PERCEPTION AND PRODUCTION OF ISOLATED AND COARTICULATED MANDARIN TONES BY AMERICAN LEARNERS

By

YUNJUAN HE

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2010

To my grandma

ACKNOWLEDGMENTS

First of all, I would like to thank the chair of my committee, Dr. Ratree Wayland, who is very knowledgeable at doing research on tones. It was her to attract me to apply for PhD program at the University of Florida. During my PhD study, she provided lots of valuable advice. Second, I would like to thank Dr. Caroline Wiltshire. From taking a course on research method in linguistics with her, I had the opportunity to start thinking about the topic of my dissertation and finished the first draft of the research proposal. Third, I would like to thank Dr. Andrea Pham and Dr. Zhihui Fang, for accepting to be part of my committee and providing me with their insightful comments. Fourth, I would like to thank Dr. Cynthia Shen, Ms. Elinore Fresh and Ms. I-Chun Peir from the department of department of languages, literatures & Cultures, who helped me to recruit American students to participate in my study. I would like to thank my classmates Ye Han, Weihua Zhu, Lili Gai for their academic or personal help when I was in Gainesville and Tampa. Last but the least, I would like to thank my grandma Rongjie Liu, my father Yedong He, my mother Weiping Yu and my husband Feng Li for being supportive and always believe in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

PERCEPTION AND PRODUCTION OF ISOLATED AND COARTICULATED
MANDARIN TONES BY AMERICAN LEARNERS

By

Yunjuan He

May 2010

Chair: Ratree Wayland
Major: Linguistics

This study tested the effects of language learning experience, tonal condition, tonal sequence, tonal context and syllable position on American English speakers' ability to perceive and produce Mandarin Chinese coarticulated tones in disyllabic words. It showed that experienced learners with more language learning experience outperformed inexperienced ones on perception and production of coarticulated tones. Tonal condition, that is, whether a tone in isolation or in coarticulation, affected the perception and production of all tones. Tones in isolation were identified and produced with higher accuracy rate than tones in coarticulation. Tonal sequence, that is, whether the two syllables of the word had the same tone or two different tones, was found to affect the perception and production of Tone 1. Tone 1 in identical tonal sequence was identified and produced with higher accuracy rate than it was in nonidentical tonal sequence. Tonal context, that is, whether the pitch of the two tones at their intersection was compatible or conflicting, affected the perception and production of Tone 1. Tone 1 in compatible context was perceived and produced better than it was in conflicting tonal context. In terms of the effect of syllable position, it was found that syllable position

affected Tone 2 and Tone 4 perception and production accuracy and syllable position affected Tone 3 perception accuracy but not production accuracy. The perception of these tones was better in final syllable position. The production of Tone 2 was better in final syllable position, whereas the production of Tone 4 was better in initial syllable position.

For the perception of coarticulated tones, tonal direction errors decreased with increased learning experience, but tonal height errors remained. However, for the production of coarticulated tones, the ability to produce tones in terms of both tonal direction and tonal height quickly improved with increased learning experience.

The relationship of perception and production of isolated tones for American learners was still not clear. Perception and production were not correlated for either group of learners. Perception of coarticulated tones predicted production for inexperienced but not experienced learners. This finding was inconsistent with the hypothesis that accurate perception precedes accurate production.

CHAPTER 1
INTRODUCTION

In the United States, Mandarin Chinese is very commonly taught as a foreign language at the college level. Because Mandarin is a tonal language, perceiving and producing tones accurately is very important for American speakers in their acquisition of Mandarin Chinese.

When learning Chinese as a foreign language, students get input of the target language mainly from their teachers in the classroom. In the current Chinese language classroom, teachers focus on introducing the tonal direction of Mandarin tones in isolation and provide practice on perceiving and producing tones in isolation on monosyllabic words (Xing, 2006). However, according to one statistical analysis (Wang et al, 1986) among the 31,159 Mandarin words surveyed, 22,941 (74%) were disyllabic words, and only 12% were monosyllabic words. The remaining 14% of words contain more than two syllables. Therefore, perceiving and producing tones on monosyllabic words alone is not sufficient for L2 Mandarin learners. Practicing tone use on disyllabic words is necessary for those who actually want to communicate successfully with native speakers.

The tonal direction and height of the four Mandarin lexical tones is quite stable when they occur in isolation, whereas these acoustic features may be changed in coarticulated tonal environments (in the current study, the coarticulated tonal environment is the disyllabic word) (Xu 1994, 1997). Therefore, the gap between the tonal practice in classroom and the use of tones in the real world poses a great problem for learners. American learners may not be able to perceive and produce coarticulated

tones on disyllabic words accurately even after they have acquired Mandarin tones in isolation.

The effect of language learning experience has been examined on adults' perception and production of L2 sounds. The main research question asked in these studies is whether the performance of experienced speakers of the target language is better than inexperienced speakers on the acquisition of L2 sounds.

Most previous research on segmental features has suggested that the ability to perceive and produce L2 sounds may improve with increased learning experience (Mackain et al.1981, Flege et al. 1995; Flege et al.,1997; He, 2004; Levy and Strange, 2008). In terms of suprasegmental features, experienced learners also outperform inexperienced learners on the acquisition of stress in English and tones in Thai and Mandarin (He et. al., 2008; Wayland and Guion, 2003; Wang et al., 1999, 2003).

In Wang et al., (1999, 2003)'s studies perception and production of isolated Mandarin tones by American speakers, experienced learners who received perception training in a language lab identified and produced tones more accurately than the inexperienced learners who did not receive the training. However, no studies have been undertaken to systematically describe how American speakers with different amounts of classroom learning experience perceive and produce Mandarin tones in coarticulation.

Research Questions

This study is aimed at finding how different amount of language learning experience affects L1 American English speakers' perception and production of L2 Mandarin tones. This study is guided by several research questions:

Research Question 1. Do experienced American English speakers perceive and produce Mandarin tones more accurately than inexperienced learners?

Research Question 2. What are the error types of perception and production of tones typical to American speakers with different amount of language learning experiences?

Research Question 3. What are the linguistic factors that prevent American speakers from perceiving and producing Mandarin tones accurately?

Research Question 4. What is the relationship between perception and production of Mandarin tones by American speakers?

Research Design

To answer the four research questions, four experiments were designed: two perception and two production. First, one perception and one production experiment of isolated Mandarin tones were completed by experienced American learners and inexperienced American learners in order to explore whether language learning experience affects the accuracy of identifying and pronouncing isolated Mandarin tones. Furthermore, the production study was designed to investigate the acoustic cues that American English speakers of L2 Mandarin tend to use to modify Mandarin isolated tones. In these first two experiments, 22 American students who had studied Chinese for three months participated as inexperienced learners, and 16 American students who had studied Chinese for fifteen months participated as experienced learners.

More importantly, this study extended the testing stimuli to longer utterances, disyllabic words, which more frequently appear in the word level of real conversation. One perception and one production experiment were conducted to assess the effects of language learning experience, syllable position, and tonal environment on American speakers' accuracy of perception and production of coarticulated Mandarin tones.

Production and perception data from 14 inexperienced learners and 9 experienced learners who were able to accurately perceive tones in isolation were included in the analyses.

Significance of the Study

This study is significant for several reasons. First, it is the first study to examine the effect of language learning experience on the perception and production of coarticulated Mandarin tones by American learners; earlier studies investigated only learners' mastery of isolated, single-syllable tones. Second, it is the first study to test the effect of tonal environment on the acquisition of Mandarin coarticulated tones on disyllabic words by L2 learners. Third, it is the first study to investigate the nature of the relationship between perception and production of Mandarin coarticulated tones by American learners. Furthering our knowledge of the acquisition of Mandarin tone by American learners will lead to a better understanding of their difficulties in perceiving and producing Mandarin coarticulated tones and help us to improve Chinese teaching pedagogy.

CHAPTER 2
LITERATURE REVIEW

In this chapter, I first introduce the features of tone in Mandarin, which are basic knowledge for understanding the present study and the methodology presented in Chapter 3, Chapter 4 and Chapter 5. I then describe the features of stress and intonation in English, two features which may be transferred to L2 Mandarin tones by American learners. Then, I present a review and critique of currently available literature relevant to my main subject: perception and production of tones by L2 speakers. Finally, I review some of the existing literature on the relationship between perception and production of L2 features.

Suprasegmental Features in Mandarin and English

Mandarin

Tone in isolation

Mandarin Chinese is a tonal language. It has four lexical tones, the main cue to which is fundamental frequency. They are

> high level tone (Tone 1), a tone [that] starts with a high F_0 value and stays around that level throughout the syllable; rising tone (Tone 2), a tone [that] starts with a low F_0, then falls slightly before rising throughout the remainder of the syllable; low rising tone (Tone 3), a tone [that] starts with an F_0 value slightly lower than the onset of Tone 2, falls to the lowest F_0 , then rises sharply to the end of the syllable; [and] falling tone (Tone 4), a tone [that] starts with the highest F_0 value of the four tones, continues to rise before reaching the maximum about one fifth of the way into the vowel, then falls sharply to the end of the syllable (Xu, 1997; p.67).

Traditionally, Mandarin tones are described by the Scale of Five Pitch Levels (Chao, 1948). This measurement is widely used in phonetic and phonological studies on Mandarin tones. Table 2-1 summarizes the description of the four tones. (5 indicates the highest possible pitch.)

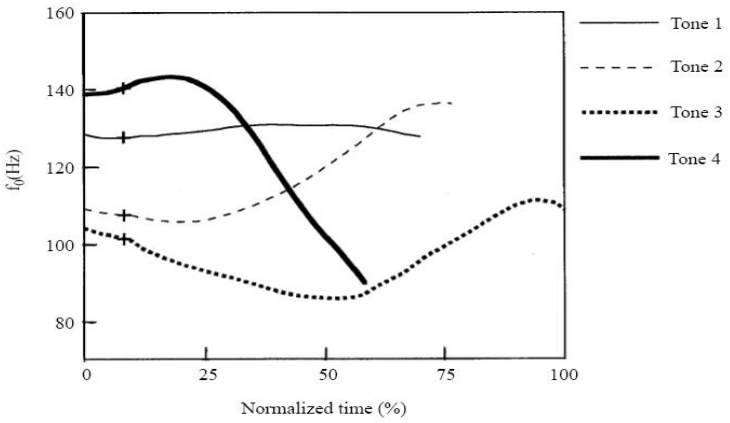


Figure 2-1. Mean F_0 contours of four Mandarin tones in the monosyllable /ma/ produced in isolation. The time is normalized, with all tones plotted with their average duration proportional to the average duration of Tone 3. (Xu, 1997; p. 67)

Table 2-1. Lexical tones in Mandarin

	Tone	Descriptive name	Pitch value	Diacritic	Pitch targets by Xu (1993)
Ma ‘mother’	Tone 1	High level	55	ˉ	high-high
Ma ‘hemp’	Tone 2	Rising	35	ˊ	low-high
Ma ‘horse’	Tone 3	Low rising	214	ˇ	low-low
Ma ‘scold’	Tone 4	Falling	52	ˋ	high-low

The fundamental frequency is the primary acoustic cue for Mandarin speakers to identify Mandarin tones ((Liu, 1924; Howie, 1976; Wu, 1986). However, the amplitude of tones also signals the category of Mandarin tones, except, Tone 1, the high flat level tone (Whalen and Xu, 1992). Tone 2, Tone 3 and Tone 4 can be identified from just the amplitude contour. Duration is also a cue to tone identity. Tone 3 has the longest duration, Tone 4 has the shortest duration, and Tone 1 and Tone 2 are in between. In addition, duration affects the perception of Tone 2 and Tone 3 (Blicher et al,1990). Specifically, it was found that Tone 2 and Tone 3 were identified more often as Tone 3 when the tones were lengthened. Besides the overall duration of the tone, Moore and Jongman (1997) found that the turning point (the point where falling pitch changes into

rising pitch) and ΔF0 (the change in pitch from onset of the tone to the turning point) are also acoustic cues to differentiate Tone 2 and Tone 3.

Tone in coarticulation

When produced in isolation, the pitch contour or the F0 value of the tone is quite stable. However, in connected speech, the basic shapes of Mandarin tones often undergo context-induced modifications (Shih, 1986; Shen, 1990; Xu, 1993, 1994). In disyllabic words, it was found by Shen (1990) and Xu, (1994 a) that the (tone value) F0 value/contour of the preceding syllable has a carry-over influence on the F0 value/contour of the following tone Furthermore, the F0 value of the following syllable also has an anticipatory influence on the F0 contour of the preceding tone Shen (1990) claimed that the bidirectional effects are symmetrical: the influences of adjacent tones on each other are similar. Also, she claimed that tonal coarticulation changes the range but not direction of F0. In other words, the pitch value of tones will be changed within their own tone category.

However, by conducting a more carefully controlled experiment, Xu (1997) found that the bidirectional effects are asymmetrical. In terms of magnitude, the carry-over effect is larger than the anticipatory effect. The carry-over effect causes the onset of the following tone to assimilate to the offset of the preceding tone: a high offset of the preceding tone raises the onset of the following tone, and a low offset lowers the onset of the following tone. On the other hand, anticipatory effect causes the offset of the preceding tone to dissimilate to the onset of the following tone: a high onset of the following tone lowers the offset of the preceding tone, and a low onset raises the offset of the preceding tone.

Besides the influence of different tonal contexts, Xu (1993) identified two additional types of contextual effects that may also affect pitch shape of a tone. One is a compatible context: the value of the pitch of the offset of the preceding tone and the value of the pitch of the onset of the following tone are similar. For example, in the phrase dàn bái zhì, (protein) the value of the offset of the preceding tone is similar to the value of the onset of the following tone. The other is a conflicting context: the value of the offset of the preceding tone and the value of the onset of the following tone are substantially different. For example, in the phrase cōng yóu bǐn (a kind of pie) the value of the offset of the preceding tone is different from the value of the onset of the following tone.

Xu (1993) investigated the perception and production of the second syllable associated either with a rising tone or a falling tone in trisyllabic words to see how different tonal contexts affect the tonal shape of the target tones. In terms of production, it was found that the slope of rising tone keeps rising in the compatible context, while the slope falls in the conflicting context. Therefore, context can actually cause the slope of a tone to change to the opposite of what is heard in isolation. The slope of a falling tone is steeper in the compatible context than in the conflicting context. Meanwhile, the perception accuracy of the category of the tone on the second syllable is higher in the compatible context than in the conflicting context.

Finally, Cheng (1986) introduces the ‘third-tone sandhi rule: If two consecutive syllables both have Tone 3, the first Tone 3 is changed to Tone 2.

English

Stress

A stressed syllable in American English is characterized by four phonetic properties: higher pitch, longer duration, higher intensity and full vowel quality (Gussenhoven, 2004). Shown in Figure 2-2 (Gussenhoven, 2004), the pitch characteristics of stress in disyllabic English words produced in citation form are 1) If the first syllable is stressed, the pitch level starts high and falls on the second syllable. The fall is somewhat truncated. 2) If the second syllable is stressed, the pitch level starts low, rises on the approach to the second syllable, and falls on the second syllable.

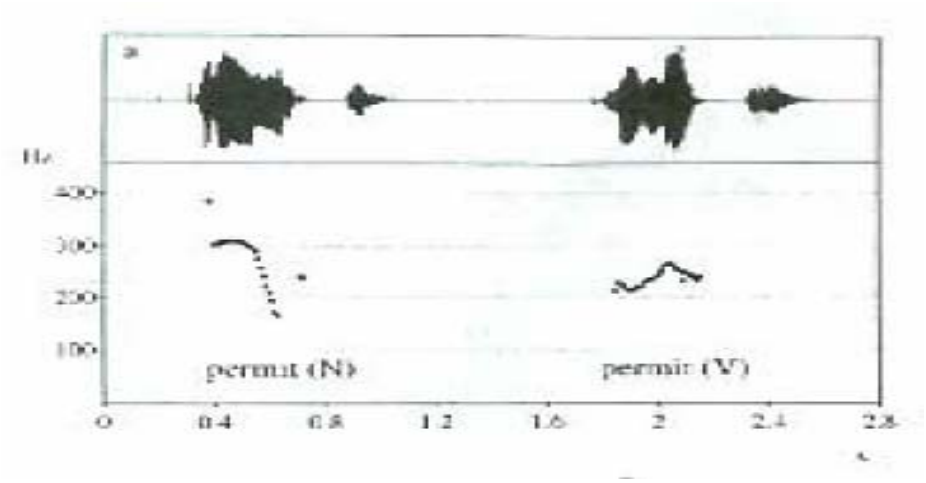


Figure 2-2. Waveforms and F0 tracks of citation pronunciations of *permit* (Noun) and *permit* (Verb), cited from Gussenhoven, 2004.

The rules of assigning stress (Cruttenden, 2000; p16) are:

- (i) Verbs and adjectives
- (a) Stress on the penultimate syllable when the final syllable has a short vowel in an open syllable or is followed by no more than one consonant
- (b) Otherwise, stress is on the final syllable (subject to rule (iii) below)
- (ii) Nouns

- (a) if the final syllable has a short vowel, disregard it and apply rules under (i) above
- (b) if the final syllable has a long vowel, it is stressed (subject to (iii) below)
- (iii) In words of more than two syllables with a long final vowel
- (a) Stress may optionally occur on the antepenultimate rather than the final syllable.

Intonation

Though English is not a tonal language, it does use intonation to express certain expressions. Unlike Mandarin tones, English intonation can be expressed over several words. In one sentence, a nuclear tone 'involves the major part of the meaning contributed by the pitch pattern of an intonation-group' (Cruttenden, 2000; p.50). According to the description of Cruttenden (2000), English has seven types of nuclear tones, which Cruttendon divides into two groups: falling tones and rising tones. Falling tones are more common in sentence final position than in sentence non-final position. They are 1) High falling: to show more interest, more excitement, or more involvement. 2) Low falling: to show less interest, less excitement and less emotion. 3) rising-falling: to mark the completeness of a declaration. Within certain environments, falling nuclear tones can express the feeling of being impressed, the intention to gossip, or the intention to challenge.

Rising tones often occur in sentence non-final position. They are: low rising, high rising, fall-rising and mid level. There are two subgroups of rising tones. One group contains dependent rising tones, whose meaning depends on the relationship with another tone. The other group contains independent rising tones, which have their own meaning regardless of surrounding tones. Dependent rising tones are 1) Low rising: the most oratorical, typical of a formal reading style. 2) High rising: more casual. 3) Fall-

rising: to emphasize the contrastive nature of the subject. 4) Mid level: no other meaning except of non-finality. Independent rising tones are 1) Low rising: to express the feeling of uncertainty. 2) High rising: to soothe, reassure or patronize. 3) Fall-rising: to express self-justification or to appeal.

The common combinations of the nuclear tones in a sentence are 1) Rising tone + falling tone. 2) Falling tone + low rising tone, which is used in tag questions. The listener will get the sense of disagreement. 3) Falling tone + falling tone, which is also used in tag questions. The listener will get the sense of being required to have agreement.

Effect of Learning Experience on the Acquisition of L2 Sounds

As indicated in the literature reviewed above, little research has been conducted on the effects of learning experience on the perception and/or production of suprasegmental features, such as stress, tone and intonation. Therefore, I broaden the scope of my literature review to include the effects of learning experience on the acquisition of L2 sound features, including both segmental and suprasegmental features. In the following paragraphs, I review studies of the effect of learning experience on the acquisition L2 sound features first, beginning with perception and then continuing with production.

