# Comparing Monosyllabic and Disyllabic Training in Perceptual Learning of Mandarin Tone

Yingjie Li
University of Kansas

Goun Lee
Sungkyunkwan University

Joan A. Sereno
University of Kansas

**Abstract**
Although computer-assisted auditory perceptual training has been shown to be effective in learning Mandarin Chinese tones in monosyllabic words, tone learning has not been systematically investigated in disyllabic words. In the current study, seventeen native English-speaking beginning learners of Chinese were trained using a high variability phonetic training paradigm. Two perceptual training groups, a monosyllabic training group and a disyllabic training group, were compared and accuracy in identifying the tonal contrasts in naturally produced monosyllabic and disyllabic words (produced by native Mandarin Chinese speakers) was evaluated. Results showed that after only four training sessions in a two-week period, beginning learners of Chinese significantly increased their tonal identification accuracy from the pretest (72%) to posttest (80%). The current findings overall show significant differences between the monosyllabic perceptual training group and disyllabic perceptual training group. Although native English-speaking learners in both training groups made improvements in their tonal identification performance in general, when examining learning for the two types of stimuli (monosyllabic and disyllabic stimuli), the results showed distinct patterns in learners' performance. While both training groups improved tonal perception, training with disyllabic stimuli (disyllabic training group) was much more effective (especially for the disyllabic

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), ***A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn*** (pp. 303-319). Dept. of English, School of Communication & Culture, Aarhus University.
© The author(s), 2019.

stimuli) and significantly helped native English-speaking participants to acquire the tones. These results illustrate limitations of the current tone teaching based solely on monosyllabic words. Instead, the current results advocate for incorporating more common and variable disyllabic words in the classroom in order to achieve native-like tone acquisition.

## 1. Introduction

While it is important for language learners to acquire the correct pronunciation of a target language (Jenkins, 2004), it is especially crucial to acquire native-like pronunciation of tone for language learners of Chinese since Mandarin Chinese is a tonal language in which tone is a key component of the lexicon used to distinguish word meaning. Accurately perceiving and correctly producing tones is of critical importance for Chinese language learners to communicate successfully in the language. In the present study, American learners of Chinese were trained using a high variability phonetic training paradigm, in which two training groups were contrasted: a group trained with monosyllabic stimuli and a group trained with disyllabic stimuli. Accuracy in identifying tonal contrasts before and after training in naturally produced monosyllabic and disyllabic words (produced by native Mandarin Chinese speakers) was evaluated.

## 2. Perceptual training

Native English learners of Chinese have difficulty perceiving and producing tones in Mandarin Chinese since the phonemic tone feature is not in part of their native language system (Miracle, 1989; Shen, 1989; Shen & Lin, 1991; Sun, 1998; Jongman, Wang, Moore, & Sereno, 2006; Lee, Tao, & Bond, 2010; He, 2010; He & Wayland, 2010, 2013; Chang, 2011; Hao, 2012). These studies have analyzed native English learners' perception of Mandarin tone in isolation and found that American listeners have particular difficulty differentiating Tone 2 (T2) and Tone 3 (T3), attributing the confusion to American listeners assigning more weight to F0 height than F0 direction when perceiving Mandarin T2 and T3 in isolation (Gandour, 1983; Gottfried and Suiter, 1997; Lee, Tao, & Bond, 2009).

While it is vital to understand tones of monosyllabic words in an isolated environment, tones in disyllabic words are equally, or even more significant, since disyllabic words are dominant in the vocabulary of modern Mandarin Chinese (Duanmu, 1999). Disyllabic words and their connected tones are used most often in Chinese people's daily life rather than monosyllabic words, with disyllabic tones mirroring the tones perceived and produced at the sentence level more than isolated tones. Few studies have examined tones in disyllabic words and tones at the sentential level (Sun, 1998; Guo & Tao, 2008; He, 2010; He & Wayland, 2010, 2013). These researchers found, as expected, that across learning experience and proficiency level, American learners did significantly better at identifying tones in monosyllabic words than in disyllabic words. Moreover, native English learners' accuracy rate of tone perception was systematically improved according to their learning experience: the higher the proficiency level or the longer they studied Mandarin Chinese, the better their accuracy was. When examining learners' identification performance of the four phonemic tones across both monosyllabic and disyllabic words, Sun (1998) found that T2 and T3 were identified significantly poorer than T1 and T4 across all four proficiency level groups. Similarly, He (2010) found that for both monosyllabic and disyllabic tonal contrasts, T3 was the most difficult to identify, then T1, T2 and T4 by inexperienced learners while T2 was the most difficult to identify among the four tones by experienced learners.

