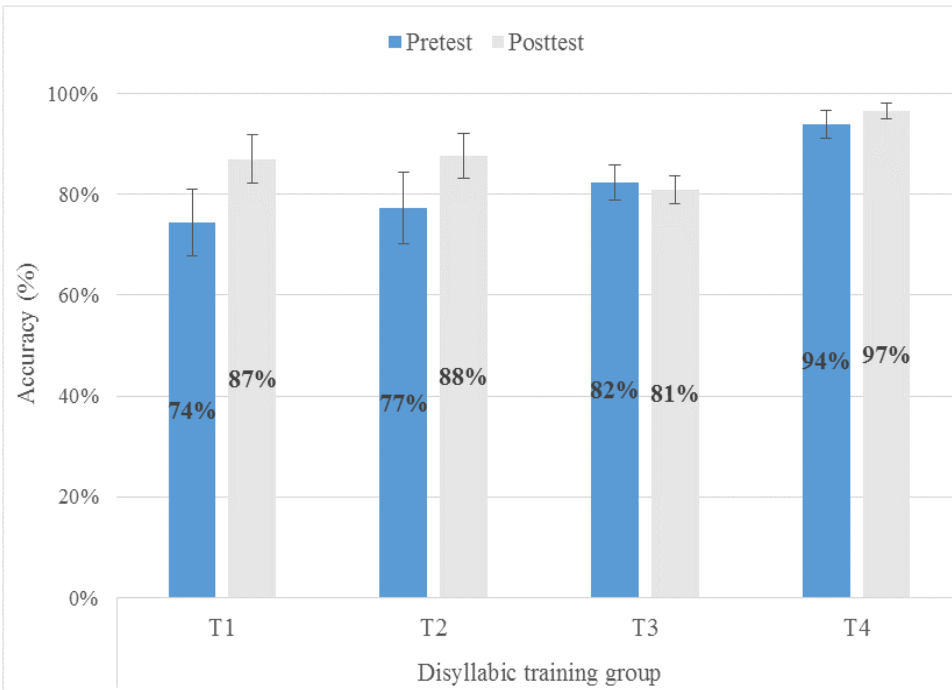


Figure 5: Accuracy rate and standard errors (SE) of monosyllable stimuli by native English-speaking learners of Chinese in disyllabic training group in pretest and posttest.

The second 2-way ANOVA, which results are shown in Figure 5, tested the effects of Test and Tone in the disyllabic training group. It yielded a trend for Test [$F(1,15)=4.162$, $p=.081$], a main effect of Tone [$F(3, 21)=3.951$, $p=.022$], and an interaction of Test X Tone [$F(3,21)=3.162$, $p=.046$].

The main effect of Tone [$F(3, 21)=3.951$, $p=.022$] showed that averaged across the two tests, the disyllabic training group learners' tone identification showed significant differences. Pairwise comparison with Bonferroni adjustment revealed that among the four tones, T4 (95%) was significantly better than T3 (82%) ($p<.001$), but not significantly different than T2 (82%, $p=.213$), nor T1 (81%, $p=.216$). T1, T2 and T3 were comparable ($p>.99$)

The interaction of Test X Tone [$F(3,21)=3.162$, $p=.046$] demonstrated a significant difference in improvement from pretest to posttest in the four tones in the disyllabic training group. T1 accuracy rate increased from 74% in the pretest to 87% in the posttest; T2 increased from 77% to 88%; T3 decreased from 82% to 81%; and T4 increased from 94% to 97%. Paired t-test showed that T1 made significant improvement when compared to T3 ($p=.032$) and T4 ($p=.031$) after training. T2 made numerical improvement compared to T3 ($p=.121$) and T4 ($p=.134$). There was no significant improvement from pretest to posttest in T3 and T4. These improvements indicate that training in disyllabic stimuli improved the learners' tonal perception in T1 and marginally in T2.

Overall, there was a significant training effect in monosyllable stimuli by the monosyllabic training group from pretest (87%) to posttest (92%), and a marginally significant training effect by disyllabic training group (82% to 88%). Also, the disyllabic training group learners did significantly better on T4 than other three tones in monosyllabic stimuli. Moreover, in the disyllabic training group, the learners' tonal perception of T1 improved significantly after training. But in the monosyllabic training group, there was no significant difference in improvement of individual tones after training.

4.1.1.2 *Tone confusions in monosyllable stimuli*

Native English-speaking learners' tone confusion of four individual tones is presented in the Table 2 and Table 3. Error rates for each tone pair were investigated in two directions. For example, for tone pair T1 & T2, the percentage of errors for tone pair T1 and T2 represented the error rate in the direction when T1 was misidentified as T2; the percentage of errors for tone pair T2 and T1 represented the error rate in the other direction when T2 was misidentified as T1.

Table 2 shows the learners' tone confusion in pretest and posttest by the monosyllabic training group. There are total of 216 stimuli for each tone (24 monosyllables x 9 learners=216). Note that error numbers are converted to percentage.

Table 2: Confusion matrices of the four individual tones by the learners in the monosyllabic training group from pretest to posttest in percentage (some rows sum to 99% or 101% due to the rounding).

STIMULUS \ PERCEIVED	PRETEST MONOSYLLABLE STIMULI				POSTTEST MONOSYLLABLE STIMULI			
	MONO_TR_GROUP				MONO_TR_GROUP			
	T1	T2	T3	T4	T1	T2	T3	T4
T1	88	7	0	5	93	1	0	6
T2	4	82	11	3	4	87	6	3
T3	0	18	82	0	0	10	89	0
T4	2	3	0	95	1	0	0	99

From Table 2, in pretest, the most confusable tone pair was T2 & T3. There were 11% of T2 perceived as T3, and an even higher number of T3 perceived as T2 (18%) by learners in the monosyllable training group. This error rate substantially decreased after training. However, learners still had difficulty in distinguishing between T2 and T3 in the posttest, where 6% of T2 were perceived as T3, and 10% of T3 were perceived as T2.

Some tone pairs improved in one direction even though the error rate stayed the same in the other direction after training. For instance, with tone pair T1 and T2, learners perceived 7% of T1 as T2 in the pretest but the error rate decreased to 1% in the posttest. In the other direction T2 and T1, learners misidentified the same 4% of T2 as T1 in both pretest and posttest.

Similarly, for tone pair T4 & T2, 3% of T4 was misidentified as T2 in the pretest although this error rate decreased to zero after training in the posttest; in the direction of misidentifying T2 as T4, meanwhile, the error rate was 3% before and after the training.

Such tones pairs seem to improve in one direction while they resist improvement in the other direction by the learners in the monosyllabic training group. This provides evidence for asymmetrical tone confusion between these tone pairs.

Tone pair T1 & T4 did not have much change after training. A 2% of T4 was misidentified as T1 in pretest and 1% in posttest. In the other direction, a 5% of T1 was misidentified as T4 in pretest and 6% in posttest.

For some tone pairs, such as T1 & T3, T3 & T4, learners did not make any errors in both pretest and posttest. In other words, in monosyllable stimuli, the learners were able to distinguish T1 from T3, and T3 from T4 very clearly before and after training.

Table 3 shows tone confusion of monosyllable stimuli by the learners in the disyllabic training group. There are total of 192 stimuli for each tone (24 monosyllables x 8 learners=192). All tone confusions are presented as percentages.

Table 3: Confusion matrices of the four individual tones by the learners in the disyllabic training group (8 students) from pretest to posttest in percentage (some rows sum to 99% or 101% due to the rounding).

STIMULUS \ PERCEIVED	PRETEST MONOSYLLABLE STIMULI DI_TR_GROUP				POSTTEST MONOSYLLABLE STIMULI DI_TR_GROUP			
	T1	T2	T3	T4	T1	T2	T3	T4
T1	74	11	2	13	87	1	2	10
T2	7	78	10	5	3	88	7	2
T3	0	16	82	2	0	19	81	0
T4	1	4	1	94	2	0	2	96

From the error rates in Table 3, the most confusable tone pair for the learners in the disyllabic training group was T2 and T3. In the pretest, 10% of T2 were perceived as T3, and this error rate decreased to 7% after training. A 16% of T3 were perceived as T2 in pretest, and the error rate increased to 19% in the posttest. That is, after disyllabic training the confusability of perceiving a T3 as a T2 was not only persistent but also getting worse. This result indicated that disyllabic training did not help the learners' T3 identification in monosyllable stimuli very much.

Tone pair T1 & T4 showed asymmetrical tone confusion. In pretest, learners perceived 13% of T1 as T4, and the error rate decreased to 10% in posttest. In the other direction, however, the tone confusion was a lot lower. 1% of T4 were mistakenly perceived as T1 in the pretest and 2% in the posttest, which suggests that the learners in the disyllabic training group were successful in distinguishing T4 from T1.

Some tone pairs, however, showed improvement in both directions. For T1 &T2, learners misperceived 11% of T1 as T2 in the pretest, but this error rate decreased tremendously (to 2%) in the posttest, which provided evidence for great improvement after training. Similarly, 7% of T2 were misperceived as T1 in the pretest, and the error rate was reduced to 4% in the posttest.

Tone pair T2 & T4 also demonstrated symmetrical improvement after training. In one direction, 5% of T2 were misperceived as T4 in the pretest and 2% in the posttest; in the other direction, 4% of T4 were misperceived as T2 in the pretest and zero in the posttest.

Tone pairs of T1 and T3, T3 and T4 showed very low or zero error rates in both pretest and posttest, which indicated that learners in the disyllabic training group can distinguish these tones without much confusion. For instance, a 2% of T1 were misperceived as T3 in both pretest and posttest; also, only 1% of T3 were misperceived as T1 in pretest and zero in posttest. Similarly, a 2% of T3 were misperceived as T4 in pretest, and zero error rate after the training. In the other direction, 1% of T4 was perceived as T3 in pretest, and it was 2% in posttest.

Overall, from the above two training groups' results, it was clear that in monosyllable stimuli, the most confusing tone pair was T2 and T3, which were observed in both training groups from pretest to posttest. Moreover, the monosyllabic training group learners' T2 and T3 confusion was improved after training in both directions (T2 and T3: 11% vs. 6%; T3 and T2: 18% vs. 10%), but the disyllabic training group learners' T2 and T3 confusion was improved only in one direction (T2 and T3: 10% vs. 7%) while was worse in the other direction (T3 and T2: 16% vs. 19%). Such results suggest that the monosyllabic training seemed to help the learners to distinguish T2 and T3 from each other in monosyllable stimuli; however, the disyllabic training seemed to only help the learners to distinguish T2 from T3, but not T3 from T2 in monosyllable stimuli.

In addition, the least confusing tone pairs across both groups were T1 & T3, and T3 & T4, which had a very low or zero error rates in both pretest and posttest. This shows that the learners across both training groups were able to identify these tones without difficulty before and after training.

4.1.2 **Disyllable stimuli from pretest to posttest**

This section presents results of four tones by the two training groups in connected speech of two syllable words—disyllable stimuli in Mandarin Chinese. A two-way repeated measures ANOVA, with Test as a repeated measure and Training Group as a between-subjects factor, showed a main effect of Test [$F(1, 15)=6.128$, $p<.05$]. Also, a trend in interaction between Test X Training Group [$F(1, 15)=3.273$, $p=.09$] was found, which indicated there was marginal difference in the improvement from pretest to posttest depending on which training group the leaners were in. However, there was no main effect of Training Group [$F(1,15)=1.007$, $p=.331$].

In monosyllabic training group, learners' mean percent of correct identification of disyllable stimuli was 42% in pretest, and 44% in posttest with only a 2% increase. In disyllabic training group, learners' mean percent of correct identification of disyllable stimuli was 29% pretest and 39% posttest, which showed a 9% increase after training. Such results indicated that the disyllabic training assisted the learners more than those in the monosyllabic training group when identifying tones in disyllabic stimuli. In other words, disyllabic training helped learners more than monosyllabic training did in disyllabic word tone identification.

4.1.2.1 ***Individual tones in disyllable stimuli***

Because disyllable stimuli (e.g. má hù) have two tones in each stimulus, for instance, má (σ1) hù (σ2), the results followed are analyzed on tones of each syllable (σ1, σ2).

4.1.2.1.1 Individual tones at the first syllable position (σ1)

Figure 6 and Figure 7 display the tone identification performance by native English-speaking learners in two training groups (monosyllabic training group, disyllabic training group respectively) at the first syllable position in disyllable stimuli.

Results of a three-way repeated measures ANOVA, with Test (pretest and posttest) and σ1_Tone (Tone1, Tone2, Tone3, Tone4) as the within-subjects factors, and Training Group (monosyllabic training group, disyllabic training group) as the between-subjects factor, yielded a main effect of Test [$F(1, 15)=6.531$, $p=.022$], indicating that learners across both training groups did significantly better in posttest (56%) than pretest (49%) on tone identification at the first syllable position in disyllable stimuli.

It also yielded a main effect of σ1_Tone [$F(3,45)=30.913$, $p < .001$]. Pairwise comparisons with Bonferroni correction showed that, across tones at the first syllable position averaged across two tests, the accuracy rates of T1 (62%), T2 (47%,), and T4 (76%,) were significantly higher than that of T3 (24%) (with $p<.001$; $p=.001$; $p<.001$ respectively). T4 identification was also significantly better than T2 ($p=.001$), and T1 was marginally better than T2 ($p=.105$). There was no significant difference between T1 and T4 ($p=.124$). In other words, T3 was the worst tone among all four tones at the first syllable position (σ1).

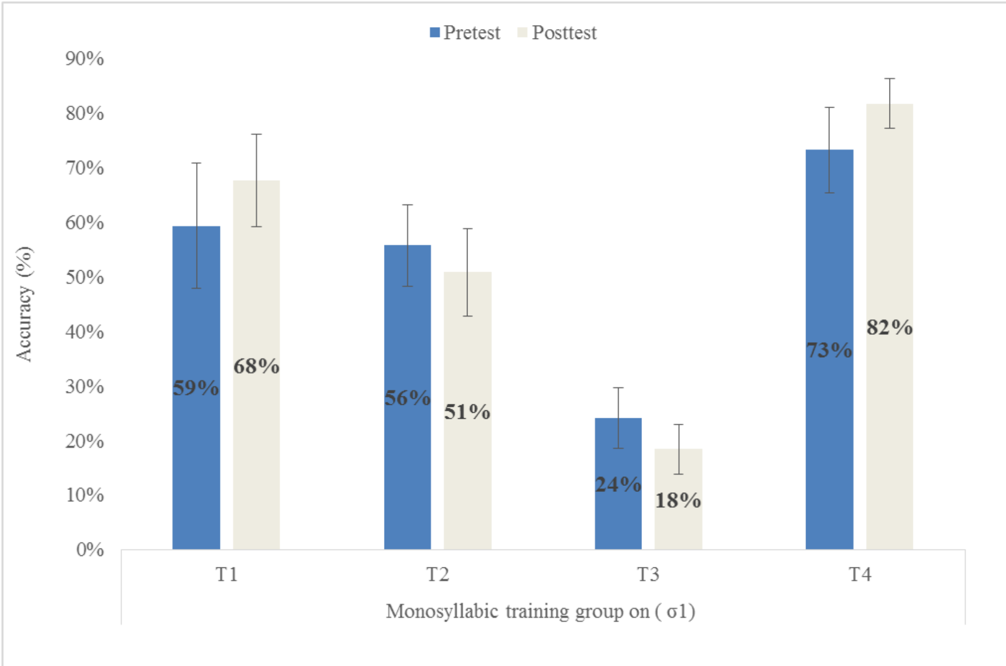


Figure 6: Accuracy rate and standard errors (SE) of four tones at the first syllable position (σ1) in disyllable stimuli by native English-speaking learners of Chinese in monosyllabic training group.