The Effect of Learning Experience on the Perception of L2 Sound Features

Flege (1997) studied the effects of experience on the perception of English vowels (/i/, /ɪ/, /e/ /æ/) by non-native speakers with various language backgrounds (German, Spanish, Mandarin, and Korean). It was found that the experienced speakers, with an average of 7.3 years length of residence (LOR) in the U.S., made significantly more use of spectral cues which changes the quality of vowels than did the inexperienced

speakers, who had an average of 0.7 years' LOR, when identifying /e/-/æ/. Experienced Mandarin and Korean speakers, like native English speakers, made significantly more use of spectral cues and fewer temporal cues (such as duration) than did the inexperienced Mandarin and Korean speakers when identifying /i/-/ɪ/.

Levy and Strange (2008) conducted a study on the perception of French vowels by L1 American English adults with and without French language experience. The results showed that the experienced group of American English adults discriminated /u-œ/, /i-y/ and /y-œ/ better than the inexperienced group did. However, there was no significant difference in the two groups' ability to discriminate /u/ and /y/. Consonantal context affected the inexperienced group's ability to discriminate these vowels, but not the experienced group. The inexperienced group performed better in the bilabial context than in the alveolar context, whereas the experienced group's performance in these two consonant contexts was comparable.

In Flege's (1984) study, experienced L1 Arabic L2 English speakers who had lived in the U.S. for an average of 5.8 years and inexperienced L1 Arabic L2 English speakers who had lived in the U.S. for an average of two months were asked to identify /s/ or /z/ as "piece" or "peas": The duration of the vowels and fricatives was manipulated, and it was found that the experienced L2 English speakers' performance more closely matched native English speakers' in all three conditions: only vowel duration increased, only fricative duration decreased and vowel duration increased + fricative duration decreased. The inexperienced L2 English speakers' behavior resembled that of the native English speakers in only two conditions. In the condition of

only fricative duration decreased, their performance is dramatically different from the English native speakers.

Only one study on discriminating suprasegmental features has been conducted: Wayland and Guion in 2003 tested naïve and experienced learners of Thai's ability to discriminate middle and low tones in Thai. This pair of Thai tones is quite confusing to American speakers. However, experienced American learners in the study performed better than American speakers without any learning experience on discriminating middle tone and low tone.

However, not all studies have found an advantage for experienced learners. In Swedish, a long vowel must combine with a short consonant in a stressed syllable, and a short vowel must be with a long consonant. McAllister et al. (2002) asked experienced and inexperienced groups of L1 English and L1 Spanish speaking adults to identify whether perceived words followed this phonological rule. Experienced and inexperienced groups of listeners did not differ significantly in their perception accuracy on this task.

The Effect of Learning Experience on the Production of L2 Sound Features

In addition to investigating the effects of language experience on perception of English vowels (/i/, /ɪ/, /e/ /æ/) by non-native speakers with various language backgrounds (German, Spanish, Mandarin, and Korean), Flege (1997) also studied the effects of language experience on the production of these vowels. Based on an intelligibility test, the production of /ɪ/ by experienced non-native subjects (average 7.3 years residency in the U.S.) was more native like than that of the inexperienced non-native subjects (average 0.7 years residency in the U.S.). Acoustic analysis revealed

that experienced L1 German speakers and experienced L1 Mandarin speakers produced a larger spectral difference between /i/-/ɪ/ and /e/-/æ/ than inexperienced speakers of the two language groups, and experienced L1 German and L1 Spanish speakers produced significantly different durations for /e/ and /æ /, which inexperienced speakers did not make.

Flege (1995) found that experienced L1 Japanese speakers of L2 English (average 21 years of residence in the U.S.) performed better than inexperienced speakers (average 2 years of residence in the U.S.) when producing English /r/ and /l/. The production of the English liquids by experienced Japanese speakers was near the native norm based on the judgment of native speakers of English, whereas inexperienced Japanese speakers frequently produced /r/ as /l/ and sometimes produced /l/ as /r/.

He (2004) tested two groups of L1 Mandarin speakers' production of L2 English /l/ in syllable-final position. One group included inexperienced L2 English speakers who were college students in China, and the other included experienced speakers of English who had studied in Canada for at least one year. It was found that experienced speakers produced more native-like syllable-final /l/s than inexperienced speakers of English.

Trofimovich and Baker (2006) investigated the production of five suprasegmental features (stress timing, peak alignment, speech rate, pause frequency, and pause duration) of six English declarative sentences by L1 Korean learners of L2 English. 30 adults with varying levels of English experience (3 months, 3 years, and 10 years of United States residence, respectively) participated in the study. The Korean learners

with extensive English language experience performed better on producing stress timing than learners with medium level of English language experience, and learners with medium level of English language experience performed better than learners with short language experience. However, the effect of language experience was not detected in the production of the other four suprasegmental features. Instead, the adult learners' age at the time of first extensive exposure to the L2 (indexed as age of arrival in the United States) influenced the production of speech rate, pause frequency, and pause duration.

In contrast to the studies discussed above, McCallister et al. (2002) found no effect of experience on production of length contrasts. Experienced (average 10 years of residence in Sweden) and inexperienced (average 3.6 years of residence in Sweden) groups of L1 English and L1 Spanish speaking adults completed a production experiment of Swedish consonant and vowel length distinctions, and the results showed that the ratios of durations of vowels and consonants were not significantly different between experienced and inexperienced groups.

He et al. (2008) investigated how native speakers of Mandarin, a tonal language, employ phonetic cues to differentiate stressed and unstressed syllables in producing English disyllabic words. Two groups of L1 Mandarin speakers with different L2 English learning experiences participated in the study: 8 inexperienced learners of English who were college students in China, and 8 experienced learners of English who had studied at a university in the U.S. for at least two years. It was found that, compared to inexperienced learners, experienced learners produced more native-like acoustic characteristics (larger ratios of amplitude and duration in stressed vs. unstressed

syllables) to stress English words. However, the ratios were still smaller than English native norms. Both groups performed similarly to the native norm on the ratio of pitch between stressed and unstressed syllables.

Conclusion

Generally speaking, previous studies show that experienced learners perform better than inexperienced learners on perceiving and producing L2 features, not only in terms of overall accuracy rate, but also in terms of which spectral cues are used in perceiving and/ or producing these L2 features. Experienced learners generally perceive more accurately than inexperienced learners. They also produce the target sounds in a more native-like fashion than inexperienced learners. However, with certain features, such as vowel and consonant length distinction in Swedish produced by English speakers and English lexical stress produced by Mandarin speakers, more learning experience does not seem to lead to improved performance. Since no study has been done to investigate the effect of learning experience on perception and production of Mandarin tones by L2 learners, this study was aimed to fill the research gap.

Perception and Production of Mandarin Tones by L2 speakers

Perception of Isolated Tones by L2 Listeners

To understand how L2 listeners identify and discriminate tones in isolation, several factors that might affect the perception of tones by L2 listeners have been studied.

Linguistic experience. It has been found that native listeners can perceive tones better than non native listeners (Gottfried et al, 1997; Lee et al, 1996; Wang, 2004, Wayland & Guion, 2003; Wayland & Guion, 2004). Cantonese listeners performed

better at differentiating Cantonese tones than Mandarin listeners and English listeners (Lee et al, 1996); Mandarin listeners performed better at identifying Mandarin tones than Cantonese listeners and English listeners (Gottfried et al., 1997, Lee et al., 1996); and Thai listeners did better at differentiating the middle and low tone contrasts in Thai than Chinese and English listeners (Wayland & Guion, 2003; Wayland & Guion, 2004). The advantage of the native listeners in perceiving the tones in their language accurately is quite straightforward. Since they are acquiring the tones from the time they are born, they are very familiar with the acoustic cues of each tone in the tone inventory of the language. This advantage allows native listeners to identify tones in a shorter time than non native speakers (Lee et al, 1996), and it also helps native listeners identify tones more accurately than non native speakers when the tones are presented with an entire syllable, center-only syllable, initial-only syllable and silent-center syllable (Gottfried et al, 1997).

Not only do non native listeners have problems with identification and differentiation of tones compared to native listeners, but they also lack categorical perception of tones. In Halle et al.'s (2003) study, Mandarin listeners perceived tones categorically, whereas French listeners perceived tones psychophysically and their results showed more variation. Furthermore, Mandarin listeners performed better at identifying and discriminating Mandarin tones than French listeners in terms of accuracy. Although the French listeners didn't show a categorical perception of Mandarin tones, they did show sensitivity to changes of pitch when the acoustic difference was salient enough. It was also found that native speakers of tonal languages have a right-ear (left-hemisphere) advantage for perception (Wang et al, 2001; Wang et

al, 2004). This results has been tested on speakers of Thai, Mandarin and Norwegian, bilingual speakers whose one of the languages is the tested language also showed left hemisphere dominance. However, the nonnative speakers without L1 tonal background, such as Norwegian and English speakers, do not show the right-ear advantage (Wang et al, 2004). Therefore, it was concluded that nonnative speakers do not have "hemispheric lateralization to native like left-hemisphere dominance". (Wang, 2004, p. 465).

It is not clear whether a listener with a tonal background will always outperform a listener without any tonal language experience when perceiving foreign tones. Some studies have supported this idea: Cantonese listeners discriminated Mandarin tones better than English listeners (Lee et al, 1996), and Wayland & Guion (2004) found that Chinese listeners were better at discriminating and identifying low and mid tones in Thai than American listeners without a background in tone languages. Wayland and Guion attributed the result to the Chinese listeners' perceptual transfer from their L1 tonal background to perceive L2 tones. The ability to track pitch direction and contour, which Chinese listeners use to perceive Chinese tones, is transferred to the discrimination and identification of the pair of Thai tones. It seems that tonal listeners use similar strategies to identify and discriminate tones which non tonal listeners do not use. Gandour (1983) claimed that listeners of tonal languages (in this case, Cantonese, Mandarin, Taiwanese, and Thai) paid more attention to the direction of tones, whereas listeners of a nontonal language (English) attached more importance to the height of the tone. Here, the ability to track the contour of the tone, which English listeners lack, seems to play a

more important role in identifying and discriminating tones than the ability to hear the tone's pitch level.

However, other studies have provided contradictory evidence: Mandarin listeners did not obtain higher tonal discrimination scores than English listeners on the perception of Cantonese tones (Lee et al, 1996). The authors suggest that Cantonese tones are more difficult than Mandarin tones for children to acquire. Differences in first language development between the two languages indicate that Cantonese tones are harder to perceive than Mandarin tones. Therefore, Cantonese listeners can discriminate Mandarin tones more easily than Mandarin listeners can discriminate Cantonese tones.

Francis et al. (2008) suggested that language background (tonal vs. nontonal) may not affect the error rate of identifying tones, but only the types of errors made when identifying tones. In their study, Mandarin speakers (tonal language speakers) and English speakers (non-tonal speakers) were trained to identify Cantonese lexical tones. Both groups performed better on recognizing Cantonese lexical tones after their training. However, there was no significant difference in the degree of improvement between the two groups. In terms of error type, Mandarin speakers more often confuse tones with similar tonal directions, whereas English speakers more often confuse tones with similar average $F_{0.}$ For example, Mandarin speakers tended to misidentify low rising (23) tone as high rising (25) tone, whereas English speakers tended to mistake it for mid level (33) tone.

Language learning experience. Experienced L2 learners whose L1 is not a tonal language may have fewer perceptual difficulties in identifying tones than inexperienced learners, since they have been exposed to the target tones for a while and have

become familiar with the phonetic properties that tones carry. Furthermore, those experienced learners could possibly reach native-like perceptual ability. One study has supported this claim: Wayland and Guion (2003) found that experienced American learners of Thai outperformed inexperienced American learners when discriminating the middle and low tones in Thai and performed almost as well as native speakers of Thai.

Perceptual training. Intensive perceptual training with certain tones will improve certain L2 listeners' ability to discriminate and identify tones (Francis et al. 2008; Wang et al, 1999; Wang & Kuhl, 2003, Wayland & Guion, 2004). American listeners who had perceptual training identified Mandarin tones more accurately after training, whereas American listeners who did not participate in training made smaller progress on tone identification (Wang et al, 1999). Surprisingly, though, younger American listeners (from 6 to 14 years old) did not make greater progress on tone identification than older American listeners (19 years old) (Wang & Kuhl, 2003). Different age groups' ability to perceive L2 tones can be improved to the same degree, as long as the group members receive the same amount of target information. In another study (Wayland & Guion, 2004), Mandarin listeners also benefitted from perceptual training in discriminating mid and low tone in Thai, whereas American listeners did not show significant improvement from the pretest to the posttest. The authors attributed the differential effect of perceptual training between the two groups of listeners to the degree of intensiveness of training. Compared to Chinese listeners with tonal language experience, American listeners without prior exposure to tones need more practice to discern acoustic properties of the pair of tones. In Francis et al. (2007), both L1 Mandarin speakers and L1 American English speakers made improvement on recognizing Cantonese lexical

tones after receiving perceptual training. In fact, there was no significant difference in the two language groups' improvement from pretest to posttest.

Tonal context. Bent (2006) tested monosyllabic and trisyllabic meaningless words (all /ra/ syllables) attached with Mandarin tones. There was an effect of tonal context on the accuracy of discrimination of tone on the second syllable within a tri-syllabic utterance. The falling and the rising tones were more easily differentiated when they are preceded by a level tone and followed by a falling tone (i.e., level + Tone + falling pattern), whereas the two tones were confused in the tonal pattern of falling + Tone + rising. The author explained this phenomenon using the Perceptual Assimilation Model (PAM). The tonal environments of level + falling + falling correspond to the L+H* L+H* intonation pattern in English, whereas the level + rising + falling pattern corresponds to the L* + H English intonation pattern. These two patterns show a two-category (TC) assimilation pattern, in which discrimination is expected to be good. On the other hand, the falling + falling+ rising and falling + rising + rising patterns are both mapped onto the same English intonation pattern, L+H* L*, and show a single-category (SC) assimilation pattern, in which discrimination is expected be poor.

Syllable structure. Native English speakers have been found to experience more difficulty in discriminating Thai middle and low tone contrasts when these tones are presented in open syllables than in closed syllables (Wayland & Guion, 2003). Differences in acoustic properties of the middle and the low tones are more salient in open syllables than closed syllables.

Talker variability. In Leather (1990), L1 Dutch speakers were tested on four Mandarin tones produced within the monosyllable /y/. Participants misidentified tones

more often when the stimuli were pronounced by multiple speakers than when all stimuli were produced by the same speaker. Leather concludes that 'too much inter-speaker variability at too early a stage may prevent the learner from discovering with sufficient accuracy the prototypical forms that the exemplars presented expound' (Leather, 1990; p98)

Speech sounds. Bent et al. (2006) investigated how tonal listeners and non tonal listeners perceive pitch in speech (Mandarin tones) and nonspeech sounds (both pure tones and saw-tooth wave pulse trains). Mandarin listeners identified Mandarin tones with a higher rate of accuracy than American listeners. Also, Mandarin listeners exhibited a small range of performance (96% to 100% correct) because of a ceiling effect, whereas the English listeners' scores varied widely and did not overlap with any of the Mandarin listeners' scores (26% to 88% correct). However, there was no significant difference between Mandarin and American listeners' performance on the nonspeech pitch discrimination.

Music experience. Musicians have been found to have an advantage for identifying and discriminating Mandarin tones than non-musicians (Alexander et al., 2005). Musicians identified and discriminated Mandarin tones with a higher rate of accuracy (89% and 87%, respectively) than non-musicians (69% and 71%). This result contradicts the claim that processing music is unassociated with processing lexical tones (Kimura, 1961). Musical experience can help listeners to perceive lexical tones more accurately (Alexander et al., 2005).

Perception of Coarticulated Tones by L2 Listeners

To the best of my knowledge, only one study has been conducted to examine how coarticulated tones are perceived by L2 listeners. Broselow et al. (1987) found that L1 English speakers found the falling tone in Mandarin harder to perceive accurately when it was on the first syllable of the disyllabic words than when it was on the second syllable. They attributed the phenomenon to L1 intonation transfer, since falling pitch rarely occurs in word-initial position in English. In terms of error type, they found that the falling tone was frequently identified as a high flat tone by American listeners when it occurred word-finally. They argued that American listeners took the falling part of the falling tone as a word-ending marker and did not associate it as a part of the tone. Therefore, they only perceived the high pitch onset at the beginning of the tone and identified the tone as a high flat tone.

Production of Mandarin Isolated Tones by L2 Speakers

Chen (1974) found that native English speakers did not produce the full pitch range of Chinese tones. Their pitch range was narrower than the native pattern. Wang et al. (2003) tested production of isolated Mandarin tones by American speakers. Their participants had taken one or two semesters of Mandarin Chinese. Between these tests (pretest and post-test), half of the participants received perceptual training. Their pre- and post-training productions were judged by multiple native speakers of Mandarin Chinese and analyzed by WAVES + ESPS software. The results indicated a specific order of ease of production: Tone 1, Tone 4, Tone 2 and then Tone 3, (ranked from easiest to hardest) which is the same order as first language acquisition of Mandarin tones by Mandarin children (Li and Thompson, 1977). In terms of types of errors, it has

been found that Tone 1 was misproduced as Tone 2 or Tone 4, Tone 2 was mainly misproduced as Tone 3, Tone 3 was often misproduced as Tone 2, and Tone 4 was misproduced as Tone 1 or Tone 3. Besides the tonal contour errors, the tonal height (which is synonymous with the term 'tonal register' in other studies on Mandarin tones) errors were also addressed. It was found that the pitch height of Tone 1 was slightly lower than native production, the onset of Tone 2 was higher and the valley of Tone 2 was lower than native norms, the valley of Tone 3 was not as low as that of the native one, and both onset and offset of Tone 4 were lower than ones of the native norms. Compared to the tonal contour errors, the tonal height errors are more difficult to overcome. In other words, pitch height and pitch contour are not acquired in parallel. For American students, the shape of pitch contour is relatively easier to produce in a more native-like manner than the value of pitch height.

In conclusion, the production of Mandarin tones on monosyllabic words in the post-test was more native-like in terms of the dimensions of tonal height and tonal contour. Therefore, the auditory training of American adults to identity Mandarin tones not only improves their perception of Mandarin tones (Wang 1999), but is also effective in improving their production of the tones without any explicit training or feedback in speech production (Wang 2003).

Production of Mandarin Coarticulated Tones by L2 Speakers

At least four papers have studied the production of Mandarin coarticulated tone by American Speakers. Shen (1989) examined the tonal production in a piece of Mandarin text by American students. The participants had studied Mandarin for one semester. A hierarchy of degrees of difficulty of producing tones is proposed: Tone 4 is the hardest,

then Tone 1 is the next hardest, and Tone3 and Tone2 are the easiest. It was assumed that Tone 4 and Tone 1 are less marked than the others in perception because they both start with high pitch and may be perceived as indicative of a stressed syllable. Therefore, Tone 4 and Tone 1 are more subjected to negatively transfer by American students. However, the author did not explain why stressed syllables are less marked than unstressed syllables and harder to produce correctly.

Miracle (1989) also examined the tonal productions of American students who had studied Mandarin for at least one academic year. The target words were monosyllabic or disyllabic words in different positions in a sentence. In contrast to Shen's (1989) finding, this study showed that the error rate of Tone 2 is the highest, which Miracle attributes to interference from the English intonation system. However, he does not thoroughly explain how the English intonation system was employed by American speakers to produce Mandarin tones. Two types of errors were categorized: tonal register (tone range) errors and tonal contour (tone direction) errors. Unfortunately, the author only described what errors were made, but did not explain *why* Mandarin tones were inaccurately modified. Table 2-2 summarizes Miracle's description of tone errors.