Improving native English learners' tonal categories in Mandarin Chinese from the onset of learning the language is clearly important. Moreover, learning not only should pay attention to monosyllabic tones but also should focus on disyllabic tone practice, including tone alternation and coarticulation among the two adjacent tones. These coarticulated tones regularly occur in Mandarin Chinese natural speech contexts, and by examining disyllabic words, English speakers may be able to improve their comprehension and pronunciation of Mandarin.

Current in-class pedagogical approaches to teach Mandarin Chinese tones often use traditional methods, such as listen-and-repeat, minimal-pair drills, and reading aloud tasks. However, short-term auditory training methods in various languages have proved to be effective in assisting learners to acquire new phonetic contrasts that do not exist in their native phonological system (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Francis, Ciocca, Ma, & Fenn, 2008; Herd, Jongman, & Sereno, 2013; Kingston, 2003; Lively, Logan, & Pisoni, 1993; Logan, Lively & Pisoni, 1991; Wang, Spence, Jongman, & Sereno, 1999).

In these high variability training procedures, language learners listen to a large variety of stimuli produced naturally by multiple native speakers of the target language. Within a short period, the learners' perception of non-native language contrasts is improved through the exposure to the target language. Furthermore, this perceptual improvement was successfully extended to the learners' production, as shown by Japanese learners of English learning /ɹ/ and /l/ (Bradlow, et al., 1997, Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Logan, et al., 1991; Lively, et al., 1993).

High variability phonetic training is not only effective at the segmental level but also at the suprasegmental level (Wang, Jongman & Sereno, 2003; Wang et al., 1999). Wang et al. (1999) found that American learners of Mandarin Chinese showed improved tone perception after training, from a pretest accuracy rate of 69% to a posttest accuracy rate of 90%, a significant improvement (21%). Wang et al. (1999) used eight 40-minute training sessions and showed improved perception of Mandarin monosyllable tonal categories. Furthermore, they found that trainees also showed generalization of the learning to new words and new speakers (Wang et al., 1999). This improvement was also retained six months after training. Wang et al. (2003) extended this perceptual improvement to Mandarin tone production. Using perceptual training techniques, the production data showed that learners' pitch contours better approximated native norms after training. Additionally, identification of trainees' post-test tone productions (compared to their pre-test productions) improved by 18%. These results indicate improved tone identification accuracy and better productions after a short perceptual training period.

Wang et al. (1999, 2003) found that through a short high variability phonetic training using monosyllabic tones in Mandarin Chinese, American beginning learners of Mandarin Chinese all improved significantly in their tonal perception and production of the four Mandarin Chinese tones in monosyllable words. But their study did not address whether the monosyllabic tone training would help learners identify tones in disyllabic words, that is words that are most often encountered in sentences and daily conversations and reflect tonal contexts more accurately. Would learners' tonal perception improve through perceptual training on disyllabic words just as they did with Wang et al.'s training on monosyllabic ones?

Therefore, the current study examined whether perceptual training can effectively be used to train native English-speaking listeners to accurately perceive common and naturalistic (involving tonal coarticulation) disyllabic words. Monosyllabic and disyllabic training will be compared in order to determine the amount of improvement in tone identification. In addition, both monosyllabic and disyllabic stimuli will be examined to determine which type of training material is most effective in helping native English learners to shape tonal categories that do not exist in their phonological inventory.

## 3. Current study

The purpose of the current study is to examine if beginning English-speaking learners' perception of Chinese Mandarin tones in both monosyllabic words and disyllabic words will be improved after perceptual training as learners gain greater proficiency in Mandarin Chinese.