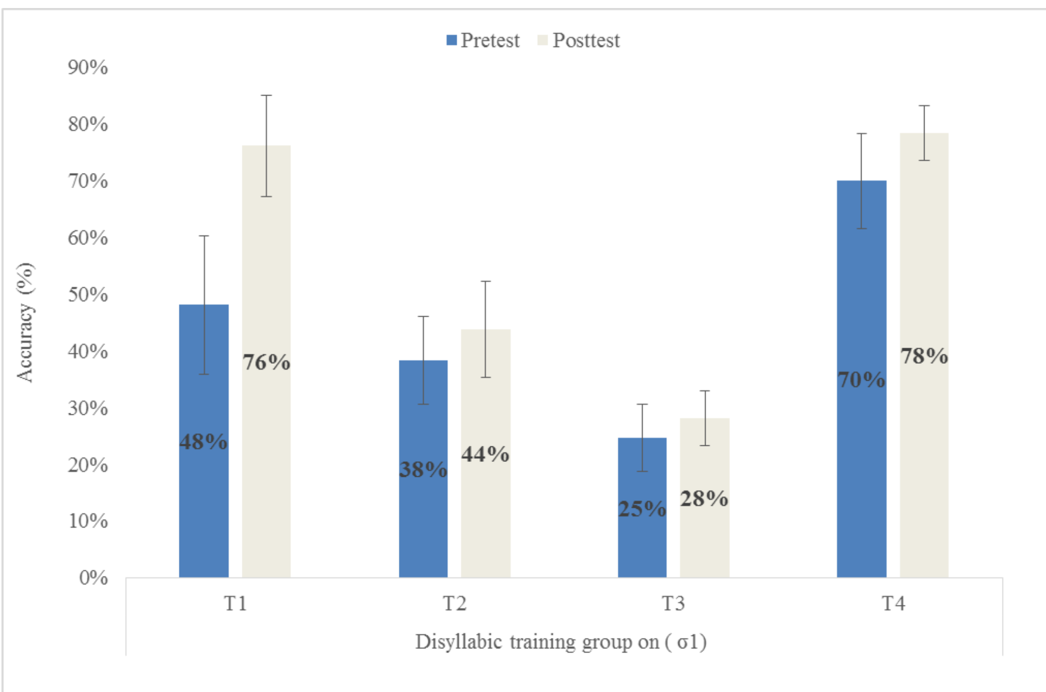


Figure 7: Accuracy rate and standard errors (SE) of four tones at the first syllable position (σ1) in disyllable stimuli by native English-speaking learners of Chinese in disyllabic training group.

There was no main effect of Training Group [$F(1,15)=.180$, $p=.677$]. The monosyllabic training group accuracy rate was 54% and disyllabic training group was 51% across two tests at the first syllable position.

The Test X σ1_Tone interaction was significant [$F(3, 45)=3.309$, $p=.028$], as shown in Figure 6 and Figure 7. The interaction suggested that there were significant improvement on the tones at the first syllable position after training. From pretest to posttest, T1 made a significant increase of 18% from 54% to 72% ($p=.009$); T2 did not make any change with accuracy rates of 47% ($p>.99$); T3 dropped 1% of accuracy rate from 24% to 23% ($p=.72$); T4 increased marginally significantly by 8% from 72% to 80% ($p=.062$).

In other words, the training helped learners improved their tone identification with T1, and marginally with T4, but not much with T2, T3when these tones were at the first syllable position of disyllable words.

Moreover, there was a trend suggested by Test X Training Group [$F(1,15)=3.798$, $p=.070$]. It showed that across four tones at the first syllable position, the learners in monosyllabic training group made an increase of 2% from pretest (53%) to posttest (55%); and the learners in disyllabic training group made an increase of 12% from pretest (45%) to posttest (57%). This indicated that when identifying tones at the first syllable position, disyllabic training helped learners more than monosyllabic training did.

No other two-way or three-way interactions were found.

In conclusion, when at the first syllable position of disyllable stimuli, native English-speaking learners in both training groups identified T1 and T4 consistently and significantly better than T3 from pretest to posttest. Among the four tones, T1 made the most improvement of 18% after training, and T4 was marginally improved 8%. Importantly, averaged across four tones, disyllabic training helped learners improve (12%) the accuracy of the four tones more than monosyllabic training did (2%) from pretest to posttest.

In addition, in Figure 6, after monosyllabic training, it is worth observing that the learners did even worse on both T2 and T3. T2 was 56% in pretest but dropped to 51% in posttest, and T3 was 24% in pretest but dropped to 18% in posttest. Such decrease in the mean accuracy suggested that monosyllabic training was not helping the learners identify T2 and T3 on the first syllable position of disyllable stimuli. On the contrary, in Figure 7, it can be observed that after disyllabic training, both T2 and T3 indeed increased their mean accuracy. T2 increased 6% from pretest 38% to posttest 44%, and T3 increased 3% from pretest 25% to posttest 28%. These results further suggest that disyllabic training was more helpful when identifying all tones (including T2 and T3) on the first syllable position of disyllable stimuli.

4.1.2.1.2 Individual tones at the second syllable position (σ2)

Figure 8 and Figure 9 illustrate the tone identification performance by native English-speaking learners in two training groups at the second syllable position in disyllable stimuli from pretest to posttest.

A three-way repeated measures ANOVA, with Test (pretest and posttest) and σ2_Tone (T1, T2, T3, T4) as within-subjects factors, Training Group (monosyllabic training group and disyllabic training group) as between-subjects factor, revealed a significant main effect of Test [$F(1, 15)=9.880$, $p=.007$]. This main effect of Test showed, averaged across two training groups and four tones at the second syllable position, the learners did significantly better in posttest with 73% accuracy rate than pretest of 67%.

A main effect of σ2_Tone [F (3, 45)=5.354, $p=.003$] suggested that, averaged across groups and two tests, there was significant difference among the four tones. The accuracy rates, from high to low respectively, were: T4 at 80%; T3 at 72%; T1 at 69%; T2 was 58%. Post hoc pairwise comparison with Bonferroni adjustment showed that there was significant difference between T4 and T2 ($p=.007$). However, there was no difference between T4 and T3 ($p=.459$), T4 and T1 ($p=.099$), T3 and T1 ($p>.999$), T1 and T2 ($p=.426$), and T2 and T3 ($p=.381$).

A main effect of Training Group [$F(1, 15)=5.317$, $p=.036$] showed that, averaged across two tests, learners in monosyllabic training group did significantly better on tone identification at the second syllable position with 77% accuracy rate than those in disyllabic training group with 62% accuracy rate. The significant mean difference was 15% between the two groups.

A significant interaction was found between Test X Training Group [$F(1, 15)=7.200$. $p=.017$]. This interaction was due to the significant difference from pretest to posttest by the two training groups. The monosyllabic training group made 1% increase from pretest 77% to posttest 78% while the disyllabic training group made a significantly greater improvement of 13% increase from pretest 56% to posttest 69%. The significant difference of improvement after training from pretest to posttest between the two training groups was 12%.

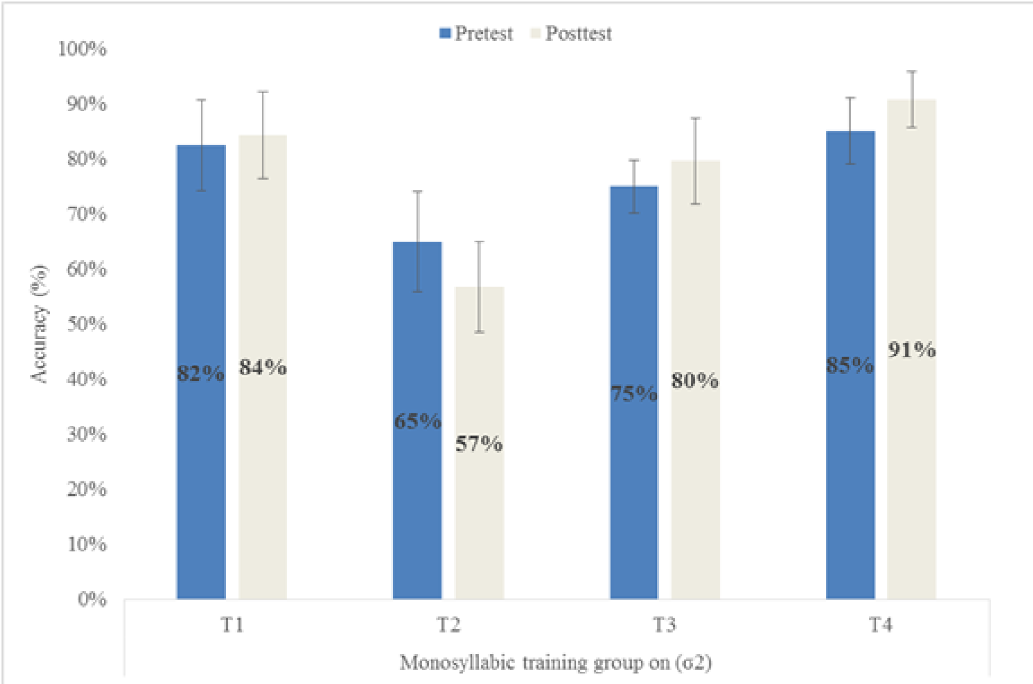


Figure 8: The percent of accuracy and standard error (SE) in individual tones by the learners of monosyllabic training group from pretest to posttest at the second syllable (σ2) of disyllable stimuli.

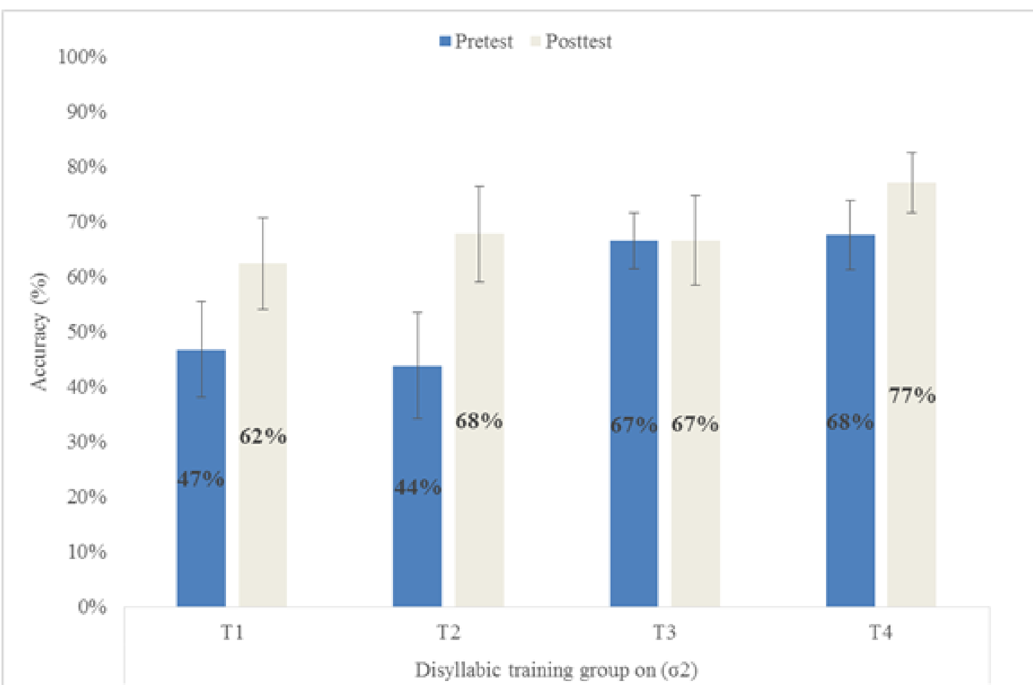


Figure 9: The percent of accuracy and standard error (SE) in individual tones by the learners of disyllabic training group from pretest to posttest at the second syllable (σ2) of disyllable stimuli.

There was no interaction of σ2_Tone X Training Group [$F(3, 45)=1.686$, $p=.183$], neither was there an interaction of Test X σ2_Tone [$F(3, 45)=.329$, $p=.805$].

However, there was a numerical trend indicated by a three-way interaction of Test X σ2_Tone X Training Group [$F(3,45)=2.413$, $p=.079$]. To decompose this three-way interaction, four separate two-way repeated measures ANOVA were conducted by dividing the σ2_Tone into four levels: T1, T2, T3, and T4, which aimed to investigate which tones on second syllable position were improved after training from pretest to posttest.

The two-way repeated measures ANOVA for T1 showed a trend of interaction between Test X Training Group [$F(1,15)=3.152$, $p=.096$], suggesting a numerical increase of T1 at the second syllable position after training, as shown in Figure 10. T1 showed 15% increase from pretest (47%) to posttest (62%) in the disyllabic training group while only 2% increase from pretest (82%) to posttest (84%) in the monosyllabic training group. The difference of improvement was 13% between the two groups.

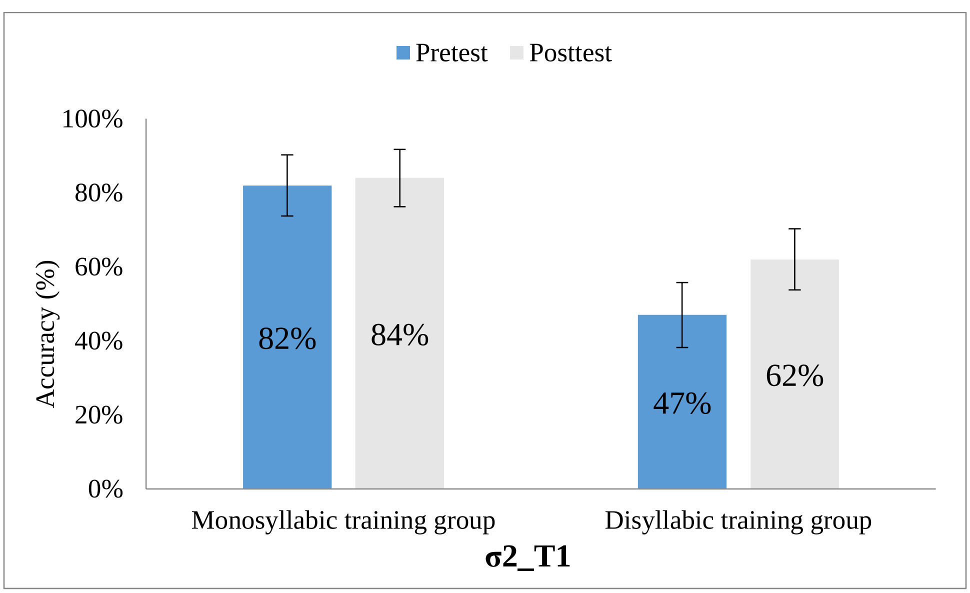


Figure 10: Means of accuracy of T1 at (σ2) the second syllable position by two training groups from pretest to posttest.

For T1 analyses, a main effect of Test [$F(1,15)=5.124$, $p=.039$] was also found, which indicated that averaged across two groups, T1 was better in posttest (74%) than in pretest (65%) after training.

A main effect of Training Group [$F(1,15)=6.836$, $p=.020$] was found as well. This indicated when identifying T1 on second syllable in disyllable stimuli, monosyllabic training group (83%) was better than disyllabic training group (55%) across two tests.

The two-way repeated measures ANOVA by T2 yielded a significant interaction of Test X Training Group [$F(1,15)=6.650$, $p=.021$] as shown in Figure 11, and no main effects found. This suggested the learners in disyllabic training group made a greater improvement of 24% from pretest to posttest than the learners in monosyllabic training group, who actually dropped 8% of the mean accuracy from pretest 65% to posttest 57%. The difference of improvement was a significant 16% between the two training groups.

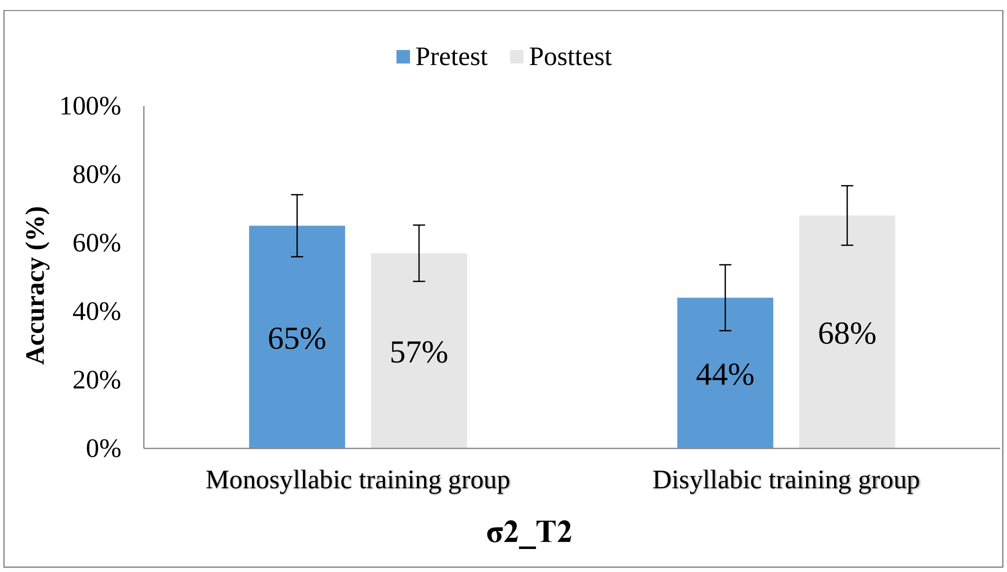


Figure 11: Means of accuracy of T2 at (σ2) the second syllable position by two training groups from pretest to posttest.