Table 2-2. Tone errors

Tone	Tonal register errors	Tonal contour errors
Tone 1	Too low	Falling tone
Tone 2	Beginning of the tone too high	Falling tone or level tone
Tone 3	Too high	Rising tone
Tone 4	Beginning of the tone is mid	N.A

Chen (1993) analyzed the tonal errors of American adult learners of Chinese by examining tones occurring in sentences.. Six students in a second-year university conversational Chinese class participated in his production experiment. Two sequential tonal patterns used by American speakers in a four-syllable sentence were detected: one is 55---33---22---53; the other is high---low---high---low--- (high), and all syllables

carried level tones. The reasons for making the two patterns were attributed to negative transfer from English prosodic features to Mandarin tones. The transfer of English intonation modifies tones into the first pattern and the transfer of English rhythm modifies tones into the second one. However, he did not explain which kind of English intonation or rhythm was negatively transferred into American speakers' production of Mandarin sequential tones. He noticed that contour tones are either distorted flatly or produced in an opposite direction; the level tone either changes the range of the tone or became a contour tone. A hierarchy of difficulty for second-language Mandarin tonal acquisition suggested in his study agrees with the first-language acquisition order reported in Li and Thompson's study (1993): Tone 1 is the easiest one to be acquired, Tone 4 and Tone 2 are next, and Tone 3 is the most difficult one. This study provides a picture of Chinese tonal acquisition among native English speakers; however, the model for labeling the tone value is not convincing. Though the author used the numbers to describe tones, he did not provide acoustic measurements to support his description, but instead relies on his own perception. Therefore, his results were very subjective.

Wang (1995) analyzed tonal errors of disyllabic words made by American students who had been studying Mandarin in China for longer than half a year. As in other studies mentioned, Tone 1 and Tone 4 were produced correctly more often than Tone 2 and Tone3. Both tonal register errors and tonal contour errors were found in the production of Tone 2, Tone 3 and Tone 4. Only tonal contour errors were found in the production of Tone 1. The most common errors for each tone are shown in Table 2-3.

Table 2-3. Most common errors of tones in disyllabic words by American students (Wang, 1995)

Tone	Tone 1	Tone 2	Tone 3	Tone 4
Main contour errors	falling rising	Low flat (1st syllable) High flat or falling (2nd syllable)	rising	high flat rising

The most interesting part of her study is that she found that Tone 2 on the initial syllable tends to be pronounced as a low tone when the pitch of the onset of the following tone is high. She attributed the error to the transfer of English stress patterns (in a disyllabic word, one syllable is stressed and another one is unstressed). In English, the stress pattern is associated with the height of pitch. The high pitch represents the stressed syllable and the low pitch represents the unstressed syllable. Wang assumed that American speakers treated the syllable starting with high pitch – in this case the second syllable -- as a stressed syllable. Consequently, the first syllable is treated as unstressed and thus pronounced with a low tone. However, this combination of high tone and low tone was not found in the production of other tone combinations.

Conclusion

American students have difficulty producing tones on disyllabic words. However, no previous studies have empirically investigated which tones in what kind of tonal context are harder to produce accurately. Though previous studies listed the main error types for the coarticulated tones, they have not revealed what environments triggered tone value changes. From previous studies, two linguistic factors which prevent American students from producing Mandarin coarticulated tones accurately have been found: 1) English intonation combinations and 2) English stress combinations. However, these two reasons have not been fully explained in previous articles. All the tested words in Shen (1989) and Chen (1993) were already familiar to their participants, and

the participants in Wang (1995) were familiar with most of the tested words. Therefore, the effect of experience might play a role in tone production. In other words, because of the effect of practice, a word with a difficult-to-produce tone pattern might have a higher possibility of being pronounced accurately than a word with an easy-to-produce tone pattern. Therefore, it is better to use words which are new to the subjects to test their ability to produce tones.

The Relationship Between Perception and Production of L2 Features

The relationship of perception and production of SLA features will be discussed with more details in the following sections.

Since very few studies have been conducted specifically on the relationship between perception and production of acquisition of tones by L2 speakers, in this section, I will broaden my review to include previous research on the acquisition of segmental features by L2 speakers. Two subsections will be included: the order of perception and production of L2 sounds and relationship between their perception and production.

The Acquisition Order of Perception and Production of L2 Features

Production precedes perception

Studies by Goto (1971) and Sheldon et al. (1982) showed that some Japanese subjects were able to produce identifiable /r/ and /l/ tokens even though they were unable to reliably identify native English /r/ and /l/ tokens. This finding led these researchers to conclude that production can precede perception in the acquisition of a non-native contrast. Flege & Eefting (1987) also found that Dutch speakers could not identify stops in a /da /-/ ta/ synthetic continuum as well as they could produce this contrast. Smith (2001) reanalyzed the results of previous studies on the perception and

production of English /l/ and /r/ by Japanese and Korean speakers. Because good production and mediocre perception performance on the contrast /l/ and /r/ can co-occur in the learners, she concluded that production development can be independent from perception development in second language acquisition.

Perception precedes production

Borden et al. (1983) examined the relationship between perception and production of English /l/ and /r/ by Korean learners of English. They found that their participants' perception developed earlier than their production and accurate perception might be a prerequisite for accurate production. The authors note that "the ability to make phonemic perceptual judgments in a /r/ - /l/ continuum that are similar to those of English speakers also seems to improve before production" (p. 516). Sheldon (1985) statistically reanalyzed the Borden et al.'s results. This reanalysis confirmed the findings of Sheldon and Strange (1982) and failed to support the argument made by Borden et al. (1983) that accurate perception is acquired before accurate production. One of the important conclusions of Sheldon's reanalysis was that the relationship between production and perception depended on the amount of time spent in the English-speaking environment by the Korean learners, so that "as the learner's time in the U.S. increases, the probability of environment of perception exceeding production decreases" (p.111). This claim is supported by the fact that the speakers studied by Sheldon and Strange (1982) were advanced learners.

Therefore, we can say that there is no conclusive connection between speech perception and production on the contrast of /l/ and /r/ by Korean and Japanese speakers. To account for the relationship between perception and production of English /l/ and /r/ by Japanese and Korean speakers, the duration of language experience

should be considered. If the speaker is a beginning learner, they will perceive more accurately than they produce, but advanced learners may produce more accurately than they perceive.

For the vowels, Bohn and Flege (1990) found that inexperienced L1 German speakers can perceive L2 English /ɛ/ and /æ/ accurately. However, they cannot produce the contrast in a native like manner. This result confirmed the conclusion of Borden et al. (1983) that inexperienced speakers will perceive L1 segments before they successfully produce them.

There are very few studies on the order of development of perception and production of suprasegmental features. Only Wang (1999) and Wang at el. (2003) tested the relationship of perception and production of Mandarin isolated tones by American students and found that perception of Mandarin tones by American speakers preceded production of Mandarin tones, since they can perceive Mandarin tones with a high rate of accuracy but their production of tonal contours do not match the native norm.

Good perception can make good production

If the statement that good perception leads to good production is true, then in order to improve their ability to produce L2 sounds, L2 learners must accurately perceive the L2 sounds. Several studies show that perception training will help learners to differentiate and/or identify L2 segments. Furthermore, the improvement of perception leads to the improvement of production. Rochet (1995) found that L1 Mandarin Chinese speakers' perception of L2 French voice onset time (VOT) categories was more accurate after receiving auditory training. Moreover, the participant's production on their

imitation task was also more native like. Logan et al. (1991) and Bradlow et al. (1997) showed that the performance of Japanese speakers on perceiving English /r/-/l/ was improved by giving them perceptual training. Bradlow et al. (1997, 1999) showed that the transfer of perceptual learning to speech production can also occur.

In terms of suprasegmental features, several studies have been conducted to test whether perceptual training will affect the performance of perceiving tones by L2 speakers. It has been shown that for American English speakers, intensive perceptual training with certain tones will improve their ability to discriminate and identify the tones (Wang et al, 1999; Wang & Kuhl, 2003, Wayland & Guion, 2004).

Later, Wang et al. (2003) reported that the production of the same participants in her 1999 study was also improved by giving them auditory training; identification of trainees' post-test tone productions improved by 18% relative to their pretest productions. They conducted two production experiments of isolated Mandarin tones by American speakers. Compared to the tonal contour errors, the tonal height errors were more difficult to overcome. In other words, pitch height and pitch contour are not acquired in parallel. For American students, the shape of pitch contour is relatively easier to produce in a native-like manner than pitch height. In conclusion, the production of Mandarin tones on monosyllabic words after training is more native like regard with the dimensions of tonal height and tonal contour. Therefore, auditorily training American English speaking adults to identity Mandarin tones not only improves their perception on Mandarin tones (Wang 1999), but is also effective in improving their tone production without any explicit training or feedback in speech production (Wang 2003). Leather (1997) found that Dutch speakers with good tone perception also produced Mandarin

tones correctly. Specifically, the tones which were perceived correctly were also produced correctly, and the tones which were perceived wrongly were also produced inaccurately.

No studies have been conducted to test whether good perception of stress and intonation will result in good production of them when learners do not receive production training.

Good perception cannot guarantee good production

Contrary to the evidence reviewed in the previous section, two studies have shown that good perception can not guarantee accurate production. In other words, perceptual accuracy gained from training is not necessarily transferable to production. According to Wang (2004), perception training helps Chinese discriminate and identify English vowel contrasts /i/-/ɪ/ and /ɛ/-/æ/ better. However, the effects of the training were not found on their accuracy of production of these vowels. Perez (2005) also showed no significant effect on production when L1 Spanish speakers received training on discrimination and identification of English vowel contrasts /i/-/ɪ/ and /a/-/ɑ/, though there was a great improvement in perception of English vowels after the training.

Furthermore, many studies have shown that accurate perception is a necessary but not sufficient condition for accurate production. It has been found that L2 English voiced stops in coda position were frequently followed by an inserted schwa, completely deleted, or simply devoiced by L1 Mandarin (Flege et al, 1992; Hansen, 2001; Wang, 1995), though there have been no studies of how Mandarin speakers perceive syllable final stops. According to my own experience as a native Mandarin speaker, we are able to perceive voiced stops in a native like manner. Mandarin speakers presumably modify

voiced stops by using deletion or feature changing strategies because of L1 language transfer and universal constraints, yet they do not experience corresponding perceptional problems. Since Mandarin does not allow any obstruents to occur in coda position, Mandarin speakers tend to use the strategies of insertion and deletion to form a CV structure that the language allows. The choice between using insertion and deletion is based on the disyllabic preference which exists in Mandarin phonology (Hansen, 2001; Wang, 1995),

There is another possible reason why L2 speakers cannot produce L2 segments accurately. According to my observation, Mandarin speakers have no difficulty perceiving the trill /r/ in Russian. However, some speakers cannot produce the trilled /r/ accurately. Apparently, the misproduction is not motivated by misperception of the sound. Here, I attribute this kind of misproduction to an articulatory disability.

During the acquisition of tones, it was found that L2 speakers cannot produce tones accurately even if they can perceive tone without any difficulty. For American English speakers, the stress pattern and intonation pattern in English have an effect on their production of Mandarin tones (Review of Chen, 1993; Wang, 1995 Section 2.2).

Wayland et al. (2006) tested the production and perception of English stress by Thai speakers. Thai speakers tested on English non-words stressed the first syllable more often than the second syllable when the non-word was categorized as a noun, whereas they stressed the second syllable more often when the non-word was a verb. Wayland et al. also found that Thai speakers tended to place stress on a syllable with a long vowel more often than those with a short vowel or coda consonant(s). However,

neither lexical class nor syllabic structure influenced their preferred patterns of stress assignment during their perception experiment.

Pickering (2001) examined the differences in English intonation between American teaching assistants and Chinese teaching assistants. It was found that Chinese teaching assistants used more level intonation and less rising intonation than American teaching assistants. The author attributed the difference in intonation patterns to the Chinese teaching assistants' L1 language transfer rather than perceptual problems.

Good production makes good perception

To the best of my knowledge, only one study has tested the effect of production training on perception of Mandarin tones. Leather (1997) gave Dutch speakers production training on Mandarin tones and then tested their perception. It was found that good production does help Dutch speakers perceive Mandarin tones correctly. Tones which were produced correctly were perceived correctly, and the tones which were produced wrongly were also perceived wrongly.

Good production cannot promise good perception

In the previous section, I reviewed studies on production and perception of English /l/-/r/. It was found that advanced Japanese speakers may produce the English /l/ and /r/ accurately despite being unable to perceive these sounds accurately. In other words, good production cannot promise good perception

Concerning suprasegmental features, none of the previous studies have been conducted under the argument that good production can promise good perception.

Speech Learning Model (SLM)

The relationship of perception and production of L2 segments is addressed in the Speech Learning Model (SLM) (Flege, 1995). In the model, it was proposed that good

production is preceded by good perception of L2 segments. This prediction is tested by giving perceptual training to L2 learners and comparing their production before and after the training. Most studies showed significant improvement on the accuracy of production. Auditory perceptual training had a positive effect on the production of French voice onset time categories by Mandarin Chinese speakers (Rochet (1995). Auditory perceptual training improved the production performance of English /l/ and /r/ by Japanese speakers (Bradlow et al., 1997, 1999; Hazan et al. 2005). Audiovisual perceptual Training improved Japanese speakers' production of labial segments in English. (Hazan et al. 2005). However two studies showed that good perception does not predict accurate production. Wang (2004) revealed that perception training helps Chinese to discriminate and identify better on English vowels contrasts /i/-/I/ and /ɛ/-/æ/. However, the effects of the training were not found on their production. Perez (2005) also showed no significant effect on production by perception training on discrimination and identification of English vowel contracts /i/-/ɪ/ and /a/-/ɑ/ by Spanish speakers, though there is a great improvement on perceiving English vowels after the training. Furthermore, adequately produced L2 sounds may not have been well perceived. Smith (2001) reanalyzed the results of previous studies on the perception and production of English /l/ and /r/ by Japanese and Korean speakers. Based on the fact that good production and fair perception performance on the contrast /l/ and /r/ can co-occur on the learners, she drew a conclusion that production development can be independent from perception development in the second language acquisition.

Relationship between Perception by L1 Listeners and Production by L2 Speakers

Foreign accents are identified by errors in segments (vowels and consonants) and suprasegmental features (stress, tone and intonation). Misproduction of sounds may cause native speakers to misunderstand. As we know, misproduced vowels and consonants may change the lexical meaning of words, for example, the vowel /ɪ/ as in the word pit is substituted with the vowel /i:/ as in the word Pete, which is a typical Chinese accent of English (Chen, 1996); similarly, the consonants /r/ and /l/ are confused in production by Japanese speakers as in the words rice and lice (Goto, 1971; Sheldon et al., 1982). Inaccurate production of suprasegmental features such as tones can also cause lexical meaning changes. For example, American speakers frequently mispronounce 马 mǎ 'horse' as 麻 má 'hemp' (Wang, 2003).

Differences in stress placement by L2 learners or by speakers of different dialects can lead to misunderstandings as well. For example, Hungarian and Polish L2 learners tend to place stress on the initial syllable of an English word (Archibald, 1993). Indian English speakers do not raise the pitch or lengthen the duration of stressed vowels the way American English speakers do, and produce smaller phonetic differences between stressed and unstressed syllables than American English speakers do (Wiltshire & Moon, 2003).

All above production errors are serious because they can cause misunderstanding among native listeners.. Considering the need for adequate communication, language teachers should focus on correcting learners' pronunication, which will improve their ability to communicate.

Certain non native pronunciations will not change the meaning of the target word. The following statements were observed by the author. In terms of vowels, American speakers often produce Mandarin /a/ as English /ɑ/. However, this kind of misproduction will not cause any problems with lexical meaning change, since Mandarin does not have /ɑ/ as a phoneme in its vowel inventory. The vowel [ɑ] is an allopne of /a/. In terms of consonants, Mandarin speakers often produce English voiced stop /d/ as Mandarin voiceless unaspirated stop /t/ and produce English voiceless aspirated stop /t/ as Mandarin voiceless aspirated stop /t^h/. So far, no studies have found these production errors to cause a delay of understanding by native speakers of English. To examine misproduction of English stress, He (2006)[1] conducted a study on the production of English lexical stress by Mandarin speakers. In the study, Mandarin speakers used smaller duration ratio and amplitude difference to stress English disyllabic words. However, the pitch contours they used to indicate English stress were similar to those of native speakers. Since pitch level is one of the primary acoustic correlates of stress ((Wiltshire and Moon, 2003). it was assumed that using smaller duration ratio and amplitude difference to stress English words by Mandarin speakers would not affect English native speakers' judgment of the location of stress

Misproduction of intonation does not change any lexical meanings. However, using different intonation from the native pattern may cause pragmatic meaning changes, which not only marks the learner as a non-native speaker but also can cause native listeners to find the learner rude. Pickering (2001) pointed out that Chinese teaching assistants frequently use level tone and falling tone where rising tone is used

[1] The state I put here is from my individual study with Dr. Wilshire in fall 2006.

by American teaching assistants. For example, falling tone is frequently used by Chinese teaching assistants in response to incorrect answers from American students, where a rising tone was expected. The misuse of the tone confuses native listeners, since falling tone is used to affirm the response in English. To avoid the misunderstanding caused by the pragmatic change with foreign accent, language teachers need take the issue seriously and draw their students' attention to problems with intonation.

Regarding tonal errors, L1 American English speakers may produce native like L2 Mandarin tonal directions, but the contour of the tone is slightly different from native norms. The pitch height of Tone 1 is often slightly lower than native production, the onset of Tone 2 is higher and the offset of Tone 2 is lower than native norms, the valley of Tone 3 is not as low as that of the native one, and both onset and offset of Tone 4 are lower than ones of the native norm (Wang, 2003). It remains an empirical question whether the tonal contour mistake affects Mandarin speakers' perception.

Conclusion

We cannot draw a general conclusion regarding the relationship between perception and production of L2 sounds. Many linguistic factors affect the relationship. In terms of the effect of language training, good perception may lead to good production. However, L1 transfer and articulatory difficulty may prevent L2 speakers from producing L2 sounds accurately. Good production may lead to good perception; however, L1 transfer, such as categorical perception, may be barriers that keep L2 listeners from perceiving L2 sounds accurately. To determine whether misproduction of

L2 features causes a delay of understanding by L1 listeners, we need to consider the nature of the error and the language background of the L1 listeners.

CHAPTER 3
PERCEPTION AND PRODUCTION OF ISOLATED MANDARIN TONES

This chapter reports on a study on the perception and production of isolated Mandarin tones by American learners of Mandarin with different learning experience. The first section presents the perception experiment, including the methodology, results, analysis, and discussion. The second section presents the production experiment in the same manner. The third section discusses the relationship between perception and production of isolated Mandarin tones by American speakers.

Research Questions and Hypothesis

The present experiments tested perception and production of Mandarin isolated tones in monosyllabic words by experienced and inexperienced American learners. The experiments were aimed at answering the following questions:

Research question 1. Does American learners' production of isolated Mandarin tones improve with Mandarin learning experience in the classroom?

Wang et al.'s (1999, 2003) studies showed that American speakers performed better on perceiving and producing isolated Mandarin tones after they received perception training to identify these isolated Mandarin tones. In the present study, American learners with more learning experience in the classroom were expected to perform with greater accuracy than inexperienced American learners who had relatively less learning experience in identifying and pronouncing isolated Mandarin tones.

Research question 2. What are the typical tonal perception and production errors for American learners with different levels of language learning experience?