Three research questions are addressed. First, will disyllabic perceptual training be more or less effective compared to monosyllabic perceptual training in helping English-speaking learners shape their tonal categories and improve their tone perception of Mandarin Chinese? Second, when contrasting these two types of training materials, monosyllabic stimuli and disyllabic stimuli, which will be more effective in learning monosyllabic tones and which will be more effective in learning disyllabic tones? And finally, will training using monosyllabic material transfer to disyllabic tone identification, and will training using disyllabic material transfer to monosyllabic tone identification? The goal is to determine which perceptual training (monosyllabic or disyllabic) will help native English learners of Chinese improve their perception of Chinese words (monosyllabic and disyllabic stimuli), and to examine if there is transfer effect between two types of training in learning tones in Mandarin Chinese.

## 4. Method

Three phases were included in the Mandarin tone experiment: a pretest, a training session (either monosyllabic or disyllabic training), and a posttest. All Mandarin Chinese beginning learners participated in identical pretests and posttests, with a forced-choice identification (ID) task used. For the pretest and the posttest, both monosyllabic stimuli and disyllabic stimuli were used. For the training sessions, training (either monosyllabic training or disyllabic training) consisted of four perceptual training sessions. The monosyllabic training group was trained exclusively with monosyllabic stimuli while the disyllabic training group was trained exclusively with disyllabic stimuli. For all training sessions, immediate feedback was given after each response for both monosyllabic and disyllabic training groups.

The two training groups were compared across pretest and posttest to observe any improvement after the training. In addition, the performance for the two types of training material (monosyllabic and disyllabic training stimuli) were examined to determine which type of training material would show the most learning improvement.

## 5. Participants

Native English learners of Mandarin Chinese participated in a two-week training program. All participants were beginning learners of the Chinese language with less than two semesters (less than 7 months) of learning Mandarin. All were college students. Overall, seventeen native English learners of Mandarin Chinese participated. Nine learners participated in the monosyllabic training group, and eight learners participated in the disyllabic training group. Participants were randomly assigned to one of the two training groups. None of these seventeen learners had any history of hearing, speech, or language difficulties.

## 6. Stimuli

All the stimuli were recorded by six native Mandarin Chinese speakers, three males and three females, in order to ensure speaker variability. Two types of stimuli, monosyllabic stimuli and disyllabic stimuli, were used throughout the pretest, training, and posttest. All monosyllabic stimuli were adopted from Wang et al. (1999). These monosyllabic stimuli included all possible permissible combinations of various initial consonants and final vowels, and different syllabic structures in Mandarin Chinese (i.e. V, CV, CVNasal, VN, CGlideV, and CGVN). Each disyllabic stimulus was composed of two randomly combined syllables from the monosyllabic stimuli. Thus, every individual syllable used for the disyllabic stimuli was identical to those used in the monosyllabic stimuli. For example, the monosyllabic stimuli *–mǎ* (*"horse"*) and *– shāng* (*"injury"*) were combined to form a two-syllable word that served as a disyllabic stimulus, *–mǎ shāng*. All monosyllabic stimuli were real words in Mandarin Chinese; the randomly combined disyllabic stimuli were non-words with a decomposable meaning. To preserve the characteristics of disyllable words in connected speech, all six speakers were instructed to produce the stimuli as natural as possible, and to avoid producing any disyllable stimuli as two separate individual syllables. In total, there were 288 monosyllabic stimuli and 144 disyllabic stimuli in the current experiment.

## 7. Procedure

The present experiment consisted of three phases: pretest, training, and posttest. All the tests and training were conducted in the KU Phonetics and Psycholinguistics Laboratory. All stimuli were presented over headphones using Paradigm software (Tagliaferri, 2008) and all learners' responses were recorded in Paradigm. Seventeen native English learners of Chinese participated in the two-week training program. Each learner participated for a total of six days (Pretest; Training1; Training2; Training3; Training4; Posttest), with three sessions the first week and three sessions the second week (see Figure 1). The pretest and posttest were 60 minutes long and each training session was 30 minutes long.