In other words, disyllabic training showed greater improvement on T2 identification than monosyllabic training at the second syllable of disyllable stimuli. This significant improvement on T2 trigged the trend in the three-way interaction of Test X σ2_Tone X Training Group.

No other significant interactions were found in the T3 and T4 repeated measures analyses.

The results in analyzing Test and Training Group as repeated measures by four tones provided the evidence for the marginal three-way interaction: Test X σ2_Tone X Training Group [$F(3,45)=2.413$, $p=.079$], that this improvement was found significantly for T2 (24%) and marginally for T1 (15%). Therefore, it can be concluded that disyllabic training seemed to elicit a significant improvement in tone perception, at least for T1 and T2 on second syllable of the disyllable stimuli.

4.1.2.2 ***Tone confusions in disyllable stimuli***

Confusion between tone pairs on each syllable position were examined in order to understand the mistakes that learners made in the tone identification task. The analyses include two training groups' tone confusion of each syllable within the disyllable stimuli. The error rates for each tone pair were investigated in two directions respectively. For example, for tone pair T1 & T2, the percentage of errors for tone pair T1 and T2 represented the error rate in the direction when T1 was misidentified as T2; the percentage of errors for tone pair T2 and T1 represented the error rate in the other direction when T2 was misidentified as T1. In Chinese, there are sixteen pairs of disyllable tones (4 tones X 4 tones =16 pairs), however, due to the "Third Tone Sandhi" rule that T3 is changed to T2 before another T3, all T3 + T3 disyllable stimuli were coded as T2 + T3.

4.1.2.2.1 Tone confusions by monosyllabic training group

Monosyllabic training group learners' tone identification of the four tones in two syllables (σ1 and σ2) of disyllable stimuli were displayed in Table 3 and Table 4, respectively. There were total of 108 stimuli (12 stimuli X 9 students=108) at each of the first and the second syllable position of the disyllable stimuli. Error numbers were converted to percentage in both Table 3 and Table 4.

Table 4 showed the learners' tone confusion in percentages at the first syllable (σ1) of disyllable stimuli by monosyllabic training group.

Table 4: Confusion matrices of the four individual tones at the first syllable position (σ1) by monosyllabic training group from pretest to posttest in percentage.

STIMULUS \ PERCEIVED	PRE_DISYLLABLE_σ1 MONO_TR_GROUP				POST_DISYLLABLE_σ1 MONO_TR_GROUP			
	T1	T2	T3	T4	T1	T2	T3	T4
T1	59	19	6	16	68	14	2	17
T2	21	56	7	17	24	48	4	24
T3	2	28	19	51	4	23	15	58
T4	12	13	2	73	8	9	1	81

From Table 4, in pretest, the most confusable tone pair was T3 & T4. There were 51% of T3 perceived as T4, and this confusion was even worse in posttest with a high of 58% of error rate. This showed that learners in monosyllabic training group had great difficulty of distinguishing T3 and T4 at the first syllable position in disyllable stimuli. Fortunately, this great difficulty occurred only in one direction. For tone pair T4 and T3, the learners misidentified only 2% of T4 as T3 in pretest, and improved with even less errors rate at 1% after training.

The reason that the learners in monosyllabic training group misidentified over half of T3 as T4 was probably due to the “Half-Third Sandhi” rule, in which T3’s pitch value 213 was reduced into 21, as a low falling tone before any tone other than another T3 (Zhang, 2007). Therefore, in current disyllable stimuli, all T3s (213 as in isolation) at the first syllable position were produced as a low falling T3 (21), which resembled the contour movement of T4 (51) that is also a falling tone.

The next confusable tone pair was T3 & T2. The learners perceived 28% of T3 as T2 in pretest, and improved in posttest with 23% of T3 as T2. In the other direction, the error rate was relatively low that the learners misidentified 7% of T2 as T3 in pretest, and improved after training with 4% error rate.

Another tone pair, T2 & T1, showed consistent tone confusion in both directions. In one direction, a 21% of T2 were misperceived as T1 in pretest, and this tone confusion did not improve after training with 24% error rate. In the other direction, 19% of T1 were misperceived as T2 in pretest, and this error rate was 14% in posttest.

Tone pair T2 & T4 was also confusable to the learners’ in the monosyllabic training group. 17% of T2 were misidentified as T4 in pretest, and this confusion was worse in posttest with a 24% error rate. There was 13% of T4 misperceived as T2 in pretest, and 9% in posttest. This showed some improvement after training.

The learners in monosyllabic training group also showed confusion to tone pair T1 & T4. 16% of T1 was misidentified as T4 in pretest, and 17% in posttest without improvement.

However, in the other direction, 12% of T4 was misidentified as T1 in pretest, and 8% in posttest with 4% improvement.

The least confusable tone pair was T1 & T3. In pretest, there was 6% of T1 perceived incorrectly as T3, and was only 2% after training. 2% of T3 was misperceived as T1 in pretest, and 4% in posttest.

Table 5 showed the learners' tone confusions in percentages at the second syllable (σ2) of disyllable stimuli by monosyllabic training group.

Table 5: Confusion matrices of the four individual tones at the second syllable position (σ2) by monosyllabic training group from pretest to posttest in percentage.

STIMULUS \ PERCEIVED	PRE_DISYLLABLE_σ2 MONO_TR_GROUP				POST_DISYLLABLE_σ2 MONO_TR_GROUP			
	T1	**T2**	**T3**	**T4**	**T1**	**T2**	**T3**	**T4**
T1	**82**	**6**	**5**	**7**	**84**	**10**	**0**	**6**
T2	**3**	**65**	**29**	**4**	**2**	**56**	**35**	**6**
T3	**0**	**23**	**75**	**2**	**2**	**17**	**80**	**2**
T4	**1**	**11**	**3**	**85**	**1**	**7**	**1**	**91**

The most confusable tone pair in the table was T2 & T3. In pretest, a 29% of T2 were perceived as T3, and this confusion was even greater in posttest with 35% error rate. In the other direction, when 23% of T3 were perceived as T2 in pretest, it was improved to 17% after training in posttest.

Tone pairs, such as T2 & T4, and T1 & T2, showed less confusion when compared to T2 & T3. 4% of T2 was perceived as T4 in pretest and 6% in posttest. In the other direction, however, 11% of T4 were perceived as T2 in pretest and only 7% in posttest.

Similarly, 6% of T1 was perceived as T2 in pretest and 10% in posttest, which also showed more confusion after training. But, there was 3% of T2 perceived as T1 in pretest, and 2% in posttest.

Some tone pairs displayed very low error rate at the second syllable of disyllable stimuli by the learners in monosyllabic training group. For instance, tone pair T1 & T4, 7% of T1 were perceived as T4 in pretest, and 6% in posttest. In the other direction, there was 1% of T4 perceived as T1 in both pretest and posttest. 5% of T1 were identified as T3 in pretest, and no misidentification in posttest. For T3 & T4, there was no change from pretest to posttest with 2% of T3 misperceived as T4. In other direction, 3% of T4 were perceived as T3 in pretest and 1% after training.

Overall, when comparing Table 3 and Table 4, across the board from pretest to posttest, the learners in monosyllabic training group made improvement at both syllable positions on T1 and T4, and at the second syllable position on T3 after training. For instance, T1 at σ1, the accuracy rate was increased from 59% to 68%, and at σ2, from 82% to 84%; T4 at σ1, from 73% to 81%, and at σ2, from 85% to 91%. Also, when T3 was in σ2, it showed improvement from 75% to 80%.

However, T2 showed decreased accuracy rate from 56% to 48% at σ1, and also decreased from 65% to 56% when at σ2. Similarly, when T3 was at σ1, it decreased from 19% to 15% after training. It seemed that T2 was the most difficult tone to identify in disyllable stimuli at both syllable positions by the learners in monosyllabic training group, and all other three tones made some improvements after training.

Analyzing the tone confusion across two syllable positions, it seems that the learners in monosyllabic group had most difficulty in distinguishing T3 from T4 (error rates of 51% and 58% in pre- and post-test respectively) at the first syllable position, which may due to the "Half-Third Sandhi" rule of T3. They also had most difficulty in distinguishing between T2 & T3 at the second syllable position (error rates of 29% and 35%; and 23% and 17% respectively in both directions from pre- to post-test), and at the first syllable position (error rates of 28% to 23% in tone pair of T2 and T3 from pre- to post-test) in disyllable stimuli.

The least confusion tone pairs were T1 & T3 across the two syllable positions with very low or zero error rates, as well as T3 & T4 at the second syllable position by the learners in the monosyllabic training group.

4.1.2.2.2 Tone confusions by disyllabic training group

Disyllabic training group learners' tone identification of the four tones in two syllables (σ1 and σ2) of disyllable stimuli were displayed in Table 5 and Table 6, respectively. There were total of 96 stimuli (12 stimuli X 8 students=96) at each of the first and the second syllable position of the disyllable stimuli. Error numbers were converted to percentage in both tables below.

Table 6 showed the learners' tone confusion in percentages at the first syllable (σ1) of disyllable stimuli by the disyllabic training group.

Table 6: Confusion matrices of the four individual tones at the first syllable position (σ1) by disyllabic training group from pretest to posttest in percentage.

STIMULUS / PERCEIVED	PRE_DISYLLABLE_σ1				POST_DISYLLABLE_σ1			
	DI_TR_GROUP				DI_TR_GROUP			
	T1	T2	T3	T4	T1	T2	T3	T4
T1	48	15	9	28	76	9	0	15
T2	23	38	12	28	18	44	16	23
T3	0	18	22	60	3	22	22	53
T4	9	9	11	70	10	6	5	78

The most confusable tone pair in Table 6 was T3 & T4. In pretest, a large 60% of T3 were perceived as T4, and 53% in posttest. However, in the other direction, there was only 11% of T4 perceived as T3 in pretest, and 5% in posttest. This extremely high error rate of misperceiving T3 as T4 could be explained by the "Half-third Sandhi" rule as well, in which the learners in disyllabic training group misperceived the low falling tone T3 (pitch value: 21) as the high falling tone T4 (pitch value: 51) at the first syllable position.

However, the difference between two training groups' performance is that the monosyllabic training group actually was worse after training (51% to 58%) while the disyllabic training group made some improvement (60% to 53%) despite the difficulty of distinguishing T3 from T4. This implies that disyllabic training seemed to assist the learners more when identifying the low falling T3 at the first syllable position in disyllable stimuli than monosyllabic training did.

Tone pair T2 & T3 also demonstrated a great deal of confusion in both directions at the first syllable position by disyllabic training group learners. In pretest, 12% of T2 were misidentified as T3, and in posttest, it was 16%. In the other direction, 18% of T3 were misidentified as T2 in pretest, and 22% in posttest. However, it is noteworthy that T2's mean accuracy improved from 38% to 44%, and T3's mean accuracy stayed the same (22%) after disyllabic training, in spite of the tone confusion between T2 & T3. This type of improvement on T2 identification did not happen in monosyllabic training.

The next groups of confusable tone pairs were T2 &T4, T1 & T4, and T2 & T1. These tone pairs all showed relatively high error rates in one direction and low error rates in the other direction. For instance, 28% of T2 were misidentified as T4 in pretest and 23% in posttest; however, only 9% of T4 were misidentified as T2 in pretest, and 6% in posttest. For T1 & T4, 28% of T1 were perceived as T4 in pretest, and 15% in posttest; in the other direction, 9% of T4 were perceived as T1 in pretest, and 10% in posttest. For T2&T1, in one direction, 23% of T2 were perceived as T1 in pretest, and 18% in posttest; in the other direction, 15% of T1 were misidentified as T2 in pretest, and 9% in posttest.

T1 & T3, again, was the least confusable pair by the learners of disyllabic training group at the first syllable position. 9% of T1 were misidentified as T3 in pretest, and no misidentification in posttest. In the other direction, no misidentification in pretest, and 3% of T3 were misidentified as T1 in posttest. This easy to distinguish tone pair echoes the finding by the learners in monosyllabic training group at the first syllable position.

Table 7 showed the tone confusion in percentages at the second syllable (σ2) of disyllable stimuli by the learners of disyllabic training group.

Table 7: Confusion matrices of the four individual tones at the second syllable position (σ2) by disyllabic training group from pretest to posttest in percentage.

STIMULUS \ PERCEIVED	PRE_DISYLLABLE_σ2				POST_DISYLLABLE_σ2			
	DI_TR_GROUP				DI_TR_GROUP			
	T1	**T2**	**T3**	**T4**	**T1**	**T2**	**T3**	**T4**
T1	**47**	**20**	**10**	**23**	**63**	**15**	**6**	**17**
T2	**11**	**44**	**34**	**10**	**4**	**68**	**25**	**3**
T3	**4**	**21**	**67**	**8**	**1**	**31**	**67**	**1**
T4	**11**	**11**	**9**	**68**	**10**	**10**	**2**	**77**

The most confusing tone pair in the table was T2 & T3. There was 34% of T2 perceived as T3 in pretest, and in posttest, this error rate decreased to 25% after training. However, in the other direction, 21% of T3 were perceived as T2 in pretest, and the error rate increased to 31% after training.

The next confusing tone pairs were T1 & T4, and T1 & T2. These two tone pairs showed high error rates in one direction and a relatively low error rate in the other direction. For example, 23% of T1 were misidentified as T4 in pretest and 17% in posttest; however, 11% of T4 were misidentified as T1 in pretest, and 10% in posttest. For T1 & T2, 20% of T1 were perceived as T2 in pretest, and 15% in posttest; and in the other direction, 11% of T2 were perceived as T1 in pretest, and 4% in posttest.

Tone pair T2 & T4 demonstrated less confusion compared to previous tone pairs. In pretest, 10% of T2 were misidentified as T4, and in posttest, it decreased to 3%. In the other direction, 11% of T4 were misidentified as T2 in pretest, and 10% in posttest.

T1 & T3 and T3 & T4 were the least confusable pairs in Table 6. 10% of T1 were misidentified as T3 in pretest, and 6% in posttest. In the other direction, 4% of T3 were misperceived as T1 in pretest, and 1% in posttest. For T3 & T4, there was 8% of T3 perceived as T4 in pretest, and this decreased to 1% in posttest; 9% of T4 was misidentified as T3, and it decreased to 2% in posttest.

In conclusion, for the learners in disyllabic training group, the least confusable tone pair at both syllables was T1 & T3. This T1 & T3 easy differentiation across syllable positions and training groups was probably due to the clear difference embedded in the phonetic characteristics, for instance, T1 has a high onset while T3 has a low onset; T1 is a level tone without pitch contour, but T3 (21) is a low falling tone at first syllable position and a contour tone at the second syllable position (213).

At the first syllable position, the learners across both groups misidentified T3 as T4 the most (51% to 58%, 60% to 53% respectively), which is due to the “Half-third Sandhi” rule. At the second syllable position, the learners misidentified T2 & T3 in both directions the most as described above.

4.1.3 **The effect of three linguistic factors on disyllable stimuli**

Tone identification accuracy data was analyzed to examine the three linguistic factors, namely syllable position (initial position vs. final position), tonal context (compatible tonal context vs. conflicting tonal context), and tonal sequence (same tonal sequence vs. different tonal sequence).

4.1.3.1 ***Effects of syllable position***

Figure 12 displays the mean of accuracy at the two syllable positions by native English-speaking learners in two training groups from pretest to posttest. Results of a three-way repeated measures ANOVA, with Syllable Position (initial position, final position) and Test (pretest, posttest) as with-subjects factors, and Training Group (Monosyllabic Training Group, Disyllabic Training Group) as between-subjects factor, yielded main effects of Test [F (1,15)=18.797, p=.001] and Syllable Position [F(1,15)=85.530, p<.001]. Two significant interactions were obtained, Test X Syllable Position [F(1,15)=10.833, p=.005] and Syllable Position X Training Group [F(1,15)=9.823, p=.007].