Previous studies revealed that American speakers who receive perceptual training tend to misperceive Tone 1 and Tone 4 as each other and Tone 2 and Tone 3 as each

other (Wang, 1999). They confuse tones in their production in the same way (Wang 2003). American learners who learned tones in the classroom in the present study were expected to perform as American speakers in the lab.

Research question 3. What is the relationship between perception and production of Mandarin isolated tones at different stages of the learning experience

Flege (1995) suggested that accurate perception is the predictor of accurate production. In the Speech Learning Model (SLM), the speech production accuracy of second language (L2) learners is limited by their perceptual accuracy. Thus, in the present study, the ability to perceive isolated Mandarin tones was expected to develop faster than the ability to produce these tones.

Perception Experiment

Method

Participants

All participants were American students at the University of Florida. They all studied Mandarin Chinese from Monday to Friday, one hour each day in the classroom.

The perception experiment was conducted at the end of November 2007. At the time, the twenty-two participants in the beginning level had studied Mandarin for three months; they were labeled inexperienced learners for this experiment. The sixteen participants in the intermediate level had studied Mandarin for twelve months and were labeled experienced learners for this study. They had had no previous experience learning a tone language prior to learning Mandarin at the University of Florida. They had no speech or hearing problems. Participants earned extra credit for completing the study.

Stimuli

A set of 12 monosyllabic words containing the four lexical Mandarin tones (three words per tone) was produced by a native female speaker of Mandarin. A single speaker was used rather than multiple speakers in order to avoid the possibility that participants might misperceive the tones because they were unable to generalize pitch changes to different talkers. A female speaker was selected because of Wang et al.'s (1999) finding that American students found female talkers more intelligible than male ones. All stimulus words were the CV syllable /na/. The syllable structure and the segments chosen in the study all exist in English. This design avoids the possibility that American listeners could misperceive the tones because they were distracted by the novel segments or syllable structure rather than focusing on the suprasegmental features.

The contour of F_0, the primary acoustic cue to identify Mandarin tones, and the secondary cues of amplitude envelope and syllable duration (Fu et al., 1998; Liu & Samuel, 2004; Whalen & Xu, 1992) were kept as the speaker naturally produced them. They were not manipulated. Tone 3 did not have creaky voice, and Tone 4 did not have glottalization (Liu & Samuel, 2004), which are reported as acoustic cues to identify those two tones additionally.

Procedure

All stimuli in the perception experiment were presented in random order. Each type of tone was presented three times. For each trial, the participant viewed a syllable *na* on the screen of the computer. The participant clicked on the syllable to hear the corresponding sound and then wrote down the tone value of the syllable of each word on a provided answer sheet. Participants could adjust the volume to their comfort level.

The participants could hear the sound as many times as they needed. Before the real test, there were three warming up trials. The results from these trials were not analyzed. No feedback about their performance was provided during the test.

Results and Analysis

In this section, both tone identification accuracy rate and tone error type from the perception will be reported.

Accuracy rate

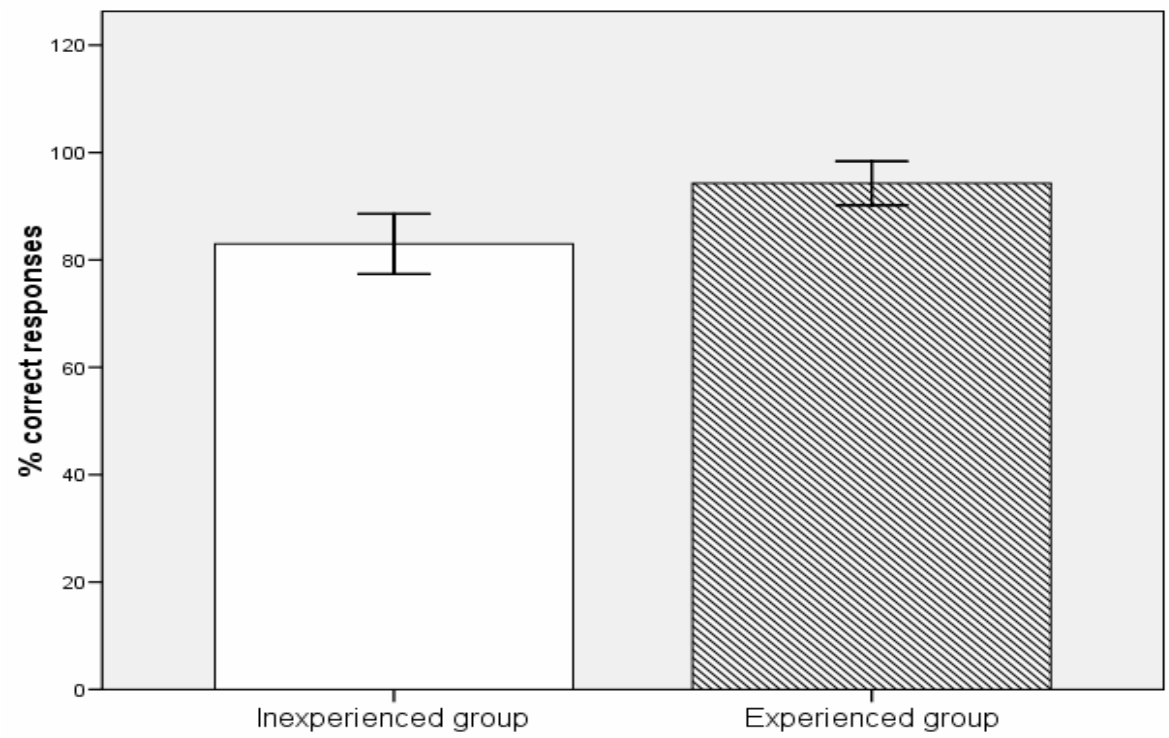


Figure 3-1. Mean percent correct identification of the four Mandarin tones for inexperienced and experienced learners of Mandarin on isolated tones. Error bars indicate standard error.

The overall results in Figure 3-1 show that experienced learners of Mandarin performed better on identifying isolated Mandarin tones than inexperienced learners (94.28% vs. 82.99% correct). A one-way ANOVA showed this difference to be significant at the .01 level [F (1, 150) = 9.213, p=0.003].

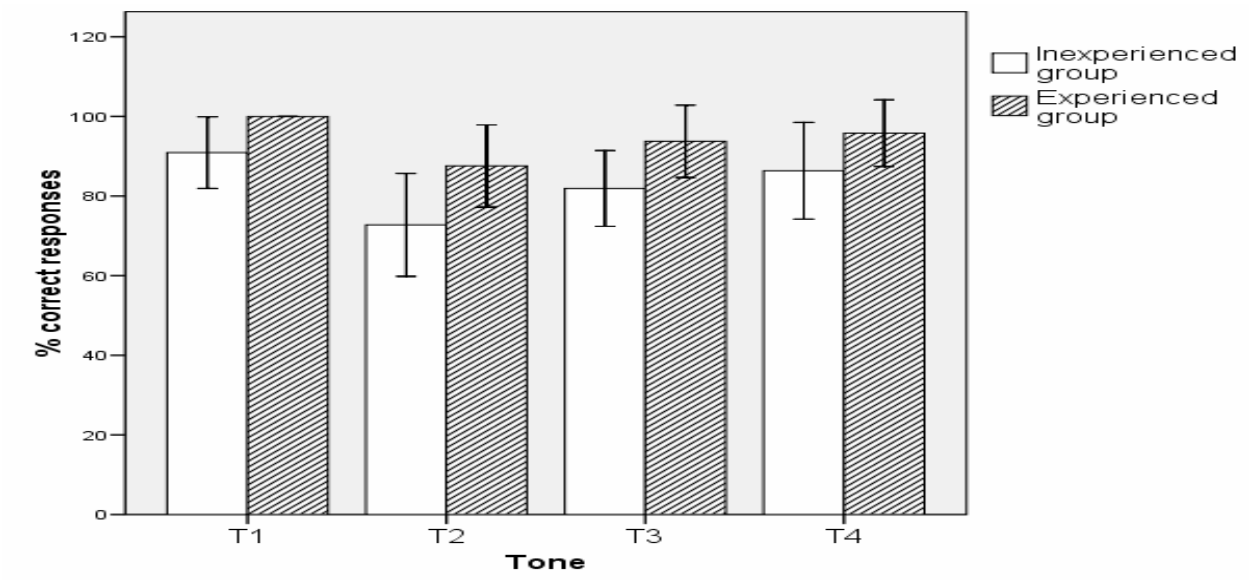


Figure 3-2. Inexperienced and experienced learners' mean percent correct identification for each tone. Error bars indicate standard error.

The accuracy rate of perception for each isolated tone by groups of American learners is illustrated in Figure 3-2. A two-way repeated ANOVA, with Tone as within subject factor and Group as between subject factor, showed a main effect for Group [$F(1, 36) = 5.86$, $p<.03$], indicating that experienced learners performed better than inexperienced learners on identifying tones overall, and a main effect of Tone [$F(3, 34) = 4.415$, $p<.02$] suggesting that their performance varied across tones. A Turkey-HSD test (Bonferroni adjusted $p<.05$) showed that their identification of Tone 2 (79%) was significantly worse than their identification of Tone 1 (94.75%), but not significantly worse than Tone 4 (90.34%) or Tone 3 (86.89%). The interaction between Group and Tone was nonsignificant [$F(3, 34) = .155$, $p= .926$], showing that experienced learners consistently outperformed inexperienced learners across all four tones. These results are consistent with those of Wang et al. (1999).

Error type

Comparing Table 3-1 to Table 3-2, it is apparent that there is no difference of error type between the two language proficiency groups. Tone 1 and Tone 4 showed bidirectional confusion patterns, since there was no significant difference between frequencies of misproducetion of Tone 1 as Tone 4 and misproduction of Tone 4 as Tone 1 by inexperienced group (χ^2=.895, p=.344). However, Tone 1 and Tone 4 were identified with very high accuracy rates by experienced group. Tone 2 and Tone 3 showed bidirectional confusion patterns, since there was no significant difference between frequencies of misproducetion of Tone 2 as Tone 3 and misproduction of Tone 3 as Tone 2 by either inexperienced group (χ^2=2.533, p=.213) or experienced group (χ^2=1.103, p=.294).

Table 3-1. Percentages of misperception made by inexperienced American learners of Mandarin, with the number of frequency in parentheses

Tone	Identified as			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	**90.9 (60)**	3.0 (2)	0	6.1 (4)
Tone 2	0	**72.7 (48)**	27.3 (18)	0 (0)
Tone 3	0 (0)	18.2 (12)	**81.8 (54)**	0 (0)
Tone 4	10.6 (7)	1.5 (1)	1.5 (1)	**86.4 (57)**

Table 3-2 Percentages of misperception made by experienced American learners of Mandarin, with the number of frequency in parentheses

Tone	Identified as			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	**100 (48)**	0 (0)	0 (0)	0 (0)
Tone 2	0 (0)	**87.5 (42)**	12.5 (6)	0 (0)
Tone 3	0 (0)	6.3 (3)	**93.8 (45)**	0 (0)
Tone 4	0 (0)	4.2 (2)	0 (0)	**95.8 (46)**

Discussion

The results of the accuracy rate of perception on isolated Mandarin tones suggest that language learning experience affects American learners' perception of Mandarin tones. The longer they study, the more accurately they perceive tones in general.

In terms of error types, it was found that the most confusing Tones for American learners are Tone 4, which is often confused with Tone 1, and Tone 2, which is frequently confused with Tone 3. These results suggest that American speakers paid more attention to the pitch value at the onset of tones than at the offset of the tones. This would explain why Tone 4 is confused with Tone 1, since both have a high pitch onset, and the low pitch offset of Tone 4 might be interpreted as simply the end of the utterance. Tone 2 and Tone 3 both have a mid-range onset pitch. Furthermore, both tones show falling + rising contours in their phonetic descriptions, though the falling part of Tone 2 is not perceived by native speakers of Mandarin. Therefore, some American speakers did not form a categorical perception on Tone 2.

Production Experiment

Method

Participants

The same participants who completed the perception experiment completed this production experiment as well.

Stimuli

The tested words in the study were non words. The participants were instructed to produce the syllable /na/ with the indicated tone. A sonorant consonant was selected because a non-sonorant consonant would perturb the F_0 of the following vowel. The vowel /*a*/ was selected to minimize vowel intrinsic pitch effect on pitch production.

Procedure

All stimuli in the production experiment were presented in random order. Each tone was presented three times. For each time, the participants viewed the word /na/ (with tone indicated) in Pinyin on the screen of the computer. Participants pronounced each word twice. The first pronunciation was not analyzed. The second try was considered indicative of what the participant really intended and would be taken into analysis. Before the real test, there were three warming up trials. The production from these warm-up trials were not analyzed. No feedback about their performance was provided during the test. A total of 456 stimuli (38 subjects x 3 tokens x 4 tones) were included in subsequent analyses.

Judgment

Two native speakers of Mandarin with a Beijing accent transcribed the tones produced by participants based on their perception. Answer sheets were provided. The produced stimuli were printed with no tonal diacritics. Judges needed to provide a tonal diacritic corresponding to the tone they heard (Wang 2003).

Results and Analysis

In this section, the results will be presented for three measures accuracy rate, error type, and duration of the vowel.

Accuracy rate

From Figure 3-3, we can see that inexperienced learners and experienced learners showed comparable performance in producing isolated Mandarin tones (84.88% vs. 85.94%). A one-way ANOVA showed no evidence of a difference in accuracy rate for the two groups [$F(1, 150) = 0.053$, $p=0.817$].

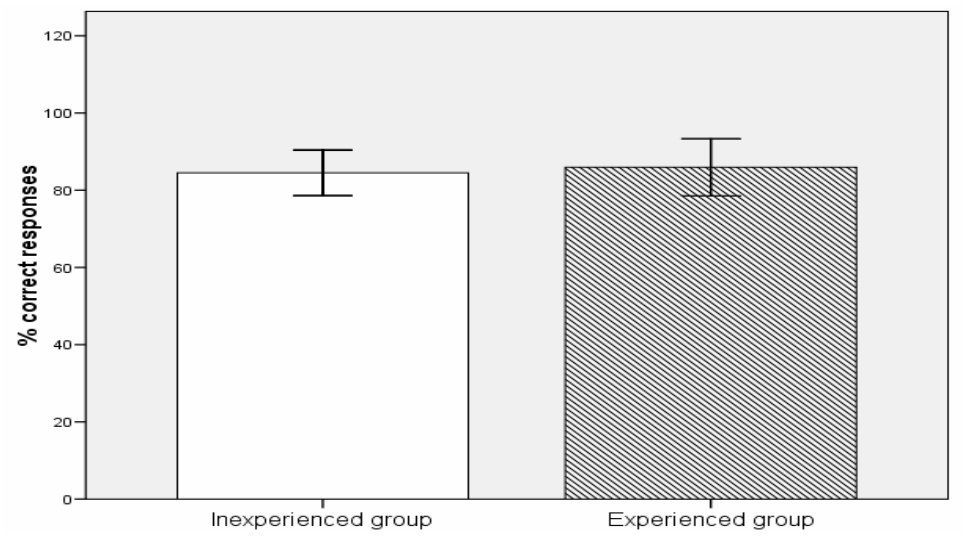


Figure 3-3. Mean percent correct production of the four Mandarin tones for inexperienced and experienced learners of Mandarin on isolated tones. Error bars indicate standard error.

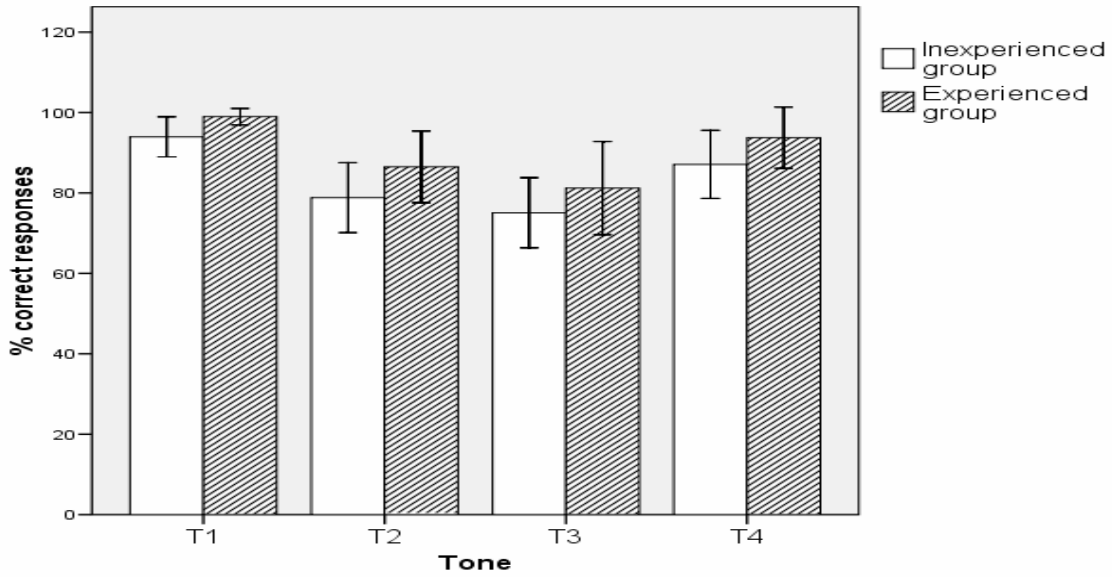


Figure 3-4. Mean percent correct production of the four Mandarin tones for inexperienced and experienced learners of Mandarin on each tones. Error bars indicate standard error.

The accuracy rate of production for each isolated tone by the two groups of American learners is illustrated in Figure 3-4. A two-way repeated ANOVA with Tone as within subject factor and Group as between subject factor showed no main effect for Group [F (1, 36) = .067, p=.797], giving us no evidence of a difference between

experienced learners and inexperienced learners in their tone production accuracy. However, there was a main effect for Tone [F (3, 34) = 9.043, p< .001]. A Tukey-HSD test (Bonferroni adjusted p<.05) showed that only Tone 3 (68.42%) was significantly worse than Tone 1 (97.39%), Tone 2 (85.13%) and Tone 4 (89.47%). The interaction between Group and Tone did not reach significance [F (3, 34) = .047, p= .986], showing that experienced learners consistently outperformed inexperienced learners across all four tones.

Error type

As shown in Tables 3-3 and 3-4, Tone 1 was produced with a very high accuracy rate by both groups of learners, and Tone 4 was sometimes produced as Tone 1 by both groups of learners. Tone 3 was more often misproduced as Tone 2 than the opposite by both inexperienced learners (χ^2=5.927, p<.02) and experienced learners (χ^2=3.920, p<.05).

Table 3-3. Percentages of production confusion made by inexperienced American learners of Mandarin, with the number of frequency in parentheses

Tone	Identified as				
	Tone 1	Tone 2	Tone 3	Tone 4	None
Tone 1	97 (64)	3.0 (2)	0 (0)	0 (0)	0 (0)
Tone 2	1.5 (1)	86.4 (57)	9.1 (6)	1.5 (1)	1.5 (1)
Tone 3	0 (0)	24.2 (16)	68.2 (45)	1.5 (1)	6 (4)
Tone 4	9.1 (7)	1.5 (1)	0 (0)	89.4 (59)	0 (0)

Table 3-4. Percentages of production confusion made by inexperienced American learners of Mandarin, with the number of frequency in parentheses

Tone	Identified as				
	Tone 1	Tone 2	Tone 3	Tone 4	None
Tone 1	97.7 (47)	2.1 (1)	0 (0)	0 (0)	0 (0)
Tone 2	6.3 (3)	85.4 (41)	8.3 (4)	0 (0)	0 (0)
Tone 3	2.1 (1)	22.9 (11)	70.8 (34)	0 (0)	4.2 (2)
Tone 4	6.3 (3)	2.1 (1)	0 (0)	91.7 (44)	0 (0)

Duration of the tone

According to previous studies on duration of Mandarin lexical tones (Fu & Zeng, 2000; Howie, 1976; Tseng, 1990), Tone 3 is the longest tone, Tone 2 and Tone 1 are shorter, and Tone 4 is the shortest tone. In the following paragraphs, the durations of Tone 2, Tone 3 and Tone 4 produced on the syllable /na/ were measured. Using both waveforms and wide band spectrograms generated by Praat (Boersma and Weenink 2007), the duration of the tone was measured from the beginning of the nasal consonant to the end of the vowel /a/. The data were analyzed (by using t-tests

Table 3-5. The mean duration of tones produced by learners who confused Tone 2 as Tone 3, in ms.