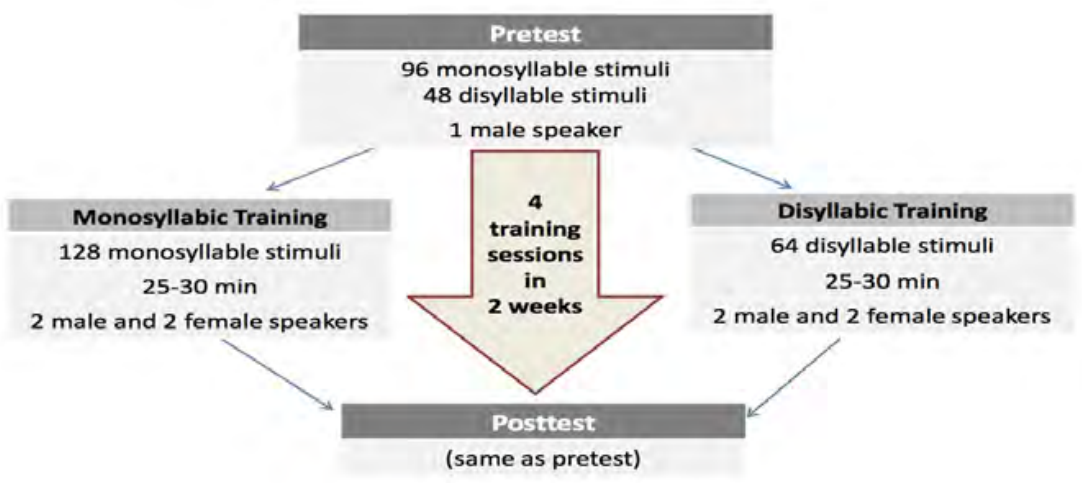


Figure 1. Experimental procedure.

The pretest consisted of two parts, monosyllable word identification and disyllable word identification. All stimuli were produced by a male native Chinese speaker (speaker 1). For both parts, learners indicated which Mandarin Chinese tones they heard. No feedback was provided. The pretest lasted about 60 minutes, approximately 30 minutes for each part.

For the monosyllable pretest, learners heard a monosyllabic stimulus and were instructed to give their tone identification response by pushing the corresponding button that represented one of the four tones (1=Tone1, 2=Tone2, 3=Tone3, and 4=Tone4). All tonal diacritics and numbers were labeled on the buttons on the keyboard. There were 96 monosyllabic stimuli in the pretest. Stimuli in the monosyllable pretest were the same 96 monosyllabic stimuli as in Wang et al. (1999) study. All monosyllabic stimuli were real words in Mandarin Chinese. There were 24 monosyllable words for each of the four phonemic Mandarin tones. All monosyllabic stimuli were presented with a 3 second inter-trial interval (ITI). Learners' accuracy during the identification task were recorded in Paradigm (Tagliaferri, 2008).

For the disyllable pretest, the learners heard a disyllabic stimulus and they were asked to indicate their tone identification response by pushing two corresponding buttons (one response followed by the other response) that represented the tone of the first syllable followed by the tone of the second syllable (1=Tone1, 2=Tone2, 3=Tone3, and 4=Tone4). All tonal diacritics and numbers were labeled on the buttons on the keyboard. There were 48 disyllabic stimuli in the pretest. Each disyllabic stimulus was composed of two randomly combined syllables from the monosyllabic stimuli. Thus, every individual syllable used for the disyllabic stimuli was identical to those used in the monosyllabic stimuli. The randomly combined disyllabic stimuli were non-words with a decomposable meaning. There were 3 disyllable words for each of the 16 (4 tones X 4 tones =16 pairs) combinations. In order to directly compare identification of the disyllable and monosyllable stimuli, accuracy for each syllable of the disyllabic stimuli was tabulated. So if a T1 + T4 was presented and the response was T2 + T4, the first syllable was recorded as incorrect and the second syllable was recorded as correct. Also, due to a productive third tone sandhi rule in Mandarin, for one of the sixteen pairs (Tone3 + Tone3), the first Tone 3 syllable is systematically produced as a Tone 2 when followed by a Tone 3 syllable. For these stimuli, the correct identification was Tone 2 + Tone 3. As with the monosyllabic part, the ITI was 3 seconds, all disyllable tonal diacritics and numbers were labeled on the buttons, and no feedback was given. Learners' accuracy in the identification task were recorded in Paradigm (Tagliaferri, 2008).