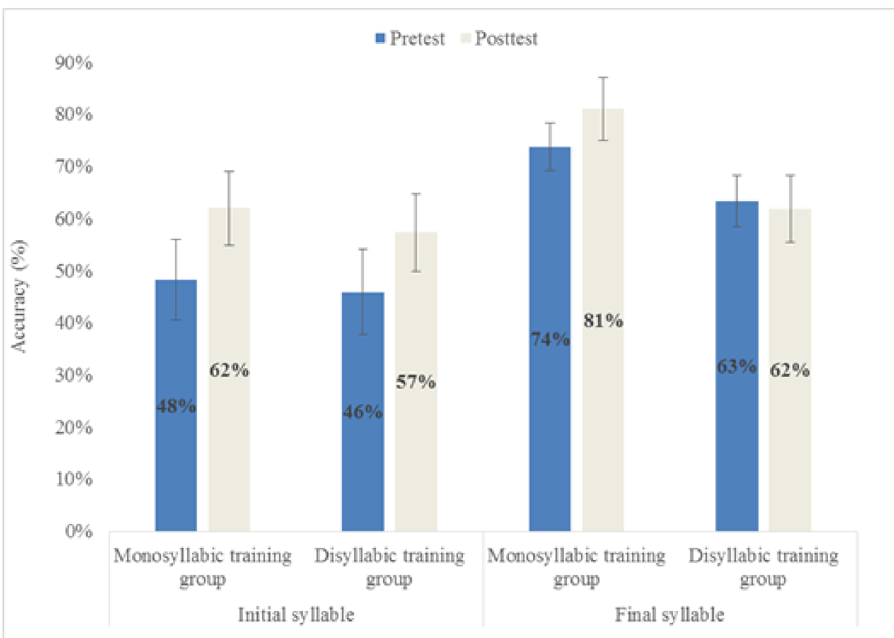


Figure 12: Accuracy rate and standard errors (SE) at two syllable positions—initial syllable and final syllable by native English-speaking learners of two training groups in pretest and posttest.

The main effect of Test [F (1,15)=18.797, p=.001] demonstrated that averaged across two training groups and two syllable positions, the learners did significantly better in posttest (66%) than pretest (58%), which suggested that the high variability training helped learners when identifying tones in disyllable stimuli.

The main effect of Syllable Position [$F(1,15)=85.530$, $p<.001$] showed that native English-speaking learners across two training groups and both tests did significantly better on tones in the final position (70%) than the tones in initial position (53%).

Test X Syllable Position [$F(1,15)=10.833$, $p=.005$] showed the 13% improvement of accuracy at initial position from pretest (47%) to posttest (60%), which was significantly higher than the 2% improvement at final position from pretest (69%) to posttest (71%). This difference of improvement suggested that after training, in disyllable stimuli, the learners' tone perception improved more at the initial position than at the final position. In other words, learners across groups had difficulty in improving their tone perception at the final position despite the fact that the tones on the final syllable seemed to have a high accuracy rate before training.

Syllable Position X Training Group [$F(1,15)=9.823$, $p=.007$] demonstrated, collapsed the two tests, the difference of the accuracy rate (3%) at initial position between monosyllabic training group (55%) and disyllabic training group (52%) was significantly different from the difference of accuracy rate (14%) at final position between the two groups (77% vs. 63%). This interaction suggested that the learners in the monosyllabic training group did better on tones at the final position than the tones at the initial position.

Overall, these results suggested, in disyllable stimuli, learners were significantly more accurate when identifying tones in final syllable position than in initial position. Also, training showed significant improvement on tones at the initial position from pretest to posttest.

4.1.3.2 ***Effects of tonal context***

Figure 13 shows how native English-speaking learners in two training groups performed in the tone identification task in two tonal contexts, compatible and conflicting, from pretest to posttest.

A three-way repeated measures ANOVA, with Test (pretest, posttest) and Tonal Context (compatible, conflicting) as within-subjects factors, and Training Group as between-subjects factor, obtained main effects of Test [$F(1,15)=5.552$, $p=.032$] and Tonal Context [$F(1,15)=14.183$, $p=.002$], and a marginal interaction between Test X Training Group $F(1,15)=3.091$, $p=.099$].

No other main effects or interactions were found.

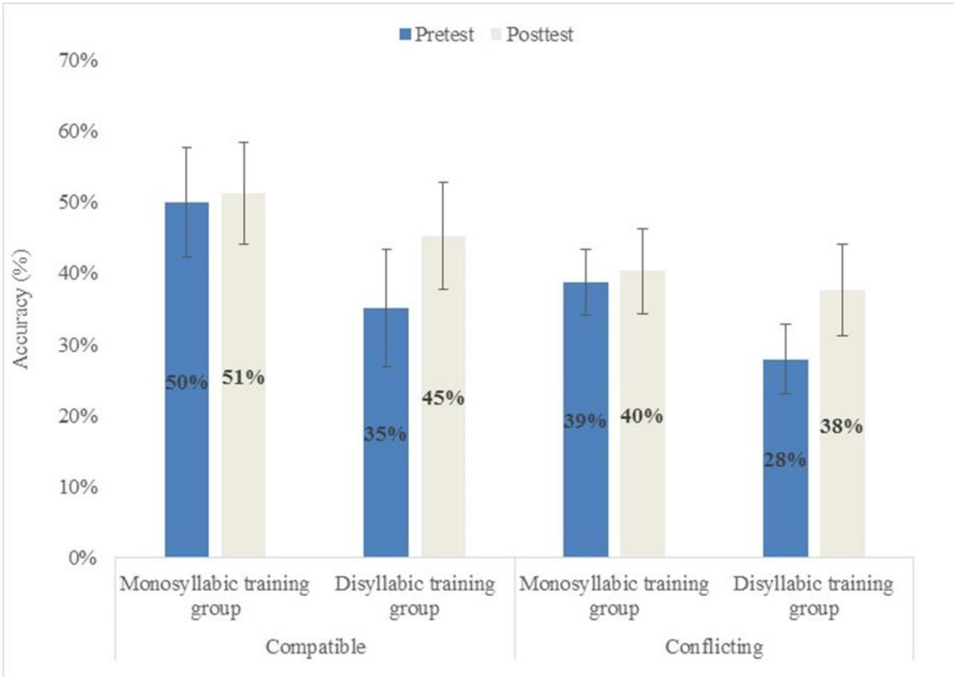


Figure 13: Percentage of accuracy and standard errors (SE) at compatible and conflicting tonal context by native English-speaking learners of two training groups in pretest and posttest.

The main effect of Test showed that averaged across two training groups and two tonal contexts, the learners did significantly better after training in posttest (44%) than in pretest (38%).

The main effect of Tonal Context indicated, averaged across tests and training groups, the learners did significantly better in compatible tonal context (45%) than in conflicting tonal context (36%) with a 9% higher accuracy rate. That is to say, the learners can identify tones in compatible tonal contexts better than in conflicting tonal contexts.

A marginal Test X Training interaction demonstrated that across two tonal contexts, the learners in disyllabic training group made a numerically larger improvement (10%) from pretest (32%) to posttest (42%) than the improvement (1%) made by monosyllabic training group from pretest (45%) to posttest (46%). Such results indicated disyllabic training helped learners more than monosyllabic training did when identifying tones across two tonal contexts from pretest to posttest.

4.1.3.3 ***Effects of tonal sequence***

Figure 14 displays two groups of native English-speaking learners' tone identification in two tonal sequences, namely same tonal sequence and different tonal sequence, in disyllable stimuli from pretest to posttest.

A three-way repeated measures ANOVA, with Test (pretest and posttest) and Tonal Sequence (same and different) as within-subjects factors, Training Group (monosyllabic training group and disyllabic training group) as between-subjects factor, obtained a main effect of Tonal Sequence [$F(1,15)=19.630$, $p<.001$], and an interaction of Tonal Sequence X Training Group [$F(1,15)=6.252$, $p=.024$]. No other main effect or interactions were found.

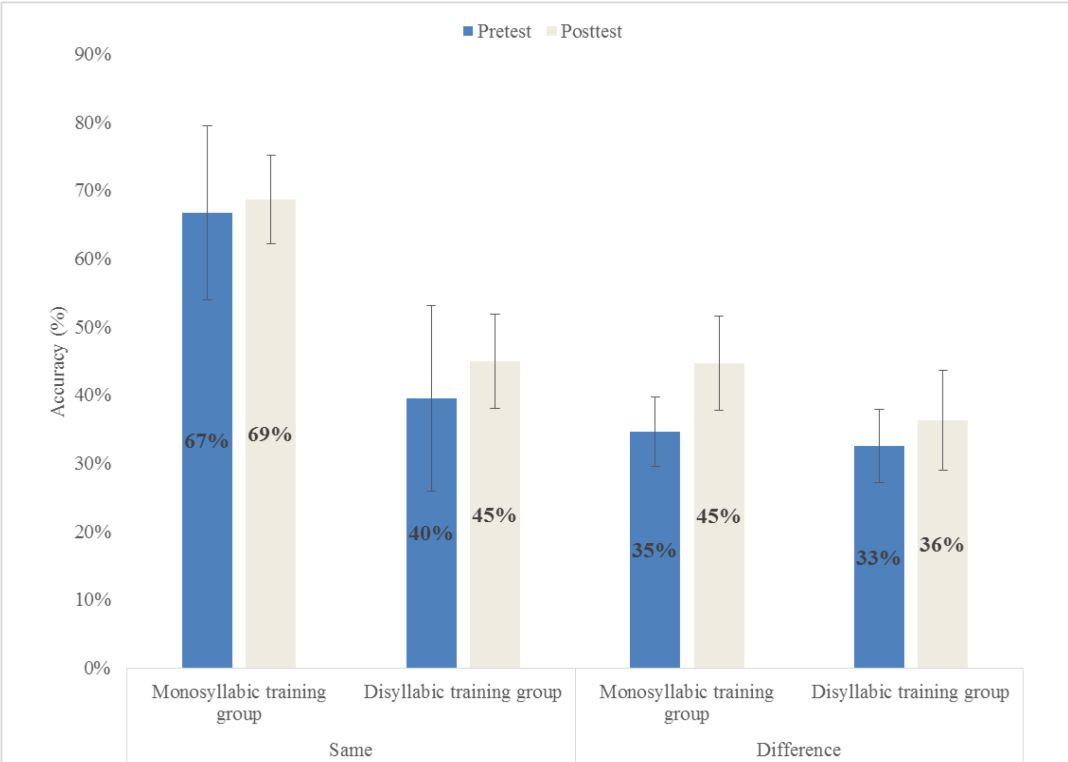


Figure 14: Mean percent of accuracy and standard errors (SE) by native English-speaking learners of two training groups in same and different tonal sequences in pretest and posttest.

The main effect of Tonal Sequence indicated that, averaged across training groups and tests, the accuracy rate of same tonal sequence (55%) was significantly better than the accuracy rate of different tonal sequence (37%).

The learners in monosyllabic training group did substantially worse in the different tonal sequence (39%) than in the same tonal sequence (68%). The learners in the disyllabic training group had a mean of 34% in the different tonal sequence and 44% in the same tonal sequence. So, this significant difference of 19% between the same and different sequences by two training groups showed that learners in monosyllabic training groups were generally worse at identifying tones in the different tonal sequence than tones in the same tonal sequence.

4.2 Generalization test

A generalization test including both new monosyllable stimuli and disyllable stimuli was given to native English-speaking learners in both training groups. The purpose of the generalization test was to examine if the training effect can be generalized both to different stimuli than were used in the training process and to different speakers that learners had not heard before.

4.2.1 **Overall improvement in pretest, posttest, and generalization test**

Tone perception accuracy rates in both monosyllable stimuli and disyllable stimuli by native English-speaking learners of two training groups are shown in Figure 15. A Greenhouse-Geisser correction was applied to report *F* values and *p* values when needed.

Due to the sickness, one participant withdrew from the generalization test in the monosyllabic training group; thus, the number of participants analyzed in following section was sixteen (8 in monosyllabic training group, and 8 in disyllabic training group).

A three-way repeated measures ANOVA, with Test (pretest, posttest, generalization test) and Stimuli (monosyllable stimuli, disyllable stimuli) as within-subjects factors, and Training Group (monosyllabic training group, disyllabic training group) as between-subjects factor, was run to investigate if the training effect that was found in pre- and post-tests could be transferred to new stimuli by a new speaker.

A main effect of Test [$F(1.259, 17.633)=9.086$, $p=.005$] was found, suggesting that the training effect was extended to new stimuli by a new speaker. The learners did significantly better in generalization test with the accuracy rate of 72% (new stimuli by new speaker) than pretest (60%). The generalization test accuracy (72%) was comparable to that in the posttest (66%) (old stimuli by old speaker). Pairwise comparisons with Bonferroni multiple adjustments showed that posttest was better than pretest (*p*=.005), and generalization test was better than pretest (*p*=.012). There was no significant difference between generalization test and posttest accuracy (*p*=.167). Listeners were able to generalize to new stimuli and a new speaker.

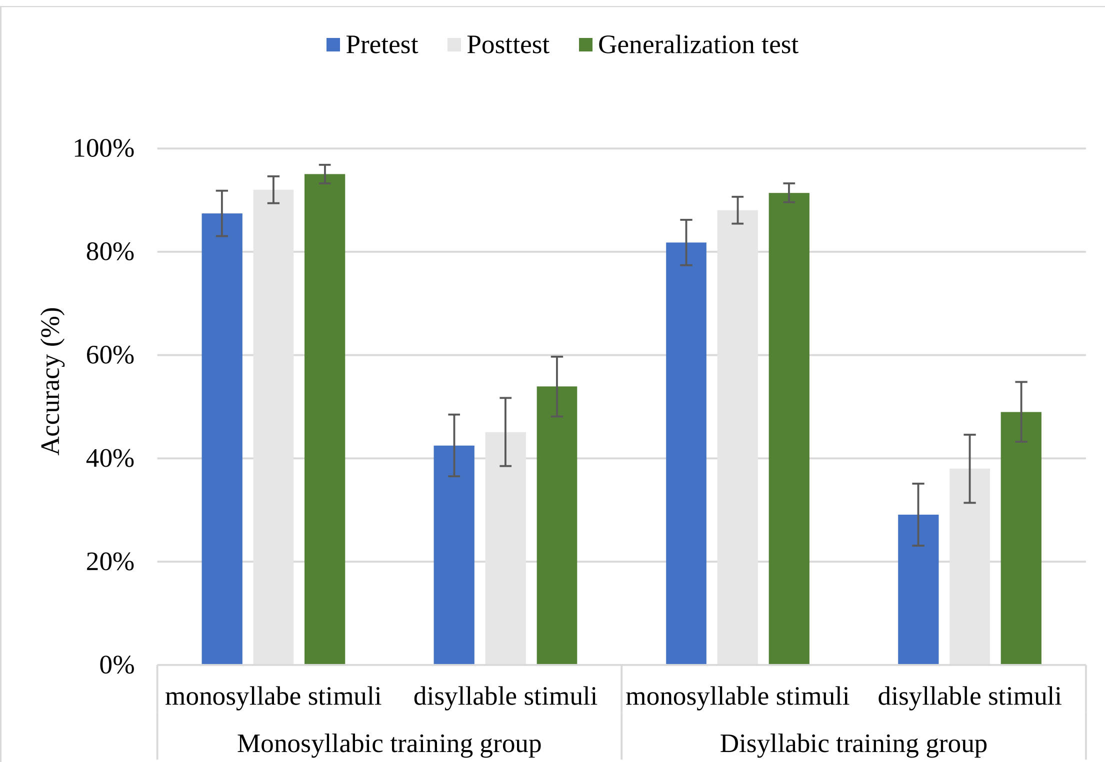


Figure 15: Percentage of accuracy and standard errors (SE) by native English-speaking learners in two training groups for monosyllable stimuli and disyllable stimuli in pretest, posttest, and generalization test.

A main effect of Stimuli [$F(1, 14)=379.094$, $p<.001$] was obtained. This indicated that the learners across both groups and tests did significantly better on monosyllable stimuli (89%) than disyllable stimuli (43%).

No other significant effects or interactions were found.

Overall, the learners from both training groups did significantly better in generalization test than in pretest, which suggests a training effect extension to new stimuli and new speaker. Across the three tests, the learners did generally better in monosyllable stimuli, which indicated that the learners' tone perception of the two different types of stimuli was different. This warrants a further investigation on tones in monosyllable stimuli and disyllable stimuli separately.