Tones	T2 produced as T2	T2 produced as T3	T3 produced as T3
Duration	617	539	703
s.d.	165	144	130

The most frequent mistake for inexperienced and experienced learners when attempting to produce Tone 2 is to produce it as Tone 3. The duration of Tone 3 produced for intended Tone 2 (557 ms) is significantly shorter than Tone 3 produced for intended Tone 3 (0.604s) (p=0.013). However, the duration of Tone 3 produced for intended Tone 2 is not significantly longer than Tone 2 produced for intended Tone 2 (0.471s) (p=0.355).

Table 3-6. The mean duration of tones produced by learners who confused Tone 3 as Tone 2, in ms.

Tones	T3 produced as T3	T3 produced as T2	T2 produced as T2
Duration	604	557	471
s.d.	92	97	116

The most frequent error for inexperienced and experienced learners attempting to produce Tone 3 is to produce it as Tone 2.The duration of Tone 2 produced for intended Tone 3 (557 ms) is significantly longer than that of Tone 2 produced for intended Tone 2 (471 ms) (p=0.034). However, the duration of Tone 2 produced for intended Tone 3 is not significantly shorter than Tone 3 (604ms) (p=0.462).

Table 3-7. The mean duration of tones produced by learners who confused Tone 4 as Tone 1, in ms.

Tones	T4 produced as T4	T4 produced as T1	T1 produced as T1
Duration	263	267	559
s.d.	105	58	195

The most frequent error for both groups of American learners attempting to produce Tone 4 is to produce it as Tone 1.The duration of Tone 1 produced for intended Tone 4 (267 ms) is significantly shorter than that of Tone 1 produced for intended Tone 1 (559 ms) (p=0.005). However, the duration of Tone 1 produced for Tone 4 is not significantly longer than that of Tone 4 produced for intended Tone 4 (263 ms) (p=1.000).

Discussion

Experienced learners were no more accurate in producing isolated tones than were relatively inexperienced learners. This result suggested that perhaps learning experience of twelve months is not sufficient to make a difference in production.

Tone 4 produced as Tone 1 might be due to a pragmatic concern since an isolated sharp falling intonation is associated with a rejection intonation, but it is not due to a perception problem since none of participants who misperceived the isolated Tone 4 were participants who misproduced the tone. The reasons to choose Tone 2 as a substitution are analyzed from a perception consideration and a production consideration. Misperception of Tone 2 as Tone 3 by participants may lead them to misproduce Tone 2 as Tone 3. However, the perception problem is a minor reason for this phenomenon, since only five out of eighteen participants who misperceived Tone 2 as Tone 3 misproduced Tone 2 as Tone 3 as well.

American speakers could misproduce Tone 3 as Tone 2 because of L1 transfer. Tone 3 is a novel tone for American speakers because it does not exist in the English intonation inventory. Therefore, they produced Tone 3 with more difficultly.

It was also found that although the pitch contour of the tone category may be produced incorrectly, the duration of the target tone category is produced correctly. Therefore, we can say that the duration of a tone category is acquired faster than the pitch contour of that tone category. Furthermore, unlike native speakers of Mandarin, American speakers of Mandarin paid more attention to duration, which is not primary acoustic character to native speakers of Mandarin, than to pitch contour when producing different tones. This is not a unique phenomenon in the field of acquisition of L2 sounds that L2 speakers paid attention to acoustic cues different from those used by native speakers to perceive or produce L2 features. For example, L1 Mandarin speakers producing L2 English only focus on the pitch difference between unstressed and stressed syllables and ignore the duration and amplitude differences (Wang 2008). In this study, the preference of using duration rather than pitch to differentiate tones might be due to their English language background. The change of pitch does not lead to the change of vowel quality in English. However, the change of duration of vowels will change the vowels quality, such as /i/-/ɪ/ (71 vs. 31 ms) (Flege et al., 1997). Therefore, American speakers are more sensitive to the difference of duration of tones.

Relationship between Perception and Production of Isolated Tones

In this section, the relationship between perception and production of Mandarin isolated tones is discussed in terms of both the two groups and individual learners within the groups.

Results

Figure 3-5 shows that both groups were highly accuracy at both perceiving and producing Mandarin isolated tones (Inexp: 82.99% & 84.50%; Exp: 94.28% & 85.94%).

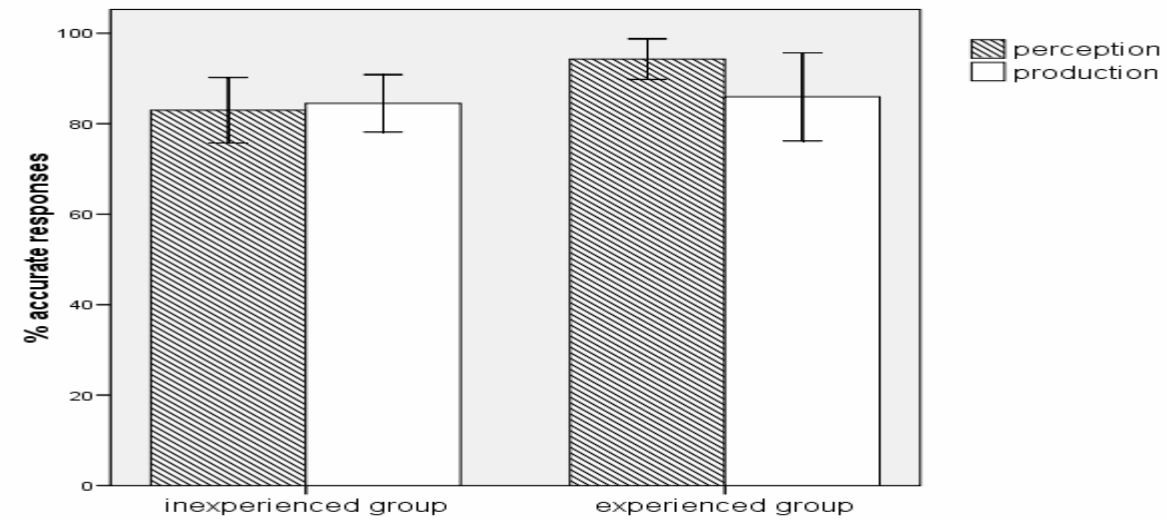


Figure 3-5. Mean percentage of accuracy of perception and production of Mandarin isolated tones by two groups. Error bars indicate standard error.

The accuracy rate of experienced learners perceiving Mandarin isolated tones is significantly better than their accuracy rate of producing the tones [t (126)=1.974, p=0.051], whereas there is no significant difference between the accuracy rates of perception and production of inexperienced learners [t (174)= -.371, p=0.711].

Table 3-8 displays the accuracy rate of perception and production of isolated Mandarin lexical tones by inexperienced American learners. According to a Pearson Correlation test, the perception and production scores are not correlated (r=0.04, p=0.372).

Table 3-9 shows that accuracy rates of perception and production of isolated Mandarin tones by experienced American learners are also not significantly correlated (r=0.013, p=0.671).

Table 3-8. Individual Inexperienced American learners' perception and production accuracy rate. These data are across all isolated lexical tones.

Participant	Perception	Production	Difference
1	100	100	0
2	100	100	0
3	100	91.75	8.25
4	75	91.75	-16.75
5	100	83.25	16.75
6	67	91.75	-24.75
7	83.5	75	8.5
8	100	91.75	8.25
9	83.5	50	33.5
10	83.25	66.75	16.5
11	83.25	66.75	16.5
12	41.75	91.75	-50
13	75	91.75	-16.75
14	75	100	-25
15	91.75	75	16.75
16	83.5	100	-16.5
17	100	91.75	8.25
18	100	100	0
19	49.75	66.5	-16.75
20	58.25	83.25	-25
21	91.75	91.75	0
22	83.5	58.5	25
Total	82.99	84.5	-1.51

Table 3-9. Individual experienced American learners' perception and production accuracy rate. These data are across all isolated lexical tones.

Participant	Perception	Production	Difference
1	83.25	91.75	-8.5
2	100	75	25
3	75	91.75	-16.75
4	100	100	0
5	91.75	100	-8.25
6	91.75	100	-8.25
7	100	83.25	16.75
8	100	100	0
9	100	100	0
10	75	75	0
11	100	66.5	33.5
12	100	100	0
13	100	91.75	8.25
14	91.75	25	66.75
15	100	83.25	16.75
16	100	91.75	8.25
Total	94.25	85.84	8.34

Discussion

Inexperienced learners' perception developed faster than production of isolated Mandarin tones, whereas experienced learners showed comparable development of both perception ability and production ability.

Overall, there is no correlation between the accuracy of perception of isolated Mandarin tones and the accuracy of production of isolated Mandarin tones by either inexperienced or experienced learners. However, the majority of individual students developed the ability to identify isolated Mandarin lexical tones faster than their ability to pronounce these tones.

Conclusion

Inexperienced learners have achieved accuracy rates of perception and production over 80%, except perception of Tone 2 and production of Tone 3. Experienced American learners of Mandarin achieved an accuracy rate over 80% across all four tones, and they performed better on perceiving – but not producing – isolated Mandarin tones than inexperienced American learners. These results suggest that learning experience of three months which inexperienced learners had is sufficient for American speakers to be familiar with Mandarin isolated lexical tones. (not sure what you mean by 'concept of Mandarin tones'. The experienced group's ability to perceptually categorize tones is stronger, as indicated by the fact that they identified tone with a higher accuracy rate.

This study also confirmed the results from Wang et al (1999, 2003) that to American learners, Tone 2 is the hardest tone to be perceived accurately and Tone 3 is the most difficult tone to be produced accurately.

Different from the previous study by Wang et al in 2003, the present study detected from errors made by American learners that they tend to use duration of tones rather F0 of tones to produce the difference among Tone 2, Tone 3 and Tone 4.

Furthermore, we analyzed the relationship between the perception and production of isolated tones by American learners in terms of their accuracy rates on identifying and producing tones. As a group, we find that the experienced group's perception ability was better than their production ability,. whereas inexperienced learners perceive and produce isolated tones at a comparable level. However, .there was no correlation between the perception and the production scores among either experienced or inexperienced learners.

CHAPTER 4
PERCEPTION OF COARTICULATED MANDARIN TONES

This chapter reports on a study on the perception of coarticulated Mandarin tones by American learners of Mandarin with different learning experience. The first section introduces the research questions and hypotheses. The second section presents the methodology, including participants, stimuli, procedure, and data analysis. The third section reports on the data and statistical analysis, and the fourth section discusses the results.

Research Questions and Hypotheses

In the present experiment, perception of Mandarin coarticulated tones in disyllabic words by inexperienced and experienced American learners was investigated. The experiment was aimed at answering the following questions:

Research question 1. Does American learners' identification of Mandarin coarticulated tones improve with Mandarin learning experience?

Because they had more learning experience, experienced learners were expected to perceive Mandarin coarticulated tones better than inexperienced learners. In other words, experienced learners were expected to identify these tones with a higher accuracy rate.

Research question 2. What are the typical tonal perception errors for American learners with different amount of language learning experience?

There are two dimensions to Mandarin tones. One is tonal direction and the other is tonal height. Therefore, it was assumed that there would be two error types, tonal direction misperception and tonal height misperception. Since tonal direction is more emphasized in the classroom teaching, the frequency of tonal direction (Xing, 2006)

misperception was expected to decrease with increasing language learning experience in classroom.

Research question 3. What are the linguistic factors that affect American learners' ability to perceive coarticulated Mandarin tones accurately?

Two linguistic factors were expected to influence American learners' ability to identify coarticulated Mandarin tones. One is syllable position (initial syllable vs. final syllable). Previous studies have not given an indication of which tones might be easier to perceive on which syllables, so no hypothesis regarding ease of perception according to syllable position was made. The other factor is tonal environment, which consists of three subordinate factors: tonal condition (isolated tones vs. coarticulated tones), tonal sequence (identical tones vs. nonidentical tones) and tonal context (compatible context vs. conflicting context).

Tones in monosyllabic word were labeled as isolated tones and tones in disyllabic words were labeled as coarticulated tones. Since pitch contours and pitch level of tones in coarticulation are altered due to coarticulatory effects of the neighboring tones; coarticulated tones were expected to receive a lower accuracy rate than those in isolation. If two identical tones occurred next to each other in a disyllabic word, such as Tone 1 + Tone 1, the tonal environment was labeled an identical tone sequence. If two different tones occurred in a disyllabic word, such as Tone 1 + Tone 2, the tonal environment was called a nonidentical tonal sequence. Considering the effect of repetition of tones, the acoustic features of tones in an identical tonal sequence should be enhanced due to the repetition. Therefore, tones in an identical tone environment

were expected to be identified with a higher accuracy rate than those in a nonidentical tone environment.

According to Xu (1994), a compatible tonal context occurs when the value of the pitch of the offset of the preceding tone and the pitch value of the onset of the following tone are similar, such as Tone 2 + Tone 4, which have a high pitch offset and onset respectively, and a conflicting tonal context occurs when the pitch value of the offset of the preceding tone and that of the onset of the following tone are substantially different, such as Tone 1 + Tone 2, which have a high offset but a low onset respectively. The tonal direction in a compatible tonal context stays as it is, whereas the degree of pitch change in a conflicting tonal context is not as substantial as in a compatible tonal context. In other words, the pitch range of the tone becomes smaller in a conflicting tonal context than in a compatible tonal context. Therefore, the acoustic features of tones are weakened or changed by the conflicting tonal context. Tones in a compatible tonal context were therefore expected to be identified with a higher accuracy rate than those in a conflicting tonal context.

In some circumstances, the pitch direction of the tone is even changed. When Tone 1 occurs after Tone 3 or Tone 4, the direction of Tone 1 changes into rising from flat due to the conflicting tonal environment (Xu, 1997). In another case, Tone 3 loses its rise and becomes a falling tone when Tone 1, Tone 2 or Tone 4 occurs after it. It was expected that the accuracy rate of identification of these tones in which tonal direction had changed would be lower than the accuracy rate for those whose tonal direction was unchanged.

Table 4-1. Mandarin tonal combinations in tonal environments (positive factors are unshaded; negative factors are shaded.)

Tonal context environment Tonal combination	Identical tone sequence	Nonidentical tone sequence	Compatible tonal context	Conflicting tonal context
T1+T1	X		X	
T1+T2		X		X
T1+T3		X		X
T1+T4		X	X	
T2+T1		X	X	
T2+T2	X			X
T2+T3		X		X
T2+T4		X	X	
T3+T1		X		X
T3+T2		X	X	
T3+T4		X		X
T4+T1		X		X
T4+T2		X	X	
T4+T3		X	X	
T4+T4	X			X

Methods

Participants

The participants in the study were all from the participant pool in the study presented in Chapter 3. The fourteen participants in the beginning level had been studying Mandarin Chinese for three months, and the fourteen participants in the intermediate level had been studying Mandarin Chinese for twelve months. All of them were able to perceive each isolated lexical Mandarin tone with at least a 67% accuracy rate (at least two correctly identified tokens out of every three stimuli). Therefore, the misperceived coarticulated tones from the current study should not be due to participants' inability to perceive isolated tones.

Stimuli

The perception study tested 45 non word stimuli, three stimuli for each of the fifteen possible disyllabic tone combinations. This design was to prevent participants from using their tonal knowledge of words with which they might already learn to identify

tones. The 45 words were produced by the same native female speaker of Mandarin mentioned in Section 3.1 and exhibited the four lexical Mandarin tones. The words were all of the structure /na.na/.

As with the isolated tone stimuli, the contour of $F_{0,}$ the amplitude envelope, and syllable duration were kept as the speaker naturally produced them. There is no creaky voice for Tone 3 and no glottalized voice quality for Tone 4.

Procedure

All stimuli in the perception experiment were presented in random order and divided into three 15-stimulus blocks. Between each block, the participant was required to rest for one minute. Each tone combination was presented three times. For each stimulus, the participant viewed the spelling nana on the computer screen. The participant clicked on the word to hear the corresponding sound and then wrote down the tone values of both syllables of each word on the provided answer sheet. Participants could adjust the volume to their comfort level and hear the sound as many times as they needed. Before the real test, there were three warm-up trials, and the results from the three warm-up trials were not analyzed.

Results and Analysis

In this section, the results will be presented from two perspectives: accuracy rate and error type.

Accuracy Rate

Three factors were considered in the study: language learning experience (inexperienced vs. experienced), syllable position (initial position vs. final position), and tonal environment (identical tonal sequence vs. nonidentical tonal sequence, compatible

tonal context vs. conflicting tonal context, changed tonal direction environment vs. unchanged tonal direction environment).

Effects of language learning experience

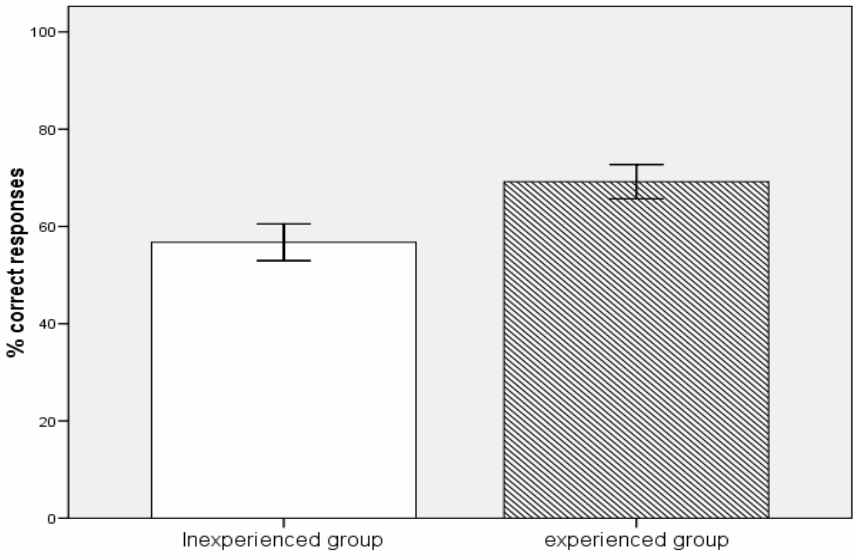


Figure 4-1. Mean percent correct identification of coarticulated Mandarin tones across all stimuli for inexperienced and experienced American learners of Mandarin. Error bars indicate standard error.

The results in Figure 4-1 show that the experienced learners of Mandarin outperformed inexperienced learners at identifying coarticulated Mandarin tones (69.21% vs. 56.75%). This difference is highly statistically significant [t (1, 838) = -4.82, p<.001].