## 8. Training sessions

Both Monosyllabic and Disyllabic training consisted of four perceptual training sessions that lasted 30 minutes each. Learners participated in a forced-choice ID task and immediate feedback was given after each response for all training sessions to help learners focus their attention on the critical acoustic cues of the four tones.

### 8.1. Monosyllabic training

The monosyllabic training group was trained exclusively with monosyllabic stimuli. There were 128 monosyllabic training stimuli, which consisted of 32 monosyllable words for each of the four tones. All were produced by four native Chinese speakers (speaker 2, speaker 3, speaker 4, speaker 5), including two male speakers and two female speakers. Overall, there were 512 stimuli in the monosyllabic training produced by the four native Chinese speakers.

For monosyllabic training, participants heard a stimulus, *"má"*, which contained a target tone (e.g., Tone 2) in a monosyllabic word, and they then indicated what they heard among four tones (1=T1, 2=T2, 3=T3, and 4=T4) by pushing the corresponding button on the keyboard. If the choice was correct, the participant would hear: "Correct! That was Tone 2, it is *má*." The next stimulus was then presented. If the response was incorrect, the participant would hear: "Uh-oh! That was má, Tone 2. Let's hear it again *má*". In each of the four training sessions, the trainees were trained with the stimuli produced by only one speaker at a time.

### 8.2. Disyllabic training

The disyllabic training group was trained exclusively with disyllabic stimuli. The monosyllabic training stimuli were used to create the disyllabic stimuli, which shared all the same syllables as those in the monosyllabic training. There were 64 disyllabic training stimuli. The same four native Chinese speakers (speaker 2, speaker 3, speaker 4, and speaker 5) produced these 64 disyllabic stimuli. In each session, the learners heard stimuli only produced by one speaker. Overall, then, there were 256 disyllabic stimuli across the four training sessions.

For disyllabic training, participants heard a disyllabic stimulus, for example, *"mǎ shāng"*, which was a Tone 3 + Tone 1 combination. The learner would then make two responses by pushing two buttons sequentially on the keyboard. The accuracy of each syllable of the disyllable stimulus was counted. Immediate feedback was given just as in the monosyllabic training. For instance, if the choice was correct, the participant would hear: "Correct! That was Tone 3 and Tone 1, it is *mǎ shāng*." The next stimulus was then presented. If either of the two responses was incorrect, the participant would hear: "Uh-oh! That was *mǎ shāng*, Tone 3 and Tone 1. Let's hear it again *mǎ shāng*. " After feedback, the next stimulus was presented.

### 8.3. Posttest

The posttest was identical to the pretest, including both monosyllabic stimuli and disyllabic stimuli. Learners indicated which Mandarin Chinese tones they heard by pushing the corresponding button for the four tones (1=Tone1, 2=Tone2, 3=Tone3, and 4=Tone4) and received no feedback. The posttest lasted about 60 minutes, approximately 30 minutes for each part.

## 9. Results

A binomial logistic regression was conducted to examine the effect of training type on the perception of Chinese tones, using the lme4 package (Bates, 2005; Bates & Maechler, 2010) in the R statistical environment (R development Core Team, 2012, Version 3.4.3). The model had Correct (1=Correct vs. 0=Incorrect) as a dependent variable, and Training Group (Monosyllabic Training vs. Disyllabic Training), Tested Stimuli (Monosyllabic Stimuli vs. Disyllabic Stimuli), and Test (Pretest vs. Posttest) as fixed effects. Subjects and Stimuli were entered as random factors. When there was a significant three-way interaction among the independent variables, we stratified the data by Tested Stimuli to probe the interaction.

The model showed a significant main effect of Test, $c^2(1)= 51.16$, $p<0.001$, indicating that there was a significant improvement after the training from pretest (72%) to posttest (80%), an 8% improvement. There was also a significant main effect of Tested Stimuli, $c^2(1)= 170.45$, $p<0.001$, indicating that the participants identified monosyllabic stimuli more accurately (90%) than the disyllabic stimuli (65%). We also found a significant two-way interaction between Tested Stimuli and Test ($c^2(1)= 9.05$, $p=0.002$), indicating that there was a greater improvement on disyllabic stimuli (9% improvement) than monosyllabic stimuli (7% improvement) after the training. We also found a significant interaction between Test and Training Group ($c^2(1)= 6.38$, $p=0.011$), indicating that the disyllabic training showed a greater improvement (11% improvement) than the monosyllabic training (6% improvement). A statistically significant three-way interaction among Tested Stimuli, Test, and Training Group was also found ($c^2(1)= 6.45$, $p=0.011$).