4.2.2 Monosyllable stimuli in generalization test

The two groups' performance on monosyllable stimuli in generalization test were analyzed in repeated measures ANOVA as displayed in Figure 16.

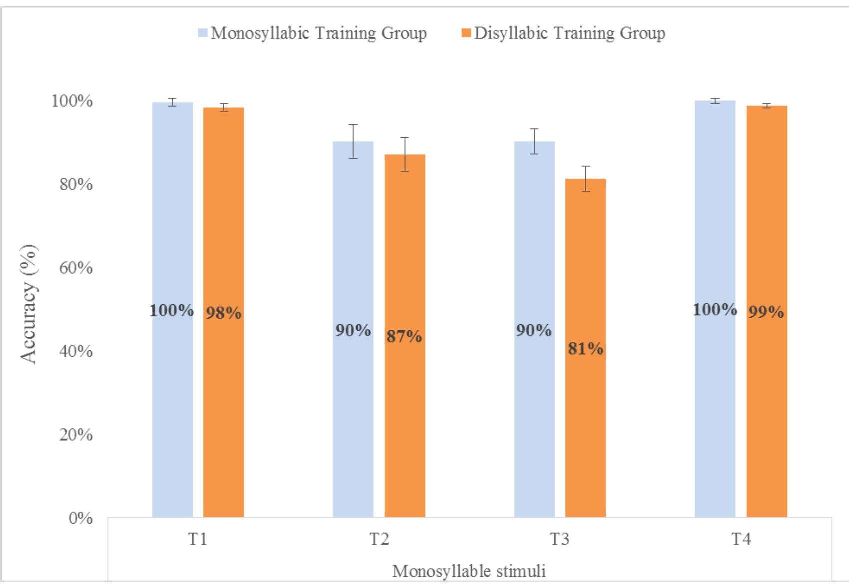


Figure 16: Percentage of accuracy and standard errors (SE) by native English-speaking learners in two training groups for monosyllable stimuli in generalization test.

A two-way repeated measures, with Tone (T1, T2, T3, T4) as within-subjects factor, and Training Group (monosyllabic training group, disyllabic training group) as between-subjects factor, was conducted to examine the two groups of learners' tonal perception performance in monosyllable stimuli, yielded a main effect of Tone [$F(1.565,21.905)=21.323$, $p<.001$], which suggested that there was significant difference among four tone's accuracy rates. Pairwise comparison with Bonferroni adjustment showed that T1 (99%) and T4 (99%) were significantly better than T2 (89%) and T3 (86%), with *p* values respectively at $p=.013$ (T1 vs. T2, T4 vs. T2), and $p<.001$ (T1 vs. T3, T4 vs. T3). There was no difference between T1 and T4 ($p>.99$), neither between T2 and T3 ($p>.99$).

No other main effects or interactions were found.

4.2.3 **Disyllable stimuli in generalization test**

In Figure 17, the learners tone performance on each tone of the disyllable stimuli in generalization test was analyzed in a three-way repeated ANOVA with Syllable (σ1, σ2), Tone (T1, T2, T3, T4) as within-subjects factors, and Training Group as a between-subjects factor.

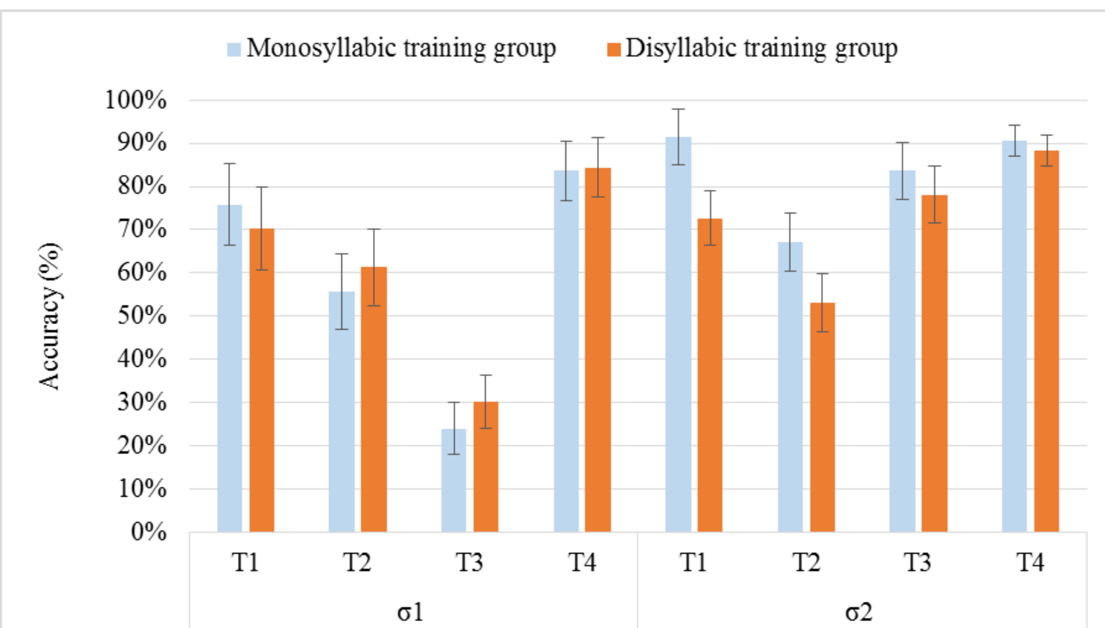


Figure 17: Accuracy rate and standard errors (SE) of four tones in two syllables in disyllable stimuli by native English-speaking learners in two training groups in generalization test.

The results showed that there was main effect of Syllable [$F(1, 14)=42.847$, $p<.001$], which indicated that the learners across groups did significantly better on tones at the second syllable σ2 (78%) than at the first syllable σ1 (62%).

There was also main effect of Tone [$F(3,42)=20.890$, $p<.001$]. Pairwise comparisons with Bonferroni multiple adjustments showed T1 (78%) and T4 (87%) were significantly better than T2 (59%) and T3 (54%). However, the perception accuracy of T2 and T3 were comparable, and accuracy of T1 and T4 were also comparable.

There were also significant interactions of Syllable X Tone [$F(3,42)=25.692$, $p<.001$], and Syllable X Training Group [$F(1,14)=5.005$, $p=.042$]. Post hoc analyses showed that the learners across both groups did significantly better on T3 at the second syllable position (80%) than at the first syllable position (27%) ($p<.001$). Moreover, the leaners in monosyllabic training group also did marginally better on T2 at the second syllable position (67%) than in the first syllable position (56%) ($p=.091$).

The learners' individual tone identification at each syllable position was analyzed in the following two sections.

4.2.4 **Individual Tones at the first syllable position (σ1)**

A two-way repeated measures ANOVA, with σ1 _Tone (T1, T2, T3, T4) as a within-subjects factor, and Training Group as a between-subjects factor, yielded a main effect of S1 (σ1)_Tone [$F(3,42)=25.535$, $p<.001$]. Pairwise comparison with Bonferroni multiple adjustments showed that, in generalization test, at the first syllable position, the learners across both training groups did significantly better in T1 (73%), T2(58%), and T4(84%) than in T3(27%), with $p<.001$ (T1 vs. T3), $p=.002$ (T2 vs. T3), and $p<.001$ (T4 vs. T3) respectively. Also, T4 was better than T2 ($p=.026$).

There were no other main effects or interactions found.

4.2.5 **Individual Tones at the second syllable position (σ2)**

A two-way repeated measures ANOVA, with σ2_Tone (T1, T2, T3, T4) as a within-subjects factor, and Training Group as a between-subjects factor, showed a main effect of σ2_Tone [$F(3,24)=15.464$, $p<.001$]. Pairwise comparison with Bonferroni multiple adjustments showed that, in generalization test, at the second syllable position, the learners across both training groups did significantly better in T1(82%), T3(81%), and T4(89%) than in T2 (60%), with $p=.002$ (T1 vs. T2), $p=.007$ (T3 vs. T2), and $p<.001$ (T4 vs. T2) respectively. The accuracy rates in T1, T3, and T4 were comparable.

No other main effects or interactions were found at the second syllable position in generalization test.

4.3 Three linguistic factors in generalization test

Three linguistic factors, syllable position, tonal context, and tonal sequence, were investigated in disyllable stimuli in generalization test. The purpose is to examine if the learners' tone perception in new stimuli by a new speaker shares the similar pattern as it was in pre- and post-tests.

4.3.1 **Effect of Syllable position**

Figure 18 presents the learners tone performance at initial syllable and final syllable in disyllable stimuli in generalization test.

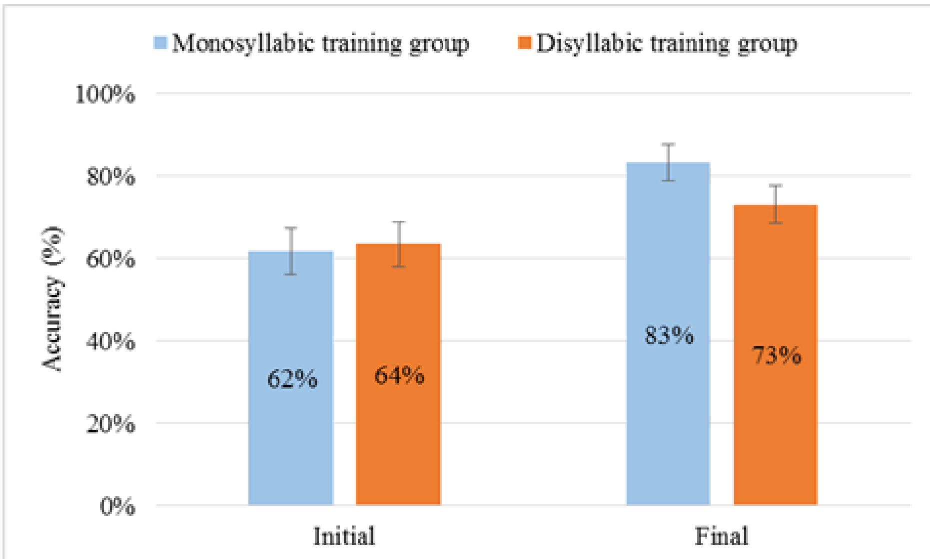


Figure 18: Percentage of accuracy and standard errors (SE) of tone perception performance by native English-speaking learners in two training groups at initial and final syllables of disyllable stimuli in generalization test.

A two-way repeated measures ANOVA was conducted, with Syllable Position (initial, final) as a within-subjects factor, and Training Group (monosyllabic training group, disyllabic training group) as a between-subjects factor.

The results yielded a main effect of Syllable Position [$F(1,14)=29.401$, $p<.001$], indicating that the learners did significantly better at tones on final syllable (78%) position than tones on initial syllable (63%) position that echoes to the finding in pre- and post-test.

A strong trend in interaction between Syllable Position X Training Group [$F(1,14)=4.327$, $p=.057$] was found. Post hoc analyses suggested the learners in monosyllabic training group did marginally better on final syllable (83%) than the learners in disyllabic training group (73%) ($p=.07$).

No other main effects and interactions were found.

4.3.2 **Effect of Tonal context**

Figure 19 shows the native English-speaking learners' tone perception performance in compatible and conflicting tonal context in the generalization test.

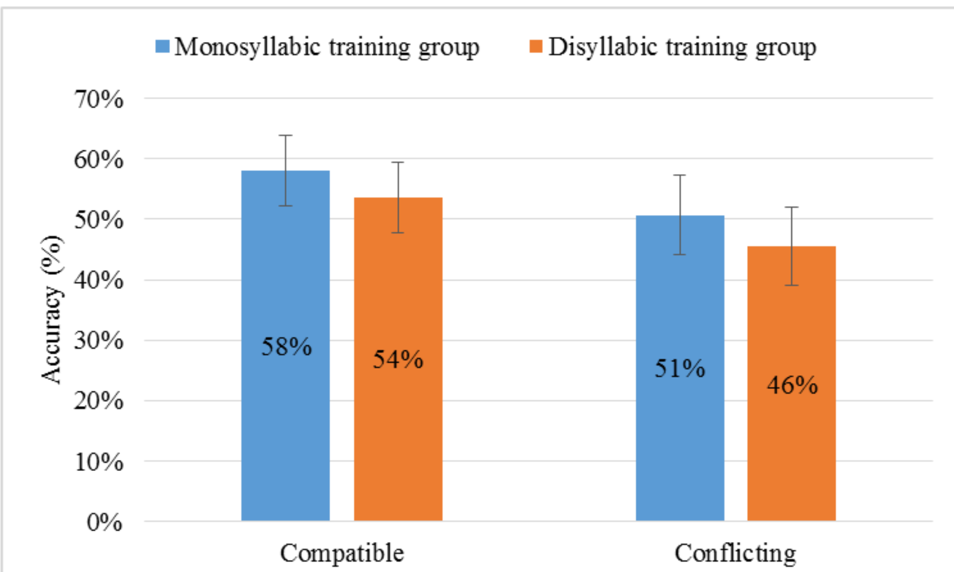


Figure 19: Percentage of accuracy and standard errors (SE) of tone perception performance by native English-speaking learners in two training groups in compatible and conflicting tonal contexts of disyllable stimuli in generalization test.

A two-way repeated measures ANOVA, with Tonal Context (compatible, conflicting) as within-subjects factor and Training Group as between-subjects factor, found a main effect of

Tonal Context [$F(1, 14)=6.672$, $p=.022$]. This result suggested that the learners did significantly better in compatible tonal context (56%) than conflicting tonal context (48%), which confirms the finding in pre- and post-test.

No other main effects or interactions were found.

4.3.3 **Effect of tonal sequence**

Figure 20 depicts the tone identification in the same and different tonal sequence by native English-speaking learners in the two training groups in the generalization test.

A two-way repeated measures ANOVA, with Tonal Sequence (same, different) as a within-subjects factor, and Training Group as a between-subjects factor, generated the learners' tonal performance.

A main effect of Tonal Sequence [$F(1, 14)=6.316$, $p=.025$] was found. This result indicated that the learners across two groups did significantly better on tones in the same tonal sequence (60%) than tones in the different tonal sequence (49%). Such result supports the previous finding in pre- and post-test.

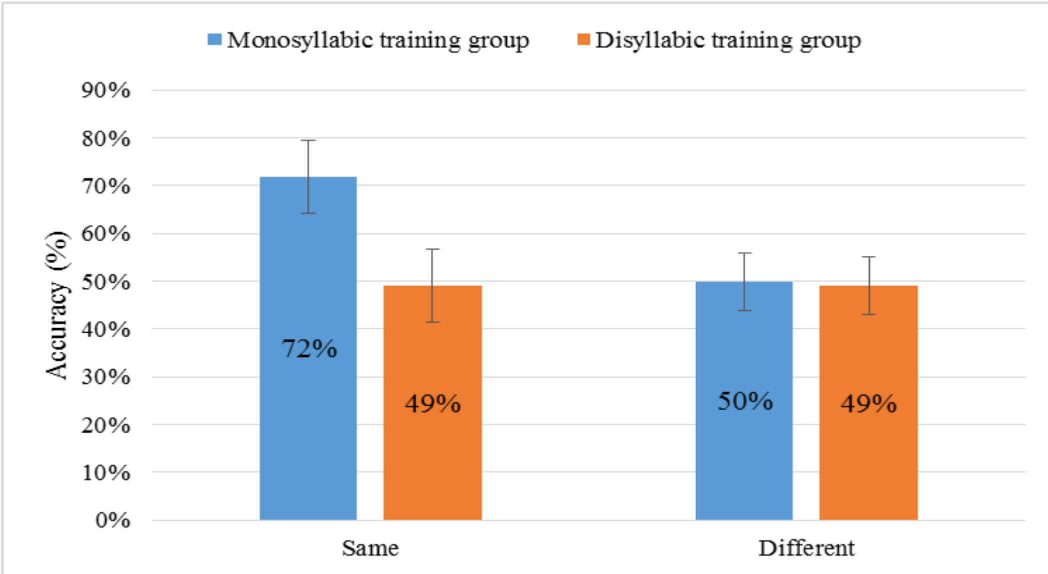


Figure 20: Accuracy rate and standard errors (SE) of the tone identification at the same and different tonal sequences by native English-speaking learners in two training groups in generalization test.