There is a significant difference between inexperienced and experienced American learners of Mandarin on identifying individual tones in a coarticulated environment. A two-way repeated ANOVA, with Tone as within subject factor and Group as between subject factor, showed a main effect for Group [F (1,166) = 21.972, p<.001], indicating significant improvement of accuracy on identifying tones from inexperienced learners (50.964%) to experienced learners (70.149%). A significant main effect of Tone was

also found [F (3,166) = 17.21, p<.001]. Post hoc analyses (Bonferroni adjusted p<.05) showed that Tone 1 perception was significantly more accurate than Tone 4. In addition, both Tone 1 and Tone 4 were significantly more accurate than Tone 2 and Tone 3. Perception accuracy of Tone 2 and Tone 3 were, however, comparable. There was no significant interaction between Group and Tone [F (1,166) = 1.340, p=.263], which means the better performance of identifying coarticulated tones by experienced learners was consistent across all four tones.

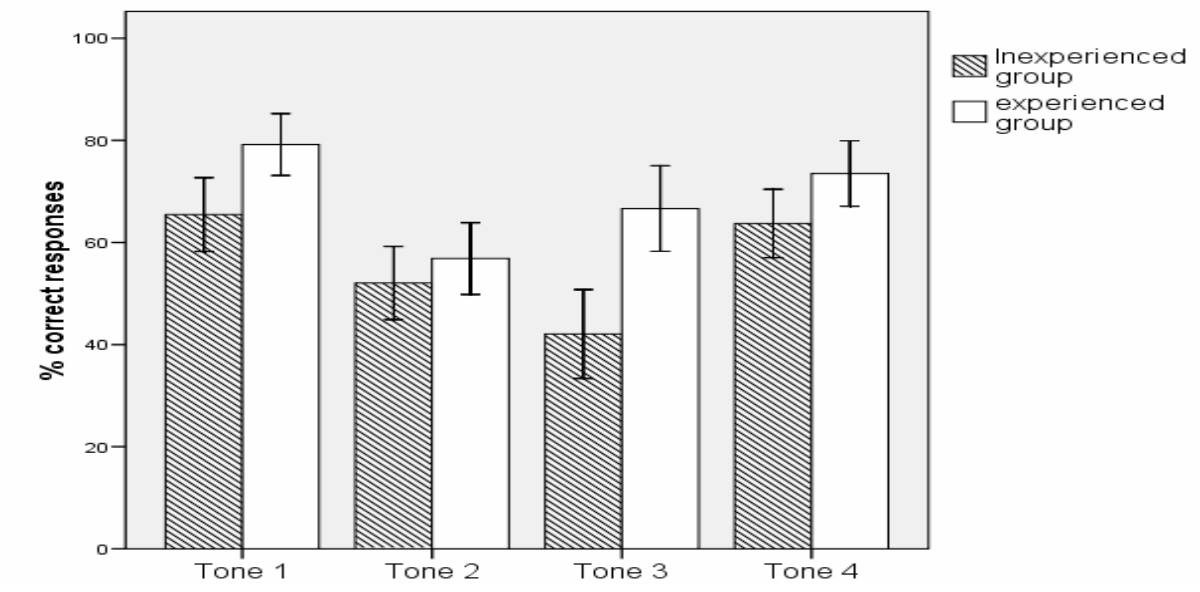
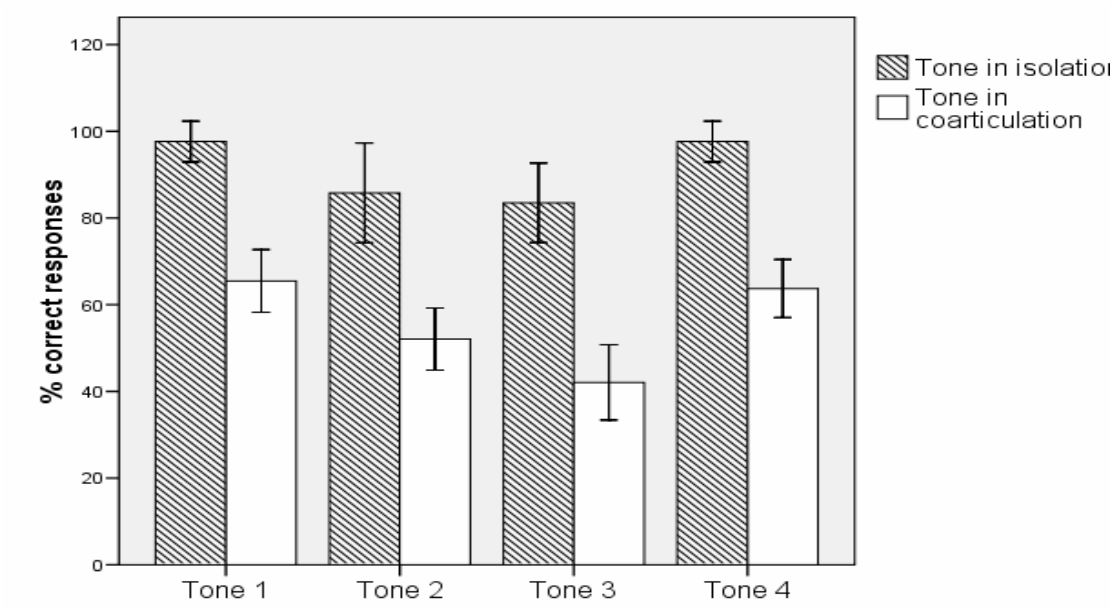


Figure 4-2. Mean percent correct identification of coarticulated tones across both syllables by inexperienced and experienced American learners of Mandarin. Error bars indicate standard error.

Effects of tonal condition

Figures 4-3 A and B show how inexperienced and experienced groups of American learners identify Mandarin tones under different tonal conditions. Overall, the condition of tones in isolation allowed greater accuracy of identification (94.089% accurate responses) for listeners than did the condition of tones in coarticulation (57.741% accurate responses). A three-way repeated measures ANOVA, with Tonal

A)



B)

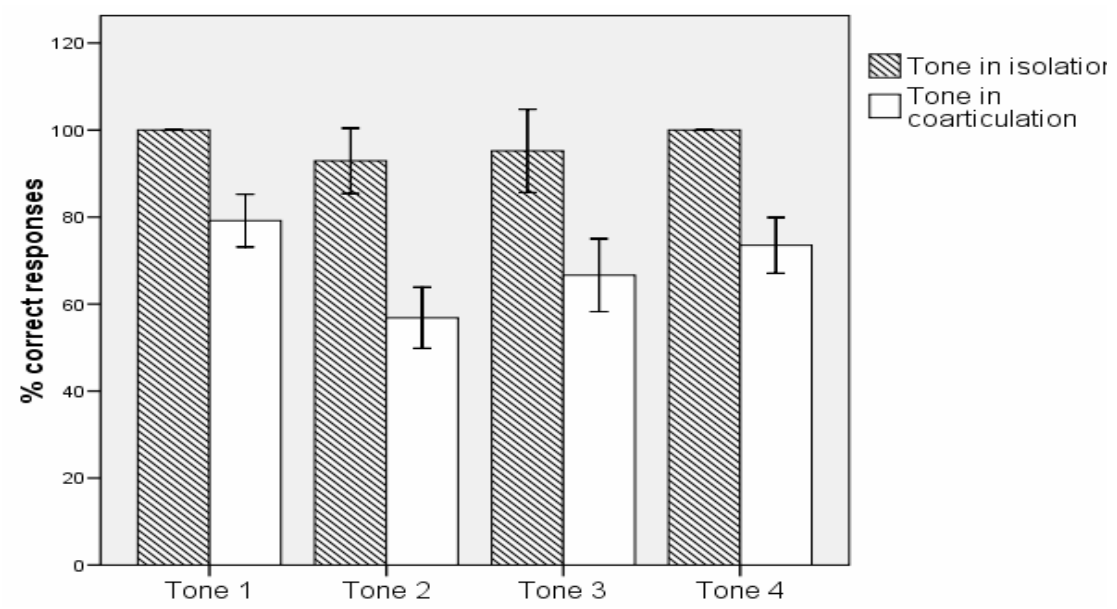


Figure 4-3. Mean percent correct perception for (A) Inexperienced and (B) Experienced groups for tones in isolation vs. tones in coarticulation. Error bars indicate standard error.

Condition (isolation and coarticulation) and Tone (Tone 1, Tone 2, Tone 3, Tone 4) as the within-subject factor and Group (inexperienced, experienced) as the between-

subject factor, yielded a main effect of Tonal Condition [F (1, 26) = 91.558, p< .001], Group [F (1, 26) = 7.739, p=.01], Tone [F (3, 24) = 20.314, p< .001]. It was also found that the interaction of Tonal Condition X Tone is significant [F (3, 24) = 14.33, p< .001]. Post hoc analyses (Bonferroni adjusted p<.05) was shown that, averaged across both groups, the accuracy rate of Tone 1, Tone 2, Tone 3 and Tone 4 were not significant different from each other, when they were in isolation. However, when they were in coarticulation, the accuracy rates of Tone 1 and Tone 4 were significantly higher than the accuracy rates of Tone 2 and Tone 3. No other significant interactions were found.

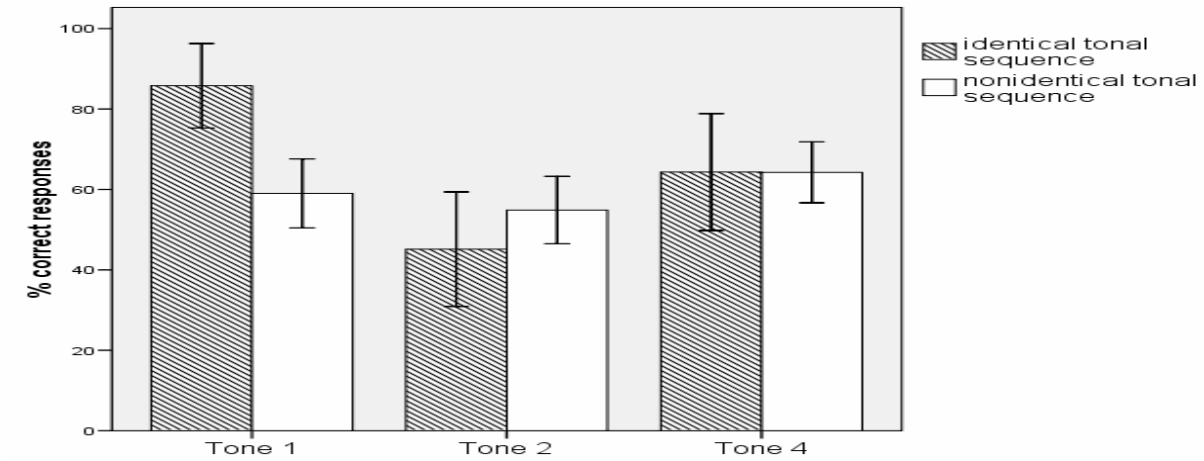
Effects of tonal sequence

Since a Tone 3 + Tone 3 series always changes to Tone 2 + Tone 3, perception of Tone 3 was always in a nonidentical tonal sequence. Therefore, only Tone 1, Tone 2 and Tone 4 will be discussed here for comparison of results under the two tonal sequences.

Figures 4-4 A and B show how inexperienced and experienced groups of American learners identify Mandarin tones under different tonal sequences. Overall, the identical tonal sequence allowed greater accuracy of identification (70.631% accurate responses) for listeners than did the nonidentical tonal sequence (64.285% accurate responses). However, the three-way repeated measures ANOVA with Tonal Environment (identical, nonidentical) and Tone (Tone 1, Tone 2, Tone 3, Tone 4) as the within-subject factors and Group (inexperienced, experienced) as the between-subject factor did not revealed the main effect of Tonal Environment [F(1, 54) = 2.739, p= .104], but it yielded the main effect of Group [F(1, 54) = 4.965, p< .05] confirming that, averaged across both tonal sequences, experienced learners out performed inexperienced learners. A marginally significant interaction of Tonal sequences X Tone

X Group [$F(2, 53) = 2.988$, $p = .059$] was also found. Further analyses were conducted to examine this interaction.

A)



B)

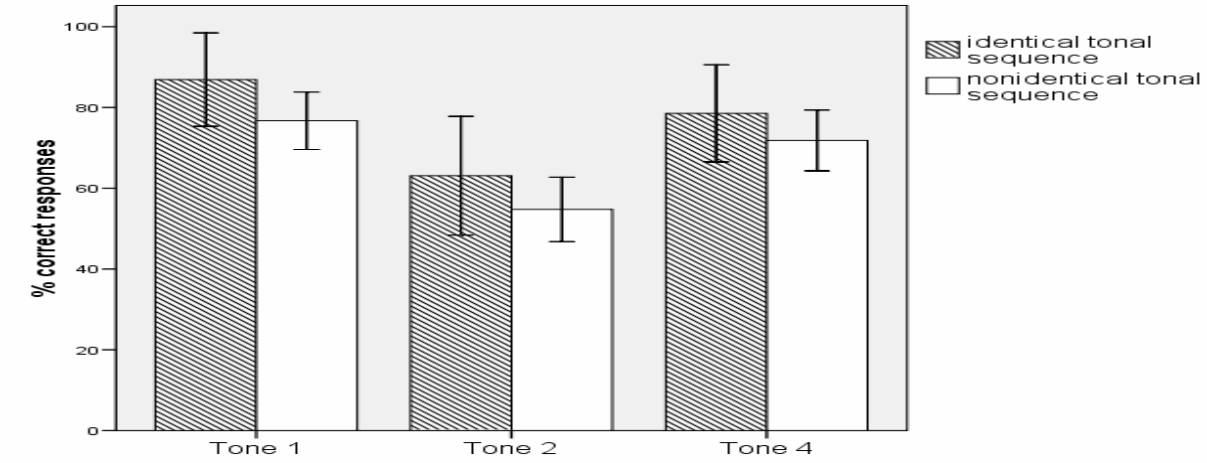


Figure 4-4. Mean percent correct perception for (A) Inexperienced and (B) Experienced groups for tones in identical tone sequence vs nonidentical tone sequence. Error bars indicate standard error.

The analyses showed that the inexperienced group (Figure 4-3 A) only made more accurate identifications under identical tonal sequences than under nonidentical tonal sequences for Tone 1 (85.79% & 59.01%) [$t(1, 109) = 3.342$, $p = .001$]. However, inexperienced listeners made more errors under identical tonal sequence for Tone 2

(45.14% & 54.88%) and made almost equally accurate judgments in both tonal sequences for Tone 4 (64.32% & 64.26%). However, the difference of accuracy rates due to the tonal sequences for Tone 2 [t (1, 108) = -1.174, p=.243] and Tone 4 [t(1, 108) = -.008, p=.993] were not significant.

Although the experienced group (Figure 4-3 B) showed a tendency to perceive tones more accurately in identical tonal sequence than nonidentical tonal sequence (Tone 1, 86.89% vs. 76.72%; Tone 2, 63.11% vs. 54.76%; Tone 4, 78.54% vs. 71.83%), none of the tones' difference in accuracy rate for the two tonal sequences was significant: Tone 1 [t(1, 108) =1.454, p=.149], Tone 2 [t(1, 108) =1.030, p=.305], or Tone 4 [t(1, 108) =.906, p=.367.

Effects of tonal context

Figures 4-5 A and B show how inexperienced and experienced groups of American learners identify Mandarin tones under different tonal contexts. Overall, the accuracy rate of identifying tones in the compatible tonal context (62.5%) is higher than the accuracy rate for the conflicting tonal context (53.87%). The three -way repeated measures ANOVA with Tonal Context (compatible, conflicting) and Tone (Tone 1, Tone 2, Tone 3, Tone 4) as the within-subject factors and Group (inexperienced, experienced) as the between-subject factor showed that there were significant main effects of Tonal Context [F(1, 54) =13.655 , p=.001], Group [F(1, 54) =9.308 , p=.004], and Tone [F(3, 52) =16.379, p< .001]. No significant interaction was found.

Even though no significant interaction among the factors was found, further analyses were conducted to investigate the effect of learning experience. It was found that inexperienced groups of listeners tend to perceive Tone 1 (72.07% vs. 58.88%), Tone 2 (58.34% vs 45.76%) and Tone 4 (66.09% vs. 61.36%) more accurately in the

A)

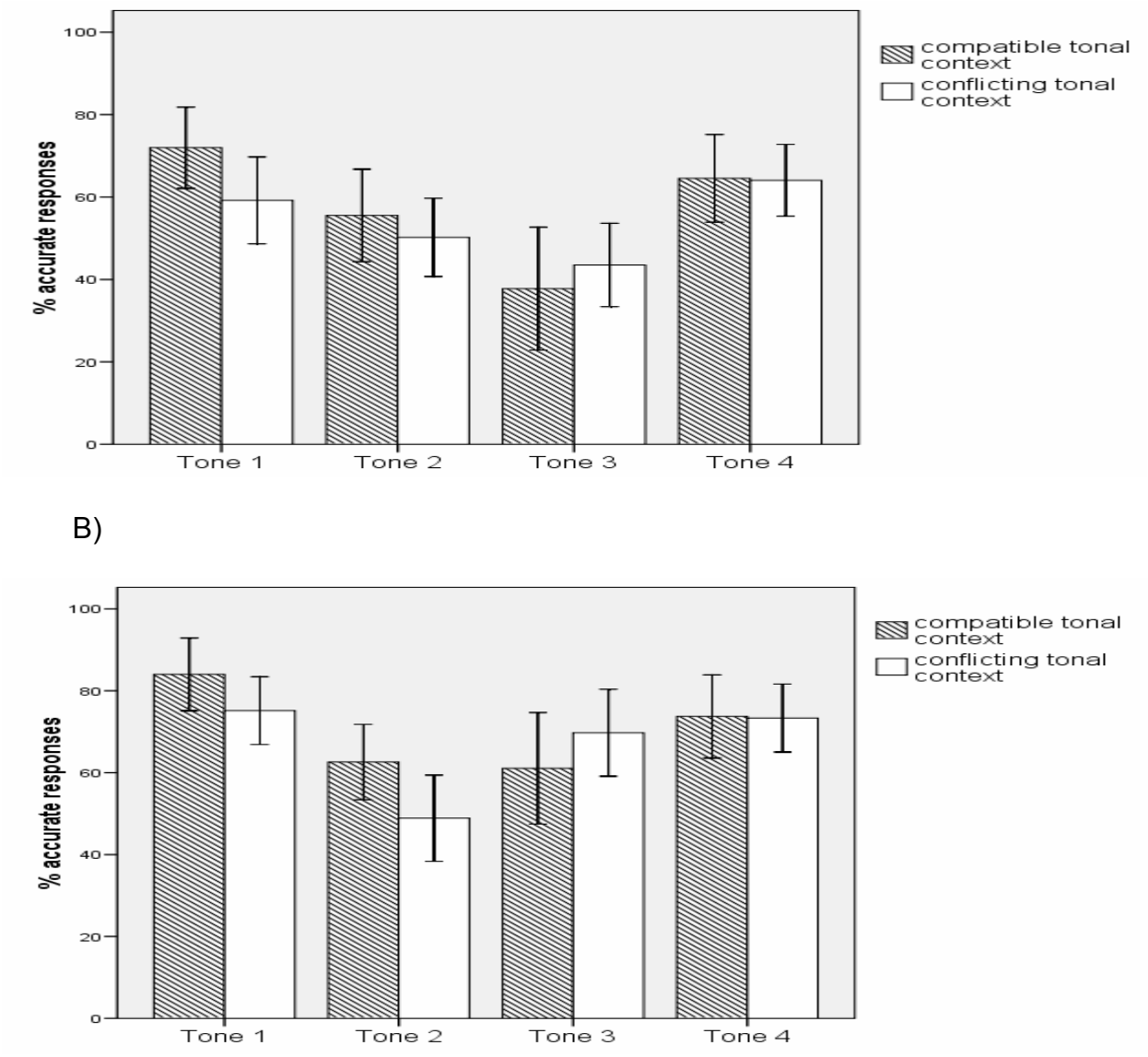


Figure 4-5. Mean percent correct perception for (A) Inexperienced and (B) Experienced groups for tones in compatible vs. conflicting tonal context. Error bars indicate standard error.

compatible tonal context than in the conflicting tonal context and Tone 3 (38.07% vs. 43.46%) less accurately in the compatible tonal context. Furthermore, t-tests indicated that the differences between the accurate perception rates of Tone 1 [$t(1, 110) = 1.844$, $p = .068$] and Tone 2 [$t(1, 110) = 1.766$, $p = .080$] which occurred in different tonal contexts were only marginally significant and the differences between the accurate perception

rates of Tone 3 [$t(1, 82) = -.584$, $p = .561$]; and Tone 4 [$t(1, 110) = 705$, $p = .482$] which occurred in different tonal contexts were not significant.