To further examine this three-way interaction, we stratified the data by Tested Stimuli, and ran two binomial logistic regressions for each stimuli type, including Training Group and Test as main effects and Subject and Stimuli as random effects. The model analyzing the monosyllabic stimuli showed a main effect of Test ($c^2(1)= 44.35$, $p<0.001$) only, indicating that there was a significant improvement on identifying monosyllabic stimuli after the training regardless of the training regime (8% improvement). For the monosyllabic test stimuli, the accuracy of pretest and posttest for the monosyllabic training group was 87% and 94%, and the accuracy of the pretest and posttest for the disyllabic training group was 82% and 90%, respectively. Figure 2 indicates the similar degree of improvement between the monosyllabic training group and disyllabic training group for the monosyllable tested stimuli.


Figure 2. Percent improvement (%) (and standard error) for the monosyllabic stimuli from pretest to posttest by native English-speaking learners of Chinese in the monosyllabic and disyllabic training groups.

The model analyzing the tested disyllabic stimuli also showed a main effect of Test ($c^2(1)=16.70$, $p<0.001$), indicating that there was a significant improvement on identifying disyllabic stimuli after the training (9% improvement) regardless of the training regime. However, for the disyllabic stimuli, we also found a significant two-way interaction between Test and Training Group ($c^2(1)= 11.86$, $p<0.001$), indicating that the disyllabic training group improved more in identifying tones in disyllabic stimuli than the monosyllabic training group did. The accuracy of the pretest and posttest for the monosyllable group was 65% and 70%, while the accuracy of the pretest and posttest for the disyllabic group was

51% and 66%, respectively. Figure 3 indicates the different degree of improvement between the monosyllabic training group and the disyllabic training group for the disyllabic stimuli, with disyllabic training showing greater gains than monosyllabic training.