There was also a significant interaction of Tonal Sequence X Training Group [$F(1,14)=6.408$, $p=.024$]. Post hoc t-tests showed that the learners in the monosyllabic training group did significantly better at the same tonal sequence (72%) than the learners in disyllabic training group (49%) ($p=.025$). However, the two groups' performances at the different tonal sequence were comparable at 50% and 49% respectively.

Overall, in generalization test, the results in analyzing the three linguistic factors in disyllable stimuli demonstrated that the learners across both groups all did better on tones at the final syllable position, in the compatible tonal context and in the same tonal sequence. Such results were a similar pattern to what was found in the pretest and posttest.

Chapter 5: Chapter Five: Discussion and Conclusion

In this chapter, the findings from the present study are first summarized and discussed based on each research question. Second, the pedagogical implications are addressed regarding teaching Mandarin Chinese tones to adult native English-speaking language learners. Lastly, the limitations of the current study are discussed, and future research on the acquisition of Mandarin Chinese tones by native English-speaking learners is proposed.

5.1 Summary and discussion of the results for Research Questions

The accuracy means addressing in Research Question 1 and 2 are displayed in Table 8.

Table 8: Overall means and means of accuracy by two training groups from pretest to posttest.

Means of accuracy	Pretest	Posttest	P value	Statistically significant (p<.05)
Overall means (across two training groups and all stimuli)	60%	65%	p=.005	Yes, significant.
Means of monosyllabic training group	64%	68%	p=.005	Yes, significant.
Means of disyllabic training group	55%	63%	p=.002	Yes, significant.

5.1.1 Research Question 1: After perceptual training, will native English-speaking learners improve their perception of tones generally in both monosyllabic words and disyllabic words in Mandarin Chinese?

The results of the current study demonstrated that, through the short two-week of high variability perceptual training, adult native English-speaking learners of Chinese were able to significantly improve their tone perception in both monosyllable and disyllable stimuli in Mandarin Chinese. There was an effect of training shown by a significant 5% increase (*p*=.005) from pretest 60% to posttest 65% in learners' overall tone perception accuracy.

In addition, learners across the two training groups generally did significantly better (*p*<.001) when identifying tones in monosyllable stimuli, with an accuracy of 87%, than in disyllable stimuli, with an accuracy of 38%. Such large accuracy gap of tonal identification between the two types of stimuli was also observed by Sun (1998) and He (2010) in their American learners' tone identification tasks. Tones in monosyllable stimuli are in an isolated environment, which means that these tones are preserved in their canonical forms, while tones in disyllable stimuli are often coarticulated and the adjacent tones' pitch values affect each other (Shen, 1990; Xu, 1994, 1997, 1998). This difference in perception accuracy of the two types of stimuli suggests that, when teaching tones in Mandarin Chinese, teachers need to not just "solely focusing on teaching and learning monosyllabic tone" as Orton (2013) pointed out in her observations. Instead, teacher and learners should give more attention to tones in disyllable stimuli that contains more contextual variability, which mirrors the tones in real conversation more than tones in monosyllable stimuli do.

More importantly, in the current study, the effect of training was not only achieved in the old stimuli by old speakers, but also extended to the new stimuli by a new speaker. In the generalization test, native English-speaking learners' tone perception accuracy was 72%, which indicated that training was generalized to new stimuli by a new speaker with a substantial 12% increase (*p*=.005) compared to the pretest accuracy of 60%. These results are similar to those obtained in the tone training by Wang et al. (1999), who trained adult American learners using Chinese monosyllable stimuli and examined their tonal perception in old and new monosyllable tones.

5.1.2 Research Question 2: Compared monosyllabic perceptual training and disyllabic perceptual training, which one will be effective in helping English-speaking learners shape their tonal category and improve their tone perception of Mandarin Chinese?

The current findings did not show significant differences between the monosyllabic perceptual training group and the disyllabic perceptual training group from pretest to posttest. When identifying tones in monosyllable and disyllable stimuli, the monosyllabic training group increased its mean of accuracy from the pretest 64% to the posttest 68%, a significant increase of 4% (*p*=.028). Similarly, learners in the disyllabic training group made a significant improvement from the pretest 55% to the posttest 63% with 8% increase (*p*=.020). In other words, both monosyllabic and disyllabic perceptual trainings were helpful for learners to build up their tonal category in Mandarin Chinese and improve their tonal perception in general. While the difference did not reach significance, one can see that disyllabic training group made double the improvement (8%) on their tonal identification overall compared to the monosyllabic training group (4%). The disyllabic training group seemed, in general, to provide more effective learning on Mandarin Chinese tones to native English-speaking learners than the monosyllabic training group did.

5.1.3 **Research Question 3: Contrasting two types of training materials in the study, monosyllabic stimuli and disyllabic stimuli, which is more effective in improving monosyllabic tones? And which is more effective in improving disyllabic tones?**

In summary, though native English-speaking learners in both training groups made improvement in their tonal identification performance in general, when contrasting the two training materials (monosyllabic stimuli and disyllabic stimuli), the results showed some distinct patterns in the learners' performance (see Table 9 for the accuracy means by two training groups on two types of test stimuli).

Table 9: Means of accuracy on two types of test stimuli by two training groups from pretest to posttest

Test Stimuli	Training Group	Pretest accuracy (%)	Posttest accuracy (%)	P value	Statistical significance ($p<.05$)
Monosyllabic test stimuli	Monosyllabic training group	87%	92%	$p=.005$	Yes, significant.
	Disyllabic training group	82%	88%	$p=.081$	Yes, marginally significant.
Disyllabic test stimuli	Monosyllabic training group	43%	45%	$p=.54$	Not significant.
	Disyllabic training group	29%	38%	$p=.048$	Yes, significant.

5.1.3.1 ***Monosyllabic training group's performance on monosyllabic test stimuli***

The native English-speaking learners in the monosyllabic training group had accuracy rate of 87% in the pretest, and they increased to 92% in the posttest with a significant 5% improvement (p=.005) on the monosyllabic test stimuli. This finding confirmed Wang et al. (1999) study that through the monosyllabic perceptual training, American learners improved their tonal perception on monosyllabic tones in Mandarin Chinese significantly with a sizable 21% increase. The reason why Wang et al.'s improvement is greater than that of the present study is due to several possible reasons. First, Wang et al. provided more training sessions to their learners than the current study did. They provided 8 sessions (40 minutes per session) of high variability phonetic training in a two-week period of time while learners in this study only had 4 training sessions (30 minutes per session) in a two-week period. More training seems to generate more learning in this case. Second, when comparing the learners' identification performance in pretest from the two studies, one can see that the native English-speaking learners in Wang et al.'s study had a relatively low accuracy rate of 69% in the pretest, while the learners in the current study had a much higher accuracy rate of 87% compared to those in Wang et al.'s. That is to say, the learners in the current study were more advanced to begin with than those learners in Wang et al.'s study. Also, in posttest, learners from both studies had two very similar accuracies, Wang et al. with 90% and the current study with 92%. In another word, it is possible that the learners in Wang et al.'s study had more room for learning from pretest 69% to posttest 90% than the learners in the current study from 87% to 92%. Lastly, Wang et al. arranged their training stimuli in a pairwise manner, which allowed for a systematic increase in difficulty of tone contrasts while the current study only presented the randomized natural training stimuli to the learners. Therefore, the targeted practice on the pairwise tone training sessions might have given an extra boost for the tone learning in the Wang et al.'s study.

5.1.3.2 ***Disyllabic training group's performance on monosyllabic test stimuli***

Similarly, the learners in the disyllabic training group also made a marginally significant 6% increase from a pretest accuracy rate of 82% to a posttest accuracy rate of 88% (p=.081).

The results for the monosyllabic test stimuli also showed that there was a trend in the three-way interaction of Test X Tone X Training Group, which was triggered by the disyllabic training group's improvement in monosyllabic tones. After the disyllabic perceptual training, the disyllabic training group learners improved their tonal perception significantly for Tone1 from 74% in the pretest to 87% in the posttest, and marginally in Tone2 from77% in the pretest to 88% in the posttest. Such results suggest that the disyllabic perceptual training seems to elicit more improvement on individual tones in the monosyllabic test stimuli, specifically for Tone1 and Tone2, than did the monosyllabic perceptual training.

From the above results, one can see that both training groups seemed to help improve the tonal perception in monosyllable stimuli. In other words, training with either monosyllabic stimuli or disyllabic stimuli is beneficial for learners to identify tones in monosyllable stimuli.

5.1.3.3 *Individual tones in monosyllabic test stimuli*

Among four phonemic individual tones, after the training, the learners across both groups identified T4 (96%) significantly better than T1 (86%), T2 (84%) and T3 (84%). Similar results were also found in the generalization test that all learners did better on T1 (99%) and T4 (99%) than on T2 (89%) and T3 (86%). These findings support what has been found in previous studies that adult learners did not perceive the four tones in isolation equally well. Sun (1998) found American learners identified both T1 and T4 better than T2 and T3 in an isolated environment. Similarly, He (2010) also found that T2 was the worst identified in monosyllable stimuli by both her low-proficiency and high-proficiency American learners of Mandarin Chinese. T1 and T4 share high onset pitch values that is perceptually salient, thus, these two tones seem to be easier to identify by the learners than T2 and T3, which share low onset pitch values. Also, Lai and Zhang (2008) suggest that by using the isolation point (IP) to examine the time difference of identifying the four tones, the IP is the earliest for T1(a high register tone), followed by T4 (a high register tone), which is then followed by T2 and T3. In other words, the learners may also use the early perceptual processing when identifying four tones, thus, T1 and T4 were easier to identify than T2 and T3.

5.1.3.4 *Two training groups' overall performance on disyllabic test stimuli*

For the disyllabic test stimuli, results from pretest to posttest showed that the monosyllabic training group did not make a significant improvement in accuracy overall from pretest at 43% to posttest at 45%. However, the disyllabic training group made a significant improvement (p=.048) from pretest accuracy of 29% to posttest accuracy of 39% on the disyllabic test stimuli. These results suggested that when trained with disyllabic stimuli (as in the disyllabic training group), it significantly helped native English-speaking participants to learn the tones better than those trained with monosyllabic stimuli (as in the monosyllabic training group). For the disyllabic stimuli, the disyllabic training was much more effective in helping to acquire the tones.

The two training groups' tone identification performance was different at the two syllable positions. The results showed, from pretest to posttest and across two groups, at the first syllable position (σ1), T3 was found to be most difficult tone to identify with a low accuracy of 24%, followed by T2 (47%), T1 (62%) and T4 (76%); at the second syllable position (σ2), T2 had the lowest accuracy of 58% among four tones, then T1 (69%), T3 (72%) and T4 (80%). Similar results were also found in generalization test that T3 was the worst among four tones at the first syllable position while T2 was the worst at the second syllable position.

5.1.3.4.1   At the first syllable position

At the first syllable position in disyllable stimuli, the results showed learners across both training groups did significantly better (p=.022) after training with accuracy rate of 56% than pretest accuracy rate of 49%.

At the first syllable position, however, the monosyllabic training group did not make significant improvement on tones from the pretest accuracy of 53% to the posttest accuracy of 55%.

This seems to indicate that teaching learners the canonical form of Mandarin tones doesn't seem to help with their learning of tones in disyllable stimuli, at least for the tones at the first syllable position.

In contrast, the disyllabic training group, at the first syllable position, made a greater increase of accuracy at 12% (p=.070) from the pretest 45% to the posttest 57%, when compared to the monosyllabic training group's 2% increase from the pretest 53% to the posttest 55%.

At the first syllable position, it seems that the disyllabic training group was more effective in helping improve the learners' tone accuracy than the monosyllabic training group was.

5.1.3.4.2 At the second syllable position

The results of the tone identification by two training groups at the second syllable position demonstrated that the learners across groups did significantly better (p=.007) in the posttest with 73% of accuracy than in the pretest with 67% of accuracy. This demonstrated that both training were effective to help the learners identify tones at the second syllable position.

At the second syllable position, the monosyllabic training group scored from pretest 77% to posttest 78% without significant improvement.

At the second syllable position, the disyllabic training group made a significant improvement from pretest 56% to posttest 69% (p=.017).

Taken together, the disyllabic perceptual training, rather than the monosyllabic perceptual training elicited a significant improvement in tone perception, on the second syllable of the disyllabic test stimuli.

5.1.4 Research Question 4: Will training using monosyllabic material transfer to disyllabic tone identification? And will training using disyllabic material transfer to monosyllabic tone identification?

The transferring of the training effect was examined in both directions, namely, how learners in the monosyllabic training group identified tones in disyllable stimuli, and how learners in the disyllabic training group identified tones in monosyllable stimuli.

The monosyllabic training group did not make a significant increase from pretest accuracy of 43% to posttest accuracy of 45% when perceiving tones in disyllable stimuli. That is to say, there was no evidence to show the transfer of learning when the learners trained with monosyllabic materials had to identify disyllabic tones.

On the other hand, the disyllabic training group made a marginally significant improvement (p=.081) identifying monosyllabic tones from an accuracy rate of 82% in the pretest to 88% in the posttest. This finding clearly showed that there was a transfer of training shown by the learners, who were trained with disyllabic stimuli, and improved subsequently their tonal accuracy in monosyllabic tone identification. These findings provided new evidence for the transfer of a training effect, in which that the learners trained in the disyllabic training group improved their tonal perception on monosyllabic tones.

5.1.5 Research Question 5: Will factors, specifically syllable position, tonal context, and tonal sequence, affect native English-speaking learners' tone perception of disyllabic words?

In disyllabic words, it was found that three linguistic factors, syllable position, tonal context and tonal sequence, did affect learners' tone identification accuracy.

5.1.5.1 *Syllable position*

Averaged across two tests and syllable positions, the two training groups' tonal identification was comparable overall. From pretest to posttest, the results from learners' performance at initial and final syllable positions found that, across training groups, the learners did significantly better at the final syllable position (p<.001) with accuracy rate of 70% than did at the initial position with accuracy rate of 53%. Moreover, the learners in the monosyllabic training group did significantly better on tones at the final syllable position (p= .007) with accuracy of 78% than at the initial position with accuracy of 55%. The learners in disyllabic training group also perceived tones at the final syllable position (63%) better than at the initial position (52%). After the training, the improvement learners across groups made at the initial syllable position (13%) was significantly higher (p=.005) than at the final syllable position (2%). The similar results were also found in generalization test that learners across groups were better at perceiving the final tones (78%) than the initial tones (63%).

Overall, native English-speaking learners identified tones better at the second syllable position than at the first syllable position. This significance of tonal accuracy at the final syllable echoes to findings by Sun (1998), and He and Wayland (2013) when investigating tone identification in disyllabic words. Such pattern is probably due to a couple of reasons: in disyllable stimuli, the tones at the final syllable tend to have longer duration than those at the first syllable in natural production (Xu and Wang, 2009). Thus, the shape of the tone is more fully represented in the final position than at the initial position which contains shorter duration. The other reason may be due to a recency effect that the tones at the final syllable were heard most recently by learners compared to the tones at the initial syllable, so the learners were able to identify the tones at the final syllable better.

In terms of learning, the learners made more improvement on initial tones than on final tones. For instance, the monosyllabic training group increased their accuracy rates from pretest 48% to 62% after training at the initial position while the increase at the final position was from the pretest 74% to the posttest 81%. A similar tonal improvement showed up for the disyllabic training group as well. The disyllabic learners increased their accuracy rates from the pretest 46% to the posttest 57% at the initial position while at the final position the accuracy rates were from the pretest 63% to the posttest 62%. Such results demonstrated that training was effective, especially for the tones at the initial syllable position.

5.1.5.2 ***Tonal effect***

From pretest to posttest in disyllable stimuli, two tonal contexts, compatible and conflicting contexts, were investigated. The learners did significantly better in compatible tonal contexts (45%) than in conflicting tonal contexts (36%) with a 9% increase (p=.002). That is to say, the learners can identify tones in compatible tonal contexts better than in conflicting tonal contexts. Moreover, in generalization test, it was found the leaners across training groups identified tones better in compatible tonal contexts with accuracy rates of 56% as compared to the conflicting tonal contexts with accuracy rates of 48%. This finding in generalization test confirms the results in pretest and posttest that compatible tonal contexts are easier than conflicting tonal contexts for learners' tone identification.