The same results were found for the experienced listeners. Although experienced groups of listeners tend to perceive Tone 1 (84.52% vs. 73.84%), Tone 2 (61.93% vs 51.77%) and Tone 4 (73.80% vs. 73.21%) more accurately in the compatible tonal context than in the conflicting tonal context and Tone 3 (64.25% vs. 81.59%) less accurately in the compatible tonal context, t-tests indicated that the differences between the accurate perception rate of tones which occurred in different tonal contexts was marginally significant on Tone 1, [$t(1, 109) = .1.781$, $p = .078$] but not significant on Tone 2 [$t(1, 110) = 1.455$, $p = .149$]; Tone 3 [$t(1, 82) = -.403$, $p < .688$]; or Tone 4 [$t(1, 110) = .092$, $p = .927$]).

In terms of change of pitch direction, in the literature review, I explained that both carry-over and anticipatory influences change the pitch ranges but not the direction of tones, except Tone 1 after Tones 3 and 4 Tone 1 rises instead of staying flat (Xu, 1997). However, this kind of modification driven by tonal context does not change the perceptual tonal category to native speakers of Mandarin.

In Figure 4-6, the accuracy rate of identifying final syllable Tone1 occurring after Tone 1 and Tone 2 was compared with the accuracy rate of identifying final syllable Tone1 occurring after Tone 3 and Tone 4. In the former tonal environment, the direction of Tone 1 was maintained, whereas in the latter tonal environment, the direction of Tone 1 was changed from flat to rising. It was shown that Tone 1 with unchanged direction was identified better than Tone 1 with changed direction (75.607% vs. 61.296%). A two-way repeated ANOVA, with Context (unchanged pitch direction and changed pitch

direction) as within group factor and Group (inexperienced group and experienced group) as between subject factor, revealed a significant main effect of Context [*F* (1, 54) =8.715, p<.01]. However, there was no significant main effect of Group [*F* (1, 54) =2.52, p=.118]. There is no interaction between Context and Group [*F* (1, 54) =1.548, p=.219]. These results suggested that both groups were more accurate at identifying Tone 1 under the unchanged pitch direction condition (i.e., after another Tone 1 or after Tone 2) than under the changed pitch direction condition.

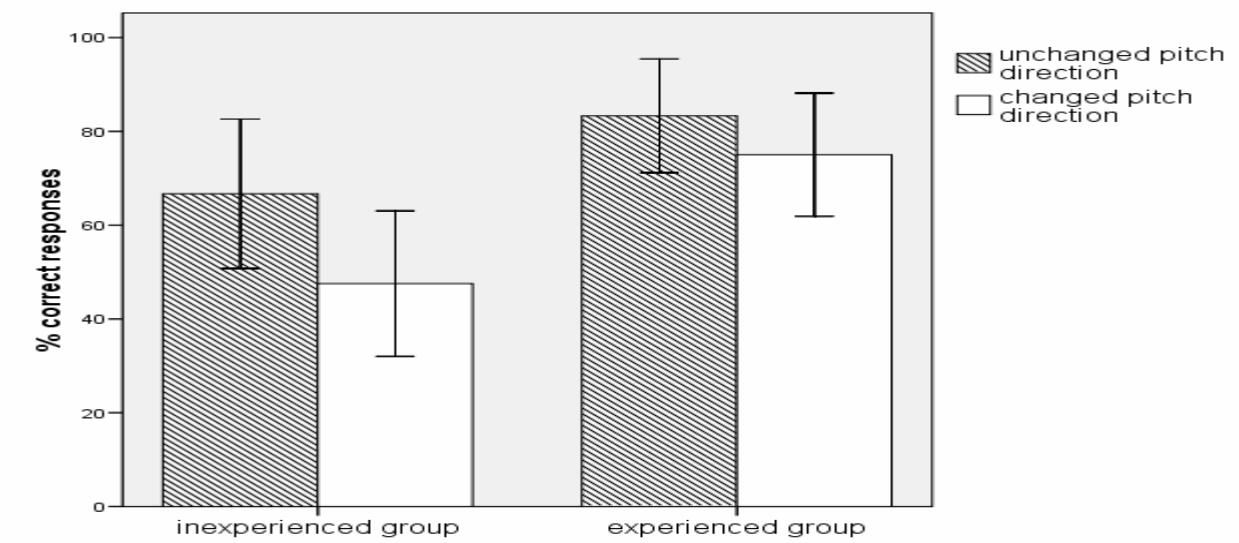
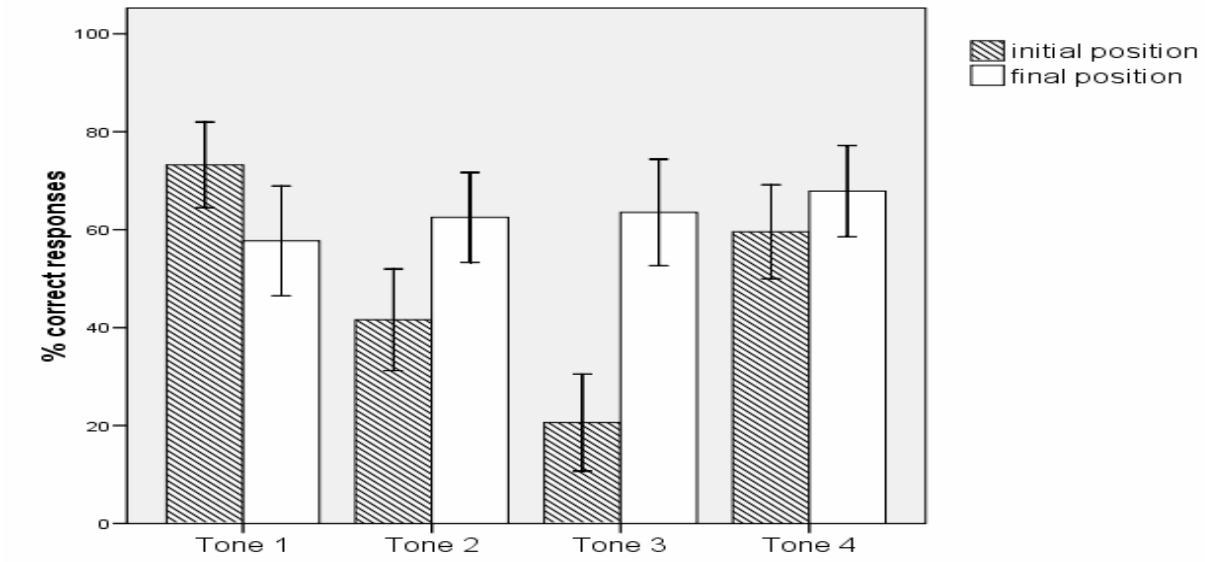


Figure 4-6. Mean percent correct identification of coarticulated final syllable Tone 1 in different tonal contexts by two groups of learners of Mandarin. Error bars indicate standard error.

Effects of syllable position

Figures 4-7 A and B show how inexperienced and experienced groups of American learners identify Mandarin tones in different syllable positions. Overall, accuracy rate for the initial syllable position (59.90% accurate responses) was lower than the accuracy rate for the final syllable position (69.05% accurate responses). The three- way repeated ANOVA yielded significant main effects of Syllable position [F (1, 82) = 33.159, p< .001], Group [F (1, 82) = 22.865, p< .001],and Tone [F (3, 80) =

A)



B)

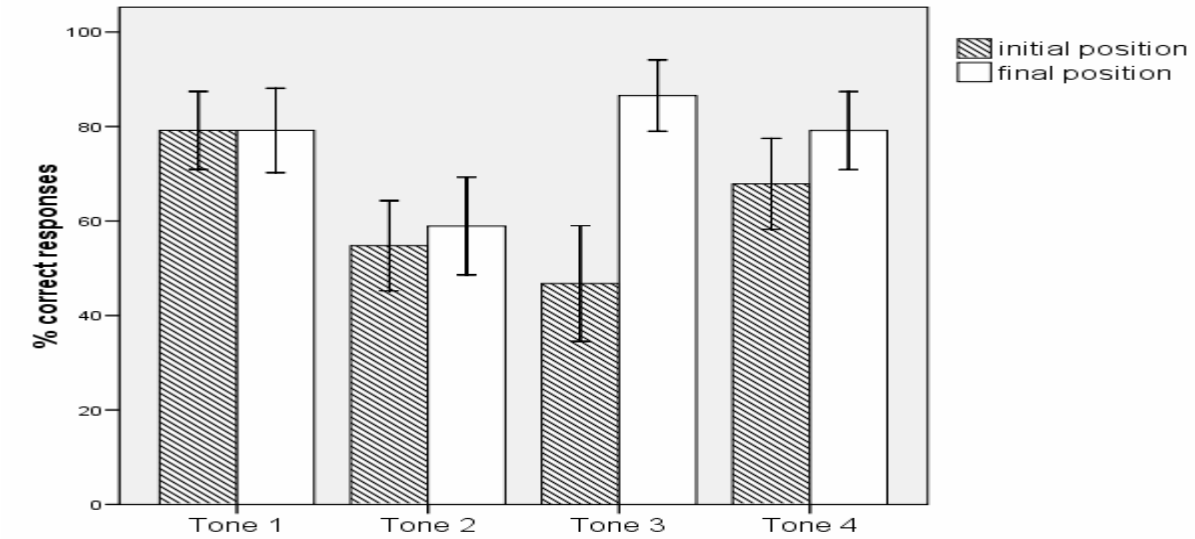


Figure 4-7. Mean percent correct perception for (A) Inexperienced and (B) Experienced groups for tones in initial vs. final syllable position. Error bars indicate standard error.

16.277, p< .001]. There was also a significant interaction of Tone X Syllable Position [F (3, 80) = 18.613, p< .000], a marginally significant interaction of Tone X Group [F (3, 80) = 2.611, p= .057], and a significant Syllable position X Group X Tone interaction [F (3, 80) = 2.984, p< .05].

Inexperienced learners identified Tone 2 (41.61% initial vs. 62.52% final), Tone 3 (20.25% initial vs. 63.52% final), and Tone 4 (59.57% initial vs. 67.88% final) better in final syllable position, whereas they identified Tone 1 (73.23% initial vs. 57.71% final) better in initial syllable position. However, results of T-test analysese showed that the difference between accuracy percentages as a function of syllable position was significant for Tone 1 [t (1,110)= 2.182, p<.04], Tone 2 [t (1, 110)=-3.019, p<.004] and Tone 3 [t (1, 82)=-5.836, p<.001], but not significant for Tone 4 [t (1, 110)=-1.243, p=.217].

Experienced learners identified Tone 2 (54.77% initial vs. 58.93% final), Tone 3 (46.76% initial vs. 86.55% final), and Tone 4 (67.88% initial vs. 79.14% final) better in final syllable position, whereas they identified Tone 1 at almost the same rate (79.18% & 79.18%)[2] in both syllable positions. T-testresults, however, showed that the difference between accuracy percentages as a function of syllable position was highly significant for Tone 3 [*t* (1, 82) =-5.524, p< .001] and marginally for Tone 4 [*t* (1, 110) =-1.777, p=.078], but not significant for Tone 1 [t (110) =.000, p=1] or Tone 2 [*t* (1, 110) =-.591, p=0.556].

Error type

Confusion matrices on identifying coarticulated tones in different syllable positions are shown in tables 4-2 to 4-5.The data was based on 168 responses for Tone 1, Tone 2, Tone 4 (12 stimuli * 14 participants), and 126 responses for Tone 3 (9 stimuli * 14

[2] Though the average mean of identification accuracy rate of initial Tone 1 is the same as the average mean of identification accuracy rate of final Tone 1, the standard deviations of these two accuracy rates are different (30.903 & 33.403).

participants). The results of Tone 3 in the tonal combination of Tone 3 +Tone 3 were not included.

As shown in the below tables, three kinds of error types were found: tonal direction misperception and tonal height misperception. The definition of each error type is as following:

Tonal direction misperception: the direction of the target coarticulated tone is different from the identified tonal category.

Tonal height misperception: the direction of the target coarticulated tone is the same as the identified tonal category, but the tonal height is different.

Table 4-2. Confusion matrix for tones in initial position identified by inexperienced learners, with the percentages in parentheses

Tone	Identified as			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	**133 (79.17)**	14(8.33)	0	21(12.5)
Tone 2	24 (14.3)	**92 (56.7)**	21(12.5)	31 (18.5)
Tone 3	12 (8.8)	22 (16.2)	**59 (43.4)**	43 (31.6)
Tone 4	20 (11.9)	14 (8.3)	20 (11.9)	**114 (67.9)**

Table 4-3. Confusion matrix for tones in final position identified by inexperienced learners, with the percentages in parentheses

Tone	Identified as			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	**134** (79.8)	*23* (13.7)	3 (1.8)	8 (4.8)
Tone 2	9 (5.4)	**99** (58.9)	*55* (32.7)	5 (3)
Tone 3	0	17(13.5)	**109** (86.5)	0
Tone 4	9 (5.4)	14 (8.3)	12 (7.1)	**133** (79.2)

Table 4-4. Confusion matrix for tones in final position identified by experienced learners, with the percentages in parentheses

Token	Identified as			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	**97** (57.7)	56 (33.3)	3 (1.8)	12 (7.2)
Tone 2	21 (12.5)	**103** (61.3)	37(22.0)	7 (4.2)
Tone 3	8 (6.3)	29 (23.0)	**80** (63.5)	9 (7.2)
Tone 4	12 (7.2)	30 (17.9)	13 (7.7)	**113** (79.2)

Table 4-5. Confusion matrix for tones in final position identified by experienced learners, with the percentages in parentheses

Token	Identified as Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	**123** (73.2)	17 (10.1)	0	28 (16.7)
Tone 2	50 (29.8)	**70** (41.7)	15 (8.9)	33 (19.6)
Tone 3	29 (23)	31(24.6)	**26** (20.6)	40 (31.7)
Tone 4	33 (19.6)	16 (9.5)	18 (14.3)	**101** (60.1)

The tonal category is defined in the Table 4-6 and how error types were categorized are shown in Table 4-7. The following paragraphs explained how the errors were categorized in details.

1) Syllable initial T1 is a flat tone, whereas the category of T2 is a rising tone and T4 is a falling tone. When syllable initial T1 is perceived as T1, T2 or T3, the error was labeled as a direction misperception.

2) Syllable initial Tone 2 is a rising tone, whereas T1 is a flat tone, and T4 is a falling tone. When syllable initial T2 is perceived as T1 or T4, (I labeled) the error as a direction misperception.

Syllable initial T3 phonetically moves in the same direction as T2. However, T3 starts lower and ends lower than T2, and turns later than T2. Therefore, T3 misperceived as T2 was labeled as a height misperception.

3) Syllable initial Tone 3 is a falling tone, whereas T1 is a flat tone, T2 is a rising tone, and T4 is a falling tone. When syllable initial T3 was perceived as T1 or T2, (I labeled) the error as a direction misperception.

Syllable initial T3 moves in the same direction as T4. However, T3 is a low falling tone, whereas T4 is a high falling tone. When T3 was misperceived as T4, (I labeled) the error as a height misperception.

4) When syllable initial T4 was perceived as T1, T2 or T3, the error was a direction misperception.

5) Syllable final T1 shows a rising direction when it occurs after T3 or T4, but the degree of rising is not as deep as Tone 2. Therefore, when T1 was identified as T2, (I labeled) the error as a height misperception.

When T1 was perceived as T3 or T4 by American learners, the misperception was categorized as a direction misperception.

6) Syllable final T2 presents a rise, whereas T1 has no direction change and T4 falls. Therefore, T2 misperceived as T1 or T4 was labeled as a direction error.

Similar to syllable final T2, syllable final T3 rises. However, syllable final T3 does not rise as high as T2. Therefore, syllable final T2 perceived as T3 is due to a confusion of tonal height.

7). Syllable final T3 perceived as T1 or T4 was labeled a direction error.

Syllable final T3 perceived as T2 was labeled a tonal height error.

8) Syllable final T4 is a falling tone, unlike T1, T2, or T3. Therefore, syllable final T4 perceived as T1, T2 and T3 was labeled as a tonal directional error.

Table 4-6. Description of Mandarin Chinese tonal categories

Tonal category	Phonological description	Phonetic description
T1	Flat	High flat
T2	Rising	Falling-rising (low onsite and high offsite) with a early turning point
T3	Falling-rising	Falling-rising (low onsite and low offsite) with a late turning point
T4	Falling	High falling

Table 4-7. Tonal confusions categorized into error types

Target coarticulated tone	Tonal direction error	Tonal height error
1) Initial high flat T1	T2 T3 T4	
2) Initial rising T2 (acoustically falling-rising with a early turning point)	T1 T4	T3
3) Initial low falling T3 (effect of tonal context)	T1 T2	T4
4) Initial high falling T4	T1 T2 T3	
5) Final rising T1 (acoustically rising after T3 or T4, effect of tonal context)	T3 T4	T2
6) Final rising T2 (acoustically falling-rising with a early turning point)	T1 T4	T3
7) Final falling-rising T3	T1 T4	T2
8) Final falling T4	T1 T2, T3	

Figure 4-8 shows that inexperienced learners tended to incur more directional misperception errors than experienced learners. This difference was non significant for initial syllable Tone 1 [χ^2=2.404, p=.121] and initial syllable Tone 4 [χ^2=2.183, p=.140]. However, it was marginally significant for final syllable Tone 1 [χ^2=3.630, p=.057] and significant for initial syllable Tone 2 [χ^2=4.325, p=.038], final syllable Tone 2 [χ^2=13.105, p=.000], initial syllable Tone 3 [χ^2=11.470, p=.001], and final syllable Tone 4 [χ^2=9.002, p=.003] Experienced learners did not make a single directional error when identifying final syllable Tone 3, whereas inexperienced learners misperceived the rising direction of final syllable Tone 3 17 times.