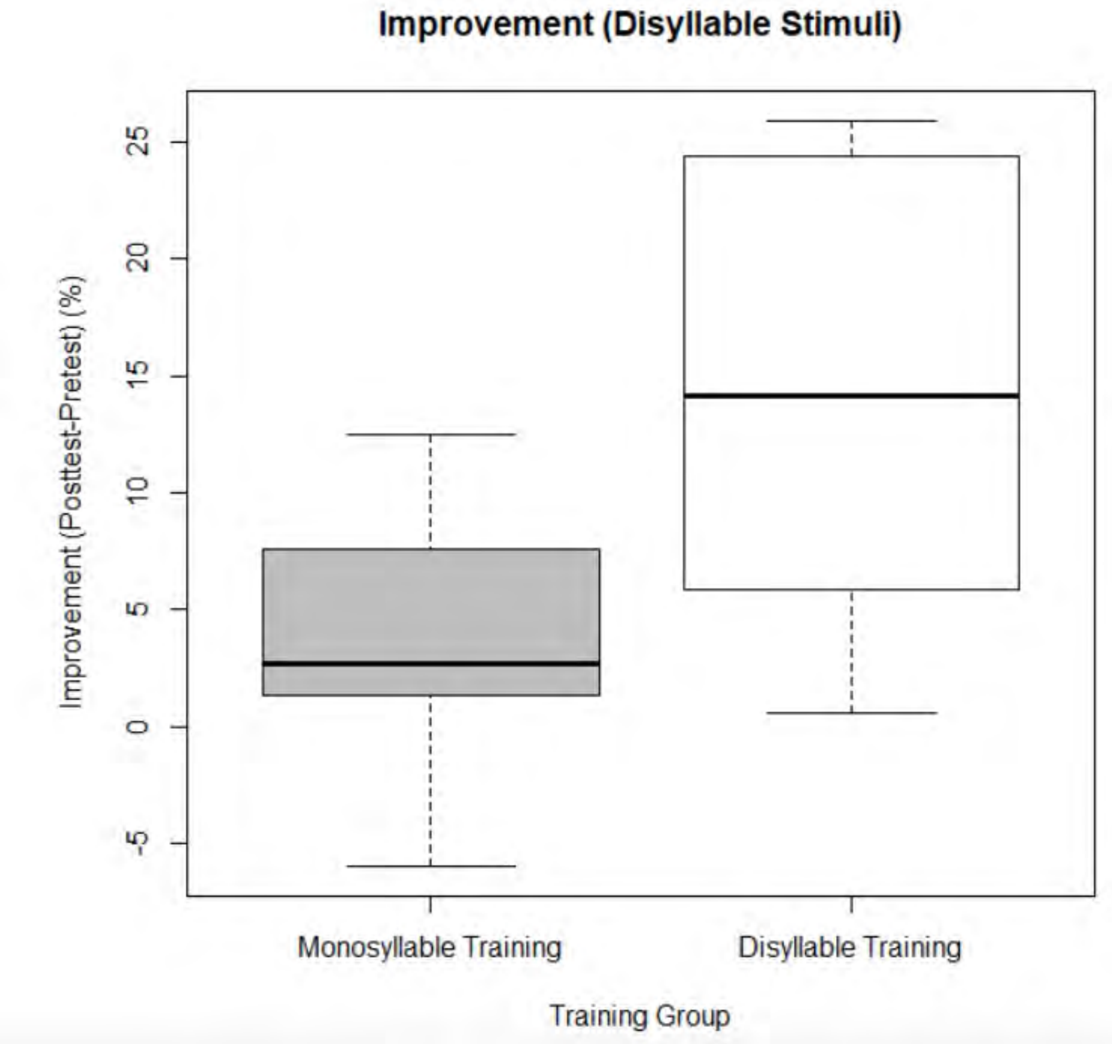


Figure 3. Percent improvement (%) (and standard error) for the disyllabic stimuli from pretest to posttest by native English-speaking learners of Chinese in the monosyllabic and disyllabic training groups.

## 10. Discussion

The results of the current study demonstrated that after high variability perceptual training, adult native English-speaking beginning learners of Chinese were able to significantly improve their tone perception in both monosyllabic and disyllabic stimuli in Mandarin Chinese. Participation in a short (four 30-minute sessions) two-week training showed a significant 8% increase ($p<0.001$) from pretest 72% to posttest 80% in learners' overall tone perception accuracy. These data are similar to Wang et al. (1999) in which examined monosyllabic perceptual training, showing a sizable 21% increase. More substantial learning in their study was most likely due to the fact that more training sessions were used (8 sessions of 40 minutes each) and also due to the fact that training stimuli for Wang et al. (1999) were arranged pairwise, which allowed for a systematic increase in difficulty of tone contrasts as learning progressed. Interestingly, while Wang et al. (1999) only used monosyllabic stimuli with monosyllabic training, the current study showed that while identification of tones in disyllabic stimuli is more challenging, there was greater improvement on disyllabic stimuli (9% improvement) than monosyllabic stimuli (7% improvement) after training. These data suggest that inclusion of more complex and variable disyllabic stimuli will not harm the beneficial aspects of high variability training.

It should be noted that learners generally did significantly better ($p<0.001$) when identifying tones in monosyllabic stimuli, with an accuracy of 90%, as compared to disyllabic stimuli, with an accuracy of 65%. Such a substantial identification accuracy gap between the two types of stimuli was also observed by Sun (1998) and He (2010). Recall that in the current study, identification of the monosyllabic stimuli in pretest and posttest is based on an isolated syllable while, for the disyllabic stimuli, listeners heard a sequence of two syllables which they were asked to identify. Differences in overall monosyllabic and disyllabic identification accuracy are likely due to the tonal environment, with tones in monosyllabic stimuli occurring in isolated environments, such that these tones are preserved in their canonical forms, while tones in disyllabic stimuli were often coarticulated with the adjacent tones' pitch (Shen, 1990; Xu, 1994, 1997, 1998) or they undergo contextually-driven phonological processes (e.g. third tone sandhi). Despite overall accuracy differences between monosyllabic and disyllabic stimuli and the challenges of disyllabic tone identification, the current results show that there was greater improvement on disyllabic stimuli (9% improvement) than monosyllabic stimuli (7% improvement) after training. Given these data showing successful improvement using disyllabic stimuli, teachers, when teaching Mandarin Chinese tones, should not shy away from providing students with disyllabic stimuli that contain more contextual variability.