The reason that the learners identified tones better in compatible contexts than in conflicting contexts may be due to the fact that the degree of adjustment between the two adjacent tones is relatively small in compatible contexts compared to conflicting contexts (Xu, 1994). As stated by Xu, a conflicting tonal context could substantially change the original tonal contours to the extent that they resemble some other tone categories. Thus, it is more difficult for leaners to identify tones that were distorted by conflicting contexts than tones in compatible contexts. The coarticulated tones that contain lots of tonal variations are difficult for learners to acquire within a short training period.

It is important, however, to realize that learners were better after training, and especially that the learners in the disyllabic training group made more improvement than those in the monosyllabic training group from pretest to posttest in both tonal contexts. From pretest to posttest, the disyllabic learners in compatible tonal contexts made a 10% increase from 35% to 45% while the monosyllabic learners barely made any improvement from 50% to 51%. Similarly, in conflicting tonal contexts, the disyllabic learners also increased 10% from 28% to 38% while the monosyllabic learners barely made any improvement from 39% to 40%. Overall, the disyllabic training seems to help the learners more when identifying both compatible and conflicting tones in disyllable stimuli than the monosyllabic training group did.

5.1.5.3 ***Tonal Sequence***

Tonal accuracies for the same and different tonal sequences in disyllable stimuli were also analyzed. It was found that the learners across the training groups did significantly better ($p<.001$) on the same tonal sequences (55%) than they did for different tonal sequences (37%). This finding was similar to the results from the generalization test with accuracy rate of 60% for the same tonal sequence and 49% for the different tonal sequence. However, this finding is different from what found by He (2010). In her results, she did not find a difference between the same and different tonal sequences by her American learners of Mandarin Chinese. In the further analysis on tones in same tonal sequence, she found that her learners did very poorly on T2+T2 and T4+T4 sequences. These two sets of same tonal sequences can also be categorized as conflicting tonal contexts, which may have created great difficulty for her learners across the two proficiency groups. .

In current study, the advantages showed in perceiving tones in the same tonal sequence may be due to a couple of reasons. The first one is the high variability phonetic training provided many exemplars of each tone to the learners, so that they could shape more robust tonal categories for all four phonemic tones after training, despite the contextual difference in these tone combinations, such as T1+T1, T2+T2, and T4+T4. The learners in the current study made great gains on tones in such same tonal sequences. The other possible reason may be due simply to the tonal repetition. For the current learners, who were at beginning level of language proficiency, it seems that same tonal sequences are easier.

From pretest to posttest, the learners in monosyllabic training group did considerably better in the same tonal sequences with accuracy of 68% than in the different tonal sequences with accuracy of 39%. This big difference between the two tonal sequences by the monosyllabic training group was found again in the generalization test with a 70% accuracy rate in the same tonal sequences and a 50% accuracy rate in the different tonal sequences. For the disyllabic training group learners, the difference in their performance on the two tonal sequences was not as great as the monosyllabic group learners. From pretest to posttest, the learners in disyllabic group had an accuracy rate of 43% in the same tonal sequences, and 35% in the different tonal sequences. In the generalization test, this difference was diminished with an equal accuracy rate of 49% at both tonal sequences.

Taken together, these findings demonstrate that the learners were generally good at perceiving tones in the same tonal sequences but bad at identifying tones in the different sequences, which embody lots of tonal coarticulation and variation. At the same time, the learners identified tones in compatible tonal contexts significantly better than in conflicting tonal contexts. Moreover, the learners perceived tones at the final syllables significantly better than those at the initial syllables. This result suggests that to improve native English-speaking learners' tonal perception of coarticulated tones, it is probably necessary to provide the learners with more perceptual training time on tones in the different tonal sequences than in the same tonal sequences, and tones in the different tonal contexts than in the same tonal contexts, and tones at the initial syllable position than at the final syllable position.

5.2 Pedagogical implication

The current study investigated the training effect by using a high variability phonetic training paradigm to facilitate native English-speaking learners to improve their tonal perception on Mandarin Chinese tones in monosyllabic and disyllabic words. The results of this study are of interest to both native English-speaking learners of Mandarin Chinese and Chinese language teachers. Such results provide a glimpse at the positive training results due to the high variability phonetic training on tone perception accuracy for native English-speaking learners of Chinese.

First, the results demonstrated that all learners improved their accuracy of tone identification significantly after the training, and this improvement was also found when perceiving new stimuli by new speakers. These data show that using a carefully designed perceptual training, learners are able to improve their tonal categorization in Mandarin Chinese in monosyllable stimuli similarly to those in Wang et al. (1999). More importantly, the present data extend these results to disyllable stimuli that have not been investigated before. Tones in disyllable stimuli more closely mirror the tones in real words and real conversation than tones in monosyllable stimuli. One may therefore conclude that the high variability phonetic training helped the learners improve their tonal perception in stimuli most resembling natural connected speech. For Chinese language teachers, this is great news that they can incorporate a training paradigm into their teaching lab to help learners of Chinese. The implementation is quite simple, without much technology training background needed for the teachers to add this into their curriculum.

Secondly, while the results show that the learners generally did better in monosyllabic tone identification than in disyllabic tone identification, the disyllabic training helped the learners more on both monosyllabic and disyllabic stimuli. These data suggest some needed changes in teaching Mandarin Chinese tone to native English speaking learners. The current in-classroom tone teaching is mainly focusing on using monosyllabic words for tonal contrasts practice (Orton, 2013). The data from the current study lend support for changes in current tone teaching in classroom. The results suggest that when the Chinese language teachers introduce the tones, they can introduce the four tones in isolation briefly, but then they should put more emphasis on introducing and practicing tones in disyllable stimuli, which carry a lot more tonal variations and coarticulation as in real conversations. Practice with disyllabic tones will not only help improve tone perception in monosyllable stimuli but also help the tones in disyllable stimuli.

Thirdly, the findings of the difficult tones when perceiving monosyllable stimuli and disyllable stimuli are meaningful for teaching and learning as well. It was found that tones in the first syllable poised more difficulty to the learners than those in the final syllables; tones in conflicting tonal contexts were harder to identify than those in compatible tonal contexts; and tones in the different tonal sequences were more challenging to perceive than those in the same tonal sequences. All above results provide a more focused and targeted direction for teachers to plan and design a more appropriate curriculum for teaching Mandarin Chinese tone.

Fourth, the current regarding the transferring of the training effect is equally important for teaching. It was found that the learners trained on disyllabic materials made great improvement in perceiving monosyllabic tones, specifically for T1 and T2. However, such a training transfer was not found for the learners who were trained on monosyllabic materials when they were to identify tones in disyllable stimuli. In fact, when identifying tones at the first syllables in disyllable stimuli, it was found that the learners in the monosyllabic training group decreased the accuracy of their Tone2 and Tone3. These results suggest that to help improve tone perception, maybe training only with monosyllabic tones is not enough. Adding disyllable stimuli that contain great variability of the tones in various phonetic environments produced by multiple native speakers in natural speech actually help the learners increase their overall tonal accuracy.

In conclusion, the current results demonstrated that some significant and effective improvements on native English-speaking learners' tonal perception in Mandarin Chinese were gained after a short 2-week of high variability phonetic training. The high variability phonetic training paradigm provided native English-speaking learners with crucial information about the language without explicit use of linguistic terminology. According to Molholt (1990), a traditional analysis of the target language's sound system and many linguistically-oriented terminologies confuse language learners who do not have training in linguistics. Therefore, many language learners, as well as language teachers, can be liberated from the complex linguistic explanations of the tones in different contexts, and can make tonal practice and learning happen in a stress-free environment. The flexible and short 30-minute training sessions used in current study are easy to access on computers by learners and can easily be incorporated into language practice by teachers. This computer-aided learning can provide learners with great convenience and self-learning efficiency, especially for beginning learners of Mandarin Chinese, who can benefit from not being given intricate lectures on tonal differences and tonal coarticulation which might discourage learners at this very beginning stage of learning the target language sound system.

5.3 Limitation and future research

The present study is the first to show that native English-speaking learners' tonal perception can be improved in disyllable stimuli by using a high variability phonetic training method.

In the current study, all participants were limited to beginning native English learners of Mandarin Chinese at a mid-west university who had less than two semesters' learning experience (considered as elementary level of proficiency) of the target language. The results of the current study cannot be generalized to the learners whose native language is not English but it is expected that similar patterns would be observed. Neither can the current results be generalized to the learners whose Chinese language proficiency is above or below elementary level. It is suggested that future studies could investigate different groups of learners (not just native English speakers) and that learners at different language proficiency levels, using the same perceptual training instruction to facilitate the training effect of improving the learners' tonal perception. It is hypothesized that similar improvements will be found.

Though both Wang et al. (1999) and the current study showed a significant training effect, Wang et al. showed a greater 21% increase comparing to this study's 5% increase. The improvement difference of the two studies may be due to a couple of reasons: longer training time and more training sessions. In the present study, the learners only had four training sessions and 30 minutes per session, less than half of the time in Wang et al. As a consequence, the fewer and shorter training sessions probably resulted in less robust significant training effects. Future studies should increase the frequency of the training sessions, as well as the training duration for each session in order to optimize observing training effects.

Finally, a production study of the learners' tonal performance should be included in future studies to determine if perceptual training effects transfer to production. Wang et al. (2003) investigated American learners' monosyllabic tone production performance after their successful perceptual training in monosyllabic tones. They found that the learners transferred tone learning to the production domain, which indicates that the new tonal categories have formed in the learners' speech system. Moreover, Herd et al. (2013) investigated the English learners' perception and production of Spanish intervocalic sounds /d, r, ɾ/, and showed that with perception-only high variability training, the

English learners' target sound production significantly improved as well. Since the purpose of learning a foreign language is to communicate, it is suggested that future studies could explore native English-speaking learners' tone production performance in both monosyllable stimuli and disyllable stimuli in order to see if perception training can be transferred to the production domain and how perception and production interacts with types of training.

5.4 Conclusion

This study investigated whether native speakers of English can be trained using a high variability phonetic training method to accurately perceive Mandarin Chinese tones in monosyllable stimuli and disyllable stimuli. The perception results clearly showed that the learners improved their tone accuracy for both monosyllable and disyllable stimuli after a short period of perceptual training. Additionally, this study investigated which training group, monosyllabic training group or disyllabic training group, would be most helpful for native English-speaking learners to establish tonal categories in their speech system. Although both groups' identification performance improved, it was found that the learners in the disyllabic training group seemed to show more learning not only on disyllabic tones but also on monosyllabic tones. Moreover, the learners in the monosyllabic training group showed little training effects for disyllabic tones but only showed improvement for monosyllabic tones. Disyllabic tones with tonal variation and coarticulation can help learners. Future tone teaching in Mandarin Chinese classes should switch the focus from teaching tones in isolation in monosyllable stimuli to tones which include coarticulation (as in disyllable stimuli) in order to improve learning and better simulate natural and realistic learning environments.

References:

Chao, Y. R., & Pian, R. C. (1955). *Mandarin primer* (p. 25). Folkways Records.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English/r/and/l: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299-2310.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. *Perception & psychophysics*, *61*(5), 977-985.

Chang, Y.-h. S. (2011). Distinction between Mandarin Tones 2 and 3 for L1 and L2 Listeners. In Z. Jing-Schmidt (Ed.), *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23). 1*, pp. 84-96. Eugene: University of Oregon.

Duanmu, S. (1999). Stress and the development of disyllabic words in Chinese. *Diachronica,* vol. 16 (1), 1-35.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268-294.

Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, *25*(2), 207-231.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95-112.

Guo, L., & Tao, L. (2008, April). Tone production in Mandarin Chinese by American students: A case study. In *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)* (Vol. 1, pp. 123-138).

Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*(2), 269-279.

He, Y. J. (2010). Perception and production of isolated and coarticulated Mandarin Tones by American learners. University of Florida, Gainesville.

He, Y., & Wayland, R. (2010). The production of Mandarin coarticulated tones by inexperienced and experienced English speakers of Mandarin. In *Speech Prosody 2010-Fifth International Conference*.

He, Y., & Wayland, R. (2013). Identification of Mandarin coarticulated tones by inexperienced and experienced English learners of Mandarin. *Chinese as a Second Language Research*, *2*(1), 1-21.

Herd, W., Jongman, A., & Sereno, J. (2013). Perceptual and production training of intervocalic/d, ɾ, r/in American English learners of Spanish. *The Journal of the Acoustical Society of America*, *133*(6), 4247-4255.

Hiller, S., Rooney, E., Laver, J., & Jack, M. (1993). SPELL: An automated system for computer-aided pronunciation teaching. *Speech Communication*,*13*(3), 463-473.

Hubbard, P. (Ed.). (2009). *Computer Assisted Language Learning: Critical Concepts in Linguistics. Present Trends and Future Directions in CALL*. Routledge.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267-3278.

Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). *Perception and production of Mandarin Chinese tones*. *Handbook of Chinese Psycholinguistics*. Li, P., Tan, L., Bates, E. & O.J.L Tzeng (Eds.). Cambridge University Press.

Kingston, J. (2003). Learning foreign vowels. *Language and Speech*, *46*(2-3), 295-348.

Lai, Y., & Zhang, J. (2008). Mandarin lexical tone recognition: The gating paradigm. *Kansas Working Papers in Linguisitcs*, 183-198.

Lee, C. Y., Tao, L., & Bond, Z. S. (2010). Identification of multi-speaker Mandarin tones in noise by native and non-native listeners. *Speech Communication*, *52*(11), 900-910.

Lee, C. Y., Tao, L., & Bond, Z. S. (2010). Identification of acoustically modified Mandarin tones by non-native listeners. *Language and speech*, *53*(2), 217-243.

Lin, Y. H. (2007). *The Sounds of Chinese.* Cambridge University Press.

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.

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.

Miracle, W. C. (1989). Tone production of American students of Chinese: A preliminary acoustic study. *Journal of the Chinese Language Teachers Association*, *24*(3), 49-65.

Molholt, G. (1988). Computer-Assisted Instruction in Pronunciation for Chinese Speakers of American English. *TESOL Quarterly*, *22*(1), 91–111. http://doi.org/10.2307/3587063

Molholt, G. (1990). Spectrographic analysis and patterns in pronunciation. *Computers and the Humanities*, *24*(1-2), 81-92.

Orton, J. (2013). Developing Chinese oral skills-a research base for practice. *Research in Chinese as a second language*, 3-26.

Quintana Lara, M. (2009). *The effects of Acoustic VisualFeedback Instruction on pronunciation of the English high front vowels by pre-service non-native English teachers* (Doctoral dissertation, UNIVERSITY OF KANSAS).

Shen, X. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association*, *24*(3), 27-47.

Shen, X.N.S. (1990). The prosody of Mandarin Chinese. *Linguistics* (Vol.118). Berkeley, California: University of California Press.

Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and speech*, *34*(2), 145-156.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of/rl/by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131-145.

Sun, S. H. (1998). *The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese* (No. 16). University of Hawaii Press.

Tagliaferri, B. (2008). Paradigm: Perception Research Systems [Computer Program]. *Retrieved from h ttp://www. paradigmexperiments. com*.

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.

Wang, Y., Jongman, A., and Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after training. *Journal of the Acoustical Society of America*, 113, 1033-1043.

Xing, J. Z. (2006). *Teaching and learning Chinese as a foreign language: A pedagogical grammar* (Vol. 1). Hong Kong University Press.

Xu, Y., & Wang, M. (2009). Organizing syllables into groups—Evidence from F 0 and duration patterns in Mandarin. *Journal of Phonetics*, *37*(4), 502-520.

Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, *95*, 2240.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of phonetics*, *25*(1), 61-83.

Zhang, J. (2007). A directional asymmetry in Chinese tone sandhi systems. *Journal of East Asian Linguistics*, *16*(4), 259-302.

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, *27*(01), 153-201.

Zhou, X., Marslen-Wilson, W., Taft, M., & Shu, H. (1999). Morphology, orthography, and phonology reading Chinese compound words. *Language and cognitive processes*, *14*(5-6), 525-565.

Appendix A: Language Background Questionnaire for English Learners of Chinese

Gender: _________

Age: _________

Native country/state: ______________

Date: __________________

What year are you? Year ___ of ☐ undergraduate ☐ graduate studies.
What is **your native** language? _________________________________

What is **your mother's** native language? _________________________________

What is **your father's** native language? _________________________________

Part I:
Knowledge of the MANDARIN language:

1. How old were you when you took your first course in Mandarin Chinese?

Experience with Mandarin instruction

	Number of years studying Mandarin	Hours of Mandarin instruction per week
Elementary school		
Middle school		
High school		
University		

2. Describe the formal instruction you are currently receiving in learning Mandarin Chinese language here at KU. Indicate course title and number of hours each course meets per week.

Course Title Number of Contact Hours

a.____

b.____

c.____

d.____

3. Have you ever used Mandarin Chinese outside of the classroom in any informal settings? If "yes", please check and provide an approximate time of the use.

____Practicing/talking Chinese with Chinese friends, ____________ hour(s) per week

____Listening to Chinese music, ___________ hour(s) per week

____Watching Chinese TV, ____________hour(s) per week

____Reading Chinese magazines/newspapers, ____________ hour(s) per week

____Traveling to China, ________ time(s) per year, for ______ days.

4. Do you have a foreign accent in Mandarin? Yes No
→ If yes, please rate the strength of your accent.

☐ No Accent ☐ Slight Accent

☐ Moderate Accent ☐ Strong Accent

5. Rank-order the four individual tones (T1, T2, T3, T4) from left to right according to the "easiest" to the "most difficult" for you to learn.

Easiest _____, _____, _____, _____, most difficult

Part II

Knowledge of OTHER languages:

Write the name of the language in the blank, and indicate your approximate abilities in each of the four areas for each language.

1. **Language: ________________________**

Speaking	***Listening***	***Reading***	***Writing***
☐ **Poor**	☐ **Poor**	☐ **Poor**	☐ **Poor**
☐ **Fair**	☐ **Fair**	☐ **Fair**	☐ **Fair**
☐ **Good**	☐ **Good**	☐ **Good**	☐ **Good**
☐ **Near-Native**	☐ **Near-Native**	☐ **Near-Native**	☐ **Near-Native**

How long have you learned/ been learning the above languages?

2. **Language: ________________________**

Speaking	***Listening***	***Reading***	***Writing***
☐ **Poor**	☐ **Poor**	☐ **Poor**	☐ **Poor**
☐ **Fair**	☐ **Fair**	☐ **Fair**	☐ **Fair**
☐ **Good**	☐ **Good**	☐ **Good**	☐ **Good**
☐ **Near-Native**	☐ **Near-Native**	☐ **Near-Native**	☐ **Near-Native**

How long have you learned/ been learning the above languages?

3. **Language: ________________________**

Speaking	***Listening***	***Reading***	***Writing***

☐ **Poor**	☐ **Poor**	☐ **Poor**	☐ **Poor**
☐ **Fair**	☐ **Fair**	☐ **Fair**	☐ **Fair**
☐ **Good**	☐ **Good**	☐ **Good**	☐ **Good**
☐ **Near-Native**	☐ **Near-Native**	☐ **Near-Native**	☐ **Near-Native**

How long have you learned/ been learning the above languages?

Appendix B: Language Background Questionnaire for Native Chinese Speakers

Gender: _________

Age: _________

Native country/state: ______________

Date: __________________

Part I

1. What is the language you use at home? (If not Mandarin Chinese, please specify the dialect, such as Cantonese, Wu dialect, etc.)

2. What is the main language you use with your friends?

3. When did you start learning English?

4. Experience with English instruction

	Number of years studying English
Elementary school	
Middle school	
High school	
University	

5. How long you have been in the United States?

Part II--Knowledge of OTHER languages:

1. Write the name of the language in the blank, and indicate your approximate abilities in each of the four areas for each language.

a. Language: ______________________

Speaking	*Listening*	*Reading*	*Writing*
☐ **Poor**	☐ **Poor**	☐ **Poor**	☐ **Poor**
☐ **Fair**	☐ **Fair**	☐ **Fair**	☐ **Fair**
☐ **Good**	☐ **Good**	☐ **Good**	☐ **Good**
☐ **Near-Native**	☐ **Near-Native**	☐ **Near-Native**	☐ **Near-Native**

How long have you learned the above language?

b. Language: ______________________

Speaking	*Listening*	*Reading*	*Writing*
☐ **Poor**	☐ **Poor**	☐ **Poor**	☐ **Poor**
☐ **Fair**	☐ **Fair**	☐ **Fair**	☐ **Fair**
☐ **Good**	☐ **Good**	☐ **Good**	☐ **Good**
☐ **Near-Native**	☐ **Near-Native**	☐ **Near-Native**	☐ **Near-Native**

How long have you learned the above language?

c. Language: ______________________

Speaking	*Listening*	*Reading*	*Writing*
☐ **Poor**	☐ **Poor**	☐ **Poor**	☐ **Poor**
☐ **Fair**	☐ **Fair**	☐ **Fair**	☐ **Fair**
☐ **Good**	☐ **Good**	☐ **Good**	☐ **Good**
☐ **Near-Native**	☐ **Near-Native**	☐ **Near-Native**	☐ **Near-Native**

How long have you learned the above language?

Appendix C: Pretest and Posttest Test Stimuli

a) 96 Monosyllabic test stimuli

	Character	Pinyin
1	参	cān
2	出	chū
3	窗	chuāng
4	嘬	zuō
5	低	dī
6	发	fā
7	姑	gū
8	郭	guō
9	憨	hān
10	齁	hōu
11	靴	xuē
12	啷	lāng
13	孬	nāo
14	拍	pāi
15	抨	pēng
16	铅	qiān
17	敲	qiāo
18	切	qiē
19	区	qū
20	烧	shāo
21	推	tuī

22	香	xiāng
23	星	xīng
24	淤	yū
25	蚕	cán
26	除	chú
27	床	chuáng
28	昨	zuó
29	敌	dí
30	罚	fá
31	轱	gú
32	国	guó
33	寒	hán
34	猴	hóu
35	学	xué
36	狼	láng
37	挠	náo
38	排	pái
39	棚	péng
40	钱	qián
41	桥	qiáo
42	茄	qié
43	渠	qú
44	勺	sháo

45	颓	tuí
46	翔	xiáng
47	型	xíng
48	鱼	yú
49	惨	căn
50	楚	chŭ
51	底	dĭ
52	法	fă
53	古	gŭ
54	裹	guŏ
55	喊	hăn
56	吼	hŏu
57	雪	xuě
58	朗	lăng
59	脑	năo
60	迫	păi
61	捧	pěng
62	浅	qiăn
63	巧	qiăo
64	且	qiě
65	取	qŭ
66	少	shăo
67	腿	tuĭ

68	想	xiǎng
69	醒	xǐng
70	雨	yǔ
71	灿	càn
72	处	chù
73	创	chuàng
74	做	zuò
75	地	dì
76	发	fà
77	故	gù
78	过	guò
79	汗	hàn
80	后	hòu
81	穴	xuè
82	浪	làng
83	闹	nào
84	派	pài
85	碰	pèng
86	歉	qiàn
87	翘	qiào
88	窃	qiè
89	去	qù
90	哨	shào

91	退	tuì
92	向	xiàng
93	姓	xìng
94	玉	yù
95	闯	chuǎng
96	左	zuǒ

b) 48 Disyllabic test stimuli

	Character	Pinyin
1	敲 香	qiāo xiāng
2	抨 出	pēng chū
3	孬 星	nāo xīng
4	憨 猴	hān hóu
5	低 钱	dī qián
6	发 狼	fā láng
7	郭 且	guō qiě
8	齁 底	hōu dǐ
9	靴 雪	xuē xuě
10	铅 汗	qiān hàn
11	拍 过	pāi guò
12	参 闹	cān nào
13	茄 区	qié qū
14	罚 姑	fá gū

15	勺烧	sháo shāo
16	昨国	zuó guó
17	翔除	xiáng chú
18	排床	pái chuáng
19	敌腿	dí tuǐ
20	挠巧	náo qiǎo
21	蚕少	cán shǎo
22	型退	xíng tuì
23	轱处	gú chù
24	桥窃	qiáo qiè
25	闯切	chuǎng qiē
26	迫窗	pǎi chuāng
27	吼淤	hǒu yū
28	脑寒	nǎo hán
29	浅学	qiǎn xué
30	左鱼	zuǒ yú
31	喊捧	hǎn pěng
32	醒古	xǐng gǔ
33	取法	qǔ fǎ
34	楚派	chǔ pài
35	雨穴	yǔ xuè
36	裹玉	guǒ yù
37	地啷	dì lāng

38	做 嘬	zuò zuō
39	发 推	fà tuī
40	姓 渠	xìng qú
41	碰 棚	pèng péng
42	翘 颓	qiào tuí
43	向 朗	xiàng lǎng
44	浪 惨	làng cǎn
45	哨 想	shào xiǎng
46	歉 故	qiàn gù
47	后 灿	hòu càn
48	去 创	qù chuàng

Appendix D: Training Stimuli

a) Monosyllabic Training Stimuli

	Character	Pinyin			
1	杯	bēi	19	潘	pān
2	奔	bēn	20	秋	qiū
3	参	cēn	21	缺	quē
4	吹	chuī	22	搔	sāo
5	春	chūn	23	沙	shā
6	聪	cōng	24	他	tā
7	粗	cū	25	窝	wō
8	爹	diē	26	先	xiān
9	蹲	dūn	27	熏	xūn
10	刚	gāng	28	真	zhēn
11	沟	gōu	29	中	zhōng
12	喝	hē	30	洲	zhōu
13	尖	jiān	31	棕	zōng
14	京	jīng	32	钻	zuān
15	究	jiū	33	层	céng
16	抠	kōu	34	锤	chuí
17	哭	kū	35	纯	chún
18	咧	liē	36	从	cóng

37	攒	cuán	57	人	rén
38	叠	dié	58	荣	róng
39	儿	ér	59	柔	róu
40	横	héng	60	如	rú
41	华	huá	61	谁	shuí
42	来	lái	62	雄	xióng
43	连	lián	63	轴	zhóu
44	铃	líng	64	足	zú
45	峦	luán	65	胆	dăn
46	埋	mái	66	顶	dĭng
47	门	mén	67	懂	dŏng
48	民	mín	68	短	duăn
49	农	nóng	69	盹	dŭn
50	奴	nú	70	耳	ĕr
51	挪	nuó	71	巩	gŏng
52	盘	pán	72	管	guăn
53	陪	péi	73	井	jĭng
54	求	qiú	74	卷	juăn
55	瘸	qué	75	苦	kŭ
56	燃	rán	76	脸	liăn

77	咧	liě	97	被	bèi
78	领	lǐng	98	笨	bèn
79	鲁	lǔ	99	彻	chè
80	美	měi	100	醋	cù
81	敏	mǐn	101	篡	cuàn
82	染	rǎn	102	蛋	dàn
83	扰	rǎo	103	冻	dòng
84	惹	rě	104	段	duàn
85	忍	rěn	105	共	gòng
86	冗	rǒng	106	贺	hè
87	软	ruǎn	107	横	hèng
88	扫	sǎo	108	话	huà
89	耍	shuǎ	109	件	jiàn
90	水	shuǐ	110	旧	jiù
91	我	wǒ	111	倦	juàn
92	朽	xiǔ	112	扣	kòu
93	枕	zhěn	113	赖	lài
94	肿	zhǒng	114	妹	mèi
95	爪	zhuǎ	115	面	miàn
96	总	zǒng			

116	念	niàn	122	肉	ròu
117	弄	nòng	123	入	rù
118	诺	nuò	124	煞	shà
119	配	pèi	125	涮	shuàn
120	绕	rào	126	踏	tà
121	热	rè	127	绣	xiù
			128	拽	zhuài

b) 64 Disyllabic training stimuli

Character		Pinyin			
1.	沙 聪	shā cōng	5.	洲 农	zhōu nóng
2.	秋 吹	qiū chuī	6.	他 纯	tā chún
3.	熏 窝	xūn wō	7.	咧 谁	liē shuí
4.	喝 奔	hē bēn	8.	缺 峦	quē luán

9. 中 我 zhōng wǒ

10. 棕 短 zōng duǎn

11. 春 肿 chūn zhǒng

12. 究 井 jiū jǐng

13. 搔 彻 sāo chè

14. 京 肉 jīng ròu

15. 杯 弄 bēi nòng

16. 钻 绕 zuān rào

17. 陪 抠 péi kōu

18. 连 蹲 lián dūn

19. 足 尖 zú jiān

20. 从 刚 cóng gāng

21. 人 横 rén héng

22. 如 攒 rú cuán

23. 盘 埋 pán mái

24. 门 荣 mén róng

25. 柔 领 róu lǐng

26. 来 枕 lái zhěn

27. 叠 扫 dié sǎo

28. 挪 软 nuó ruǎn

29. 轴 热 zhóu rè

30. 雄 笨 xióng bèn

31. 层 醋 céng cù

32. 儿 蛋 ér dàn

33. 染 沟 rǎn gōu

34. 顶 先 dǐng xiān

35. 巩 潘 gǒng pān

36. 咧 哭 liě kū

37. 脸 奴 liǎn nú

38. 爪 瘸 zhuǎ qué

39. 胆 华 dǎn huá

40. 管 民 guǎn mín

41. 敏 卷 mǐn juǎn

42. 惹 冗 rě rǒng

43. 水 盹 shuǐ dǔn

44. 总 美 zǒng měi

45. 耍 倦 shuǎ juàn

46. 耳 诺 ěr nuò

47. 扰 踏 rǎo tà

48. 鲁 妹	lǔ mèi	57. 冻 朽	dòng xiǔ
49. 贺 参	hè cēn	58. 涮 苦	shuàn kǔ
50. 赖 真	lài zhēn	59. 段 忍	duàn rěn
51. 横 粗	hèng cū	60. 篡 懂	cuàn dǒng
52. 被 爹	bèi diē	61. 件 扣	jiàn kòu
53. 面 求	miàn qiú	62. 绣 拽	xiù zhuài
54. 配 锤	pèi chuí	63. 念 共	niàn gòng
55. 旧 燃	jiù rán	64. 入 话	rù huà
56. 煞 铃	shà líng		

Appendix E: Generalization Test Stimuli

a) Monosyllabic stimuli

1 冲

chōng

2 托

tuō

3 薛

xuē

4 欢

huān

5 加

jiā

6 溜

liū

7 扑 pū

8 深

shēn

9 诗

shī

10 涛

tāo

11 挖

wā

12 弯

wān

13 西

xī

14 央

yāng

15 幽

yōu

16 亏

kuī

17 虫

chóng

18 陀

tuó

19 学

xué

20 环

huán

21 夹

jiá

22 留

liú

23 葡

pú

24 神

shén

25 时

shí

26 淘

táo

27 娃

wá

130

28 玩

wán

29 习

xí

30 杨

yáng

31 游

yóu

32 葵

kuí

33 宠

chǒng

34 妥

tuǒ

35 血

xuě

36 缓

huǎn

37 假

jiǎ

38 柳

liǔ

39 普

pǔ

40 沈

shěn

41 史

shǐ

42 讨

tǎo

43 瓦

wǎ

44 晚

wǎn

45 洗 xǐ

46 养

yǎng

47 有

yǒu

48 魁

kuǐ

49 铳

chòng

50 唾

tuò

51 穴

xuè

52 幻

huàn

53 嫁

jià

54 遛

liù

55 曝 pù

56 肾

shèn

57 世

shì

58 套 tào	61 系 xì	64 溃 kuì
59 袜 wà	62 样 yàng	
60 万 wàn	63 右 yòu	

b) Disyllabic stimuli

1	弯幽	wān yōu	14	陀洗	tuó xǐ
2	扑冲	pū chōng	15	杨穴	yáng xuè
3	托习	tuō xí	16	环遛	huán liù
4	亏留	kuī liú	17	沈欢	shěn huān
5	西史	xī shǐ	18	讨加	tǎo jiā
6	溜缓	liū huǎn	19	养娃	yǎng wá
7	挖幻	wā huàn	20	有神	yǒu shén
8	央袜	yāng wà	21	假晚	jiǎ wǎn
9	葡诗	pú shī	22	瓦妥	wǎ tuǒ
10	淘涛	táo tāo	23	宠世	chǒng shì
11	夹葵	jiá kuí	24	血右	xuě yòu
12	虫时	chóng shí	25	曝薛	pù xuē
13	游柳	yóu liǔ	26	溃深	kuì shēn

27 系 学 xì xué

28 万 玩 wàn wán

29 套 普 tào pǔ

30 嫁 魁 jià kuǐ

31 铳 样 chòng yàng

32 肾 唾 shèn tuò