More importantly, the current findings also showed significant differences between the monosyllabic perceptual training group and the disyllabic perceptual training group from pretest to posttest. Critically, these differences due to training were observed regardless of the syllabic structure of the stimuli tested. When identifying tones in both types of stimuli (monosyllabic and disyllabic), the learners in the monosyllabic training group showed a significant 6% increase from pretest 76% to posttest 82% ($p<0.001$). Similarly, learners in the disyllabic training group also showed a significant improvement from the pretest 67% to the posttest 78%, with an 11% increase ($p<0.001$). While both monosyllabic and disyllabic perceptual training was beneficial for learners to aid in building robust tonal categories in Mandarin Chinese, those learners who had disyllabic training made nearly double the improvement (11%) on their tonal identification compared to the monosyllabic training group (6%). The disyllabic training for native English-speaking learners seemed to provide more effective learning of Mandarin Chinese tones in both monosyllabic and disyllabic stimuli than did the monosyllabic training.

Interestingly, transfer effects of training were also found in current study. Learners who received the monosyllabic training improved significantly when perceiving tones not only in monosyllabic stimuli (from pretest 87% to posttest 94%), but also in disyllabic stimuli (from pretest 65% to posttest 70%) ($p<0.001$). Moreover, learners who received the disyllabic training not only showed substantial improved tone identification when identifying tones in disyllabic stimuli (from pretest 51% to posttest 66%), but also in monosyllabic stimuli (from pretest 82% to posttest 90%) ($p<0.001$). These results show that both training regimes seem to improve tonal perception, with either monosyllabic training or disyllabic training being beneficial for learners to identify Mandarin Chinese tones in monosyllabic stimuli and disyllabic stimuli. But importantly, while listeners in the monosyllabic perceptual training group exhibited similar improvement for both monosyllabic and disyllabic test stimuli (7% and 5%, respectively), listeners in the disyllabic training group showed more improvement, as expected, in the disyllable test stimuli (15%), but also showed substantial improvement in the monosyllabic stimuli (8%). Thus, when teaching the language, it may be helpful for instructors to introduce tones in disyllable words since this exposure provides learners with more typical real-world contexts exhibiting more tonal variability, and, crucially, this encourages learners to develop more robust tonal categories.

## 11. Conclusion

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

**References**

Bates, D. (2005). Fitting linear mixed models in R. *R News*, *5*(1), 27-30, http://CRAN.R-project.org/doc/Rnews/

Bates, D., & Maechler, M. (2010). Matrix: sparse and dense matrix classes and methods. R package version 0.999375-43, http://cran. r-project. org/package= Matrix.

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

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

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

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

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

Gandour, J. (*1983*). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*(2), 149-175.

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

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

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

He, Y. (2010). Perception and production of isolated and coarticulated Mandarin Tones by American Learners. University of Florida, Gainesville, Ph.D. Dissertation.

He, Y., & Wayland, R. (2010). The production of Mandarin coarticulated tones by inexperienced and experienced English speakers of Mandarin. In *Speech Prosody 2010-Fifth International Conference*, 123:1.4.

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

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

Jenkins, J. (2004). Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics*, *24*, 109-125.

Jongman, A., Wang, Y., Moore, C. and Sereno, J. A. (2006). Perception and production of Mandarin tone. In E. Bates, L. Tan, and O. Tzeng (Eds.), *Handbook of East Asian Psycholinguistics*, Cambridge: Cambridge University Press, 209-217.

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

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

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

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*(3), 1242-1255.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*(2), 874-886.

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

Orton, J. (2013). Developing Chinese oral skills-a research base for practice. *Research in Chinese as a Second Language*, *9*, 3-26.

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

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

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

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

Wang, Y., Spence, M., Jongman, A., and Sereno, J.A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*, 3649-3658.

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

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