Figure 4-9 shows that experienced learners made fewer tonal height misperception errors than inexperienced learners when they identified final syllable Tone 1 [χ^2=12.229, p=.000] and final syllable Tone 3 [χ^2=3.829, p=.05] . However, on

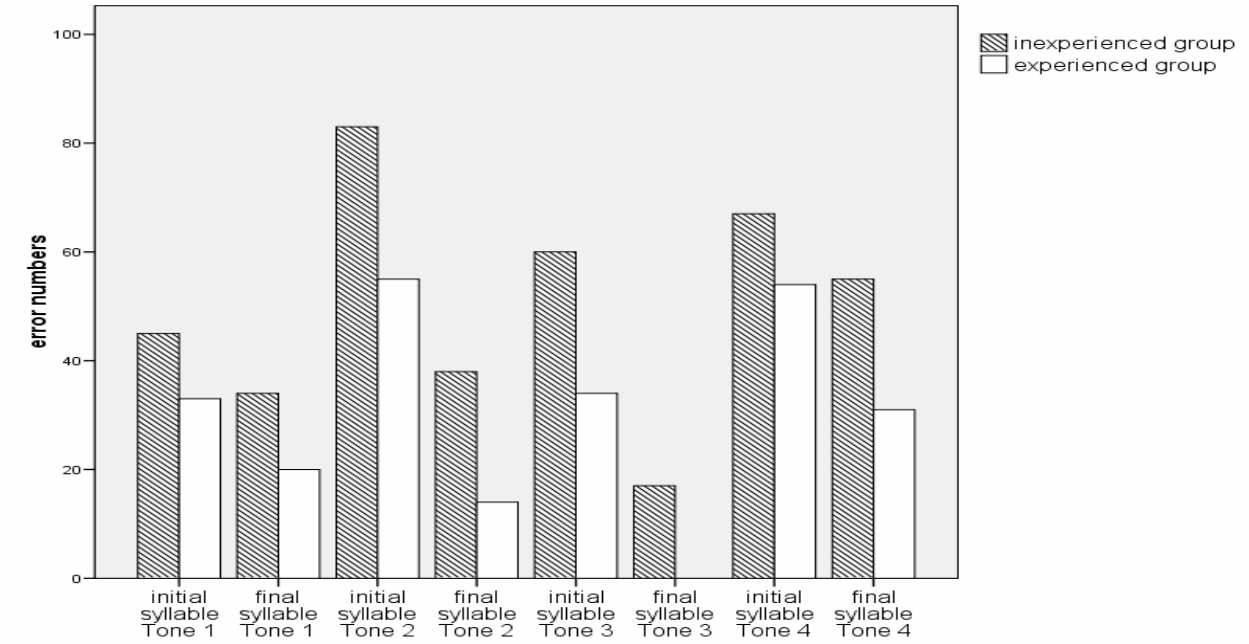


Figure 4-8. Frequency of tonal directional misidentification by two groups of American learners

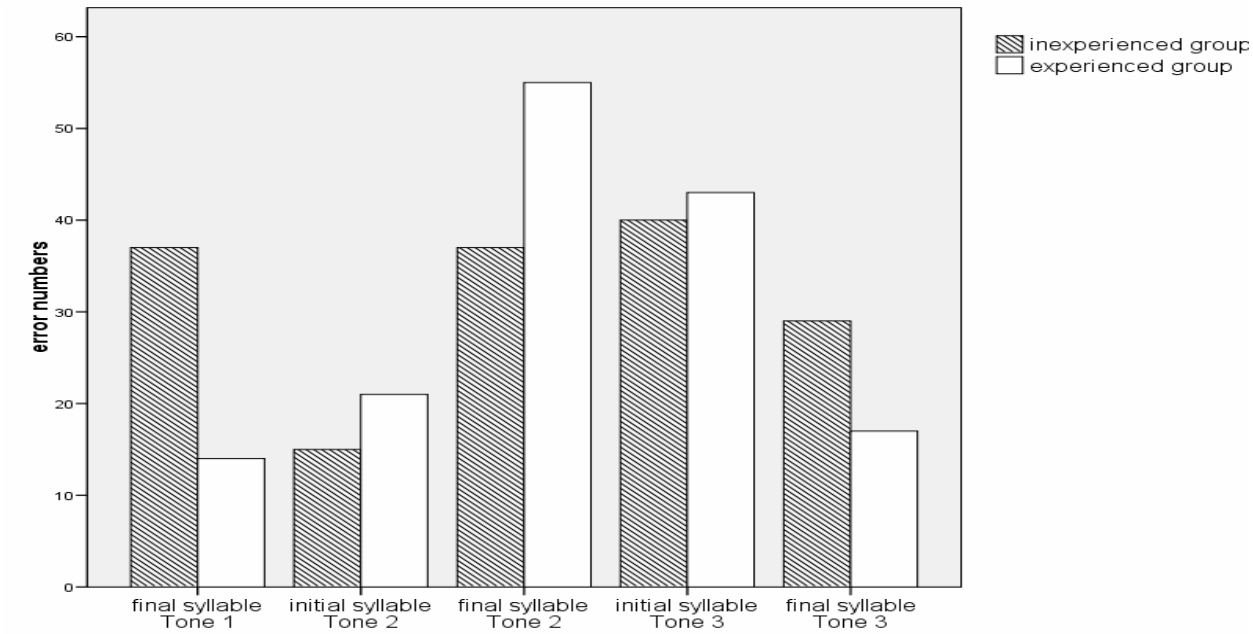


Figure 4-9. Frequency of tonal height misidentification by two groups of American learners

certain tones the experienced learners made more height perception errors than the inexperienced group, such as initial syllable Tone 2 [χ^2=1.12, p=.29], initial syllable Tone

3 [χ^2=.162, p=.688] and final syllable Tone2 [χ^2=4.850, p=.028]. However, the differences between the frequencies of misidentifying tonal height of initial syllable Tone 1 and Initial syllable Tone 3 made by two groups of learners are not significant according to chi-square tests.

Discussion

This study was designed to investigate how coarticulated Mandarin Chinese tones in disyllabic words are identified by American learners with different levels of language experience (inexperienced learners vs. experienced learners). The purpose of the perception experiment was to test whether American learners' identification of coarticulated Mandarin Chinese tones improves with language learning experience, whether the syllable position affects identification accuracy, and whether tonal context affects identification types.

Effects of Learning Experience

The results show that experienced American learners of Mandarin performed better than inexperienced learners on identifying coarticulated tones in disyllabic words, and this advantage was consistent across all four Mandarin tones. The results also show that experienced learners' performance was less influenced by the linguistic factors, such as tonal environment and syllable position. In terms of tonal environment, inexperienced learners' performance on identifying coarticulated tones was more affected by the tonal environment than experienced learners. Inexperienced learners identified Tone1 better when it occurred next to an identical tone and in the unchanged tonal direction environment than next to a nonidentical tone or in the changed tonal direction environment, whereas experienced learners did not perform statistically differently in these various environments. Regarding the effect of syllable position,

inexperienced learners showed a stronger tendency than experienced learners to perceive final syllable tone more accurately than initial syllable tone in more tonal categories.

Therefore, we can say that learning experience makes American speakers more sensitive to pitch change across tonal categories in coarticulated tonal environments but less sensitive to pitch change within tonal categories triggered by tonal environments. In other words, more exposure to the Chinese language along with learning experience helps American speakers to improve their ability to perceive tonal features from phonetic cues to phonological categories. For example, inexperienced American learners misperceive Tone 1 (as Tone 2) more frequently when it is slightly rising due to the tonal environment, but this small degree of rising does not cause Tone 1 to change into Tone 2 in native Mandarin speakers' perception. The fact that experienced American learners are more accurate than inexperienced learners at identifying Tone 1 when its pitch direction was changed due to tonal context suggested that with a longer experience with Mandarin, American learners are better at ignoring within-category differences (induced by phonetic contexts) and focus more on across-category differences. In other words, in comparison to inexperienced learners, they were able to ignore the rise in pitch at the onset of Tone 1 when it occurs after Tone 3 and Tone 4, and focus instead on the rest of the pitch contour to arrive at a constant percept on Tone 1.

In addition, learners with more learning experience may also be less influenced by their L1 background than learners with less learning experience. Although English is not a tonal language, it does use intonation to signal pragmatic meanings, such as rising

intonation for interrogative. The negative L1 influence is more often transferred by inexperienced American learners, and it makes inexperienced American learners perceive final syllable Tone 2 better than initial syllable Tone 2. The advantages that experienced learners have can be attributed to the increased L2 input that comes along with their longer learning experience. This idea is supported by Wayland et al. (2003), which reported that in discriminating the Thai middle and low tone contrast, experienced learners of Thai outperformed inexperienced learners who did not have any learning experience in Thai. Previous studies showed that not only in the classroom but also in the laboratory, more L2 perception input seems to lead to greater ability to perceive L2 tonal features, such as Mandarin and American listeners discriminating the mid tone and low tone contrast in Thai (Wayland & Guion, 2004), American listeners identifying Mandarin tones (Wang et al., 1999), and Mandarin listeners and American listeners identifying Cantonese tones (Francis et al., 2008). Extended to L2 input in general, listeners with more L2 perception experience have more advantages to perceive L2 tonal features correctly compared with inexperienced listeners.

Effects of Tonal Condition

In the present study, three types of tonal environments were tested. Two types of tonal condition were tested. The first type of tonal condition is called tones in isolation and the second type of tonal condition is called tones in coarticulation. It found that tones in isolation is identified better than tones in coarticulation and this kind of superior cross over four tones. The fact can be possibly attributed to the effect of attention distribution. It is assumed that the attention on a tone in coarticulation was less than it was in isolation.

Effects of Tonal Sequence

The type of tonal environment is defined by considering the tonal sequence. There are two types of tonal sequence: identical tonal sequence vs. nonidentical tonal sequence. Identical tonal sequence was assumed to be a positive tonal environment for American learners to identify tones, whereas nonidentical tonal sequence was assumed to be a negative tonal environment. It was assumed that identical tonal sequence helps learners to identify tones by showing tonal repetition. This was not confirmed by the results. There was no main effect of tonal sequence. However, if we look at the effect of the factor on a specific tone, only Tone 1 was identified with a significantly higher accuracy rate in identical tonal sequence than in nonidentical tonal sequence. The difference between the accuracy rates of Tone 2 and Tone 4 in two tonal sequences does not reach the significant level. The tonal combination of Tone 1 + Tone 1 is the only tonal combination in two positive tonal environments, which is identical tonal sequence and compatible tonal context, whereas Tone 2 and Tone 4 in any tonal combination are with one positive and one negative environment. Therefore, Tone 1 in the identical tonal sequence is identified better, but Tone 2 and Tone 4 in identical tonal sequence are not identified better due to the negative interaction with the tonal context that they occur in.

Effects of Tonal Context

The type of tonal environment is categorized by the relation between the offset of preceding tone and the onset of following tone (compatible and conflicting tonal context). Since a conflicting tonal context reduces the phonetic pitch range of the tones, identification might be more difficult in this type of context. Therefore, the compatible

tonal context was predicted to improve American learners' tone identification, while the conflicting context was expected to inhibit it. This study showed American learners was affected by the linguistic factor of tonal context.

For final syllable Tone 1, the tonal context changes not only the pitch but also the direction of the tone. Final syllable Tone 1 turns into a rising tone when it occurs in a conflicting tonal context. As expected, final syllable Tone 1 in conflicting context was identified less accurately the same tone in a compatible context.

Effects of Syllable Position

Syllable position affected American learners' perception of coarticulated tones. Tones in final syllable position were identified better than tones in initial syllable position. More specifically, coarticulated Tone 2 and Tone 4 were identified better in final than initial position. Here, it might be due to L1 perception transfer. There is no lexical tone in English. However, English does use intonation to differentiate statements and questions. A falling intonation makes a statement, and a rising intonation indicates a yes-no question. To perceive the pragmatic function, English listeners pay more attention to the pitch direction of sentence final intonation. Therefore, the American learners in the present study might transfer this kind of attention to perception of tones in disyllabic words. In other words, the American learners pay more attention to the tonal direction of the word final tone than the word initial tone. Therefore, the accuracy rate of identifying Tone 2 and Tone 4 is higher in word final position than in word initial position.

The effect of syllable position on identifying Tone 3 is significant for both groups of learners. Both of groups of learners indentified Tone 3 better in final syllable position.

Tone 3 changes from a falling-rising tone to a low falling tone when it occurs on initial syllable position, whereas the tonal direction of Tone 3 on final syllable position remains as the same as it is in isolation. The changed tonal direction of initial Tone 3 might cause the difficulty for both groups of learners to identify the tone.

The above results may be also due to a recency effect; final tones are remembered better than initial tones. Therefore, final syllable tones are misperceived less frequently than initial syllable tones.

We also noticed that the duration of the tone on the final syllable is longer than the one on the initial syllable. The shape of the tone on the final syllable might be more fully presented, and the fully presented tone might be recognized more easily. Therefore, the accuracy rate for the final tone is higher than the one for the initial tone.

The following argument is speculation and needs to be tested in future studies. In certain contexts, Tone 3 changes from a falling-rising tone to a falling tone, whereas Tone 1 changes from a high flat tone to a rising tone. Tone 3 changes from a complex tone to a simple tone, but Tone 1 changes from one simple tone to another. Therefore, the change of Tone 3 is more dramatic in terms of tonal types and it is more difficult to generalize to the same tonal category. Therefore, even with more learning experience, American learners still had a hard time identifying coarticulated Tone 3 in initial syllable position.

Development of Perception of Coarticulated Tones

Two types of errors were found in both groups: tonal direction misperception and tonal height misperception. Experienced learners committed fewer tonal directional errors than inexperienced learners. However, experienced learners still made a

considerable amount of tonal height errors. Therefore, the ability to identify tonal direction may improve faster than the ability to identify tonal height among English speakers. Two reasons could explain this phenomenon. The first is the influence of the language teaching method. In Chinese language classrooms, instructors focus on introducing the differences in tonal direction of Mandarin tones in isolation but not on the differences of their tonal height. They only drill their students in the classroom on identifying tonal direction. Therefore, the more time American learners spend studying in the classroom, the more practice they get in identifying tonal direction. The practice experience improves experienced learners' ability to identify tonal direction. Meanwhile, the lack of practice in identifying tonal height keeps learners from improving this skill.

The second potential reason for this unparallel development is the tonal environment. Within a disyllabic word, the tonal direction of one tone can be implied by the tonal direction of its neighboring tone. For example, the tonal directions of Tone 4 and Tone 2 in the tonal combination Tone 4 +Tone 2 are implied by the opposite directions of tones in the tonal combination. Listeners can notice that the pitch changes of the two contour tones are different from each other. However, the height of one tone is different compared with its neighboring tone. For example, Tone 3 in the tonal combination Tone 3 + Tone 2 is a low falling tone. Its falling direction is more noticeable than its low range in the tonal direction of falling and rising. Therefore, tonal height is harder to identify in a disyllabic tonal environment even for experienced American learners.

Conclusion

Experienced learners identified coarticulated tones with higher accuracy rate than inexperienced learners. Moreover, experienced learners' tonal identification ability was more influenced by tonal coarticulation and tonal repetition and less influenced by tonal context and syllable position than inexperienced learners' ability was. In other words, with increasing learning experience, American learners' ability to categorically perceive coarticulated Mandarin tones improves. The results indicate that tonal environments and syllable position play important roles in identifying coarticulated tones. This implies that in future models of tone perception, the surrounding tonal environment needs to be considered to account for identification accuracy of lexical tones by L2 learners. Furthermore, language learning experience improves American learners' ability to identify Mandarin tone direction more than their ability to identify tone height.

CHAPTER 5
PRODUCTION OF COARTICULATED MANDARIN TONES

This chapter reports on a study of the production of coarticulated Mandarin tones by American learners of Mandarin with different learning experience. The first section presents the research questions and hypotheses. The second section presents the methodology, including participants, stimuli, procedure and data analysis. The third section reports the data and statistical analysis, and the fourth section discusses the results.

Research Questions and Hypothesis

In the present experiment, production of Mandarin coarticulated tones in disyllabic words by inexperienced and experienced American learners was investigated. The experiment was aimed at answering the following questions:

Research question 1. Does American learners' production of Mandarin coarticulated tones improve with Mandarin learning experience?

Experienced learners were expected to produce Mandarin coarticulated tones better than inexperienced learners. In other words, experienced learners were expected to pronounce these tones with a higher accuracy rate.

Research question 2. What are the typical tonal production errors for American learners with different amounts of language learning experience?

Recall that Wang et al. (2003) categorized two types of errors American learners make when producing *isolated* Mandarin tones: tonal direction confusion and tonal height confusion. They also found that the tonal height confusion was more resistant to improvement. In the present study, these two types of errors were expected to occur in

American learners' production of coarticulated Mandarin tones in disyllabic words, and the tonal height confusion was expected to be harder for learners to overcome.

Research question 3. What are the linguistic factors that affect American learners' ability to produce coarticulated Mandarin tones accurately?

As with perception, two linguistic factors were expected to influence American learners' production of coarticulated Mandarin tones. One is syllable position (initial vs. final syllable position). There was no prediction regarding which tones on which syllable would be produced better. The other factor is tonal environment, including tonal condition (isolation vs. coarticulation), tonal sequence (identical vs. nonidentical tone sequence) and tonal context (compatible vs. conflicting tonal context).

Tones in monosyllabic word were labeled as isolated tones and tones in disyllabic words were labeled as coarticulated tones. Considering the effect of coarticulated environment, tones in coarticulation were expected to be produced with a lower accuracy rate than those in isolation.

Since all the participants had already learned to produce isolated Mandarin tones, they were expected to produce these four tones more easily in the identical tonal sequence by simply repeating the tones in sequence. Therefore, tones in identical tone environment were expected to be produced with a higher accuracy rate than those in nonidentical tone environment.

The concept of compatible and conflicting tonal contexts was introduced in the previous chapter. The tones in compatible tonal context are connected smoothly, whereas there is a pitch gap between the tones in conflict. American learners were expected to feel more comfortable producing coarticulated tones in a tonal environment

without a tonal gap. Therefore, tones in the compatible tonal context were expected to be produced with a higher accuracy rate than those in the conflicting tonal context. Table 4-1 categorized the possible Mandarin tonal combinations on disyllabic words into four tonal environments.

Method

Participants

The participants were selected from the pool of participants in the study of isolated tone production. Nine inexperienced American learners and nine experienced American learners participated in the study. These 18 participant (8 female and 10 male) were all part of the group of 28 who completed the coarticulation perception task. All of them attained at least a 67% accuracy rate of both perceiving and producing isolated tones in the first experiment. This step was designed to reduce the possibility that the participants' inaccurate production of coarticulated tones was due to their inability to perceive or produce Mandarin tones in isolation.

Stimuli

The study tested 48 non words. This design was intended to prevent participants from using their tonal knowledge of words with lexical meanings to produce tones. As with the perception stimuli, the target words all had the structure /na.na/. The syllable structure was chosen to avoid the possibility of misproducing tones by placing English stress on one syllable, since English speakers tend to stress CVN over CV syllable structure. There are three reasons to choose nasal /n/ as the onset consonant in the stimuli: 1) the pitch value of the vowel is not changed due to the consonant. Nasal consonant /n/ does not change the F_0 the following vowel. 2) There is no broken pitch

track within one syllable. Nasal /n/ keeps the pitch track consistent. 3) Due to the character of nasals in the spectrogram, there is a very clear boundary between the two syllables.

Procedure

All stimuli in the production experiment were presented in random order, and each possible tone combination was presented three times. The stimuli were divided into three blocks. Between each block, the participant was required to rest for one minute. This design ensured that participants' effort on producing each token was approximately equal. For each stimulus, the participant viewed one word in Pinyin on the computer screen. The same word was presented again after the participant finished their first try. The second try on the pronunciation was counted. This step was intended to ensure that the participants' tonal errors were not due to their unfamiliarity with the stimuli. In order to ensure that the participant produced coarticulated tones on disyllabic words, no pause was allowed in their production between the two syllables of each word. The 48 stimuli (The production stimuli include T3 + T3, even though the perception stimuli didn't) were preceded by five practice stimuli that were not analyzed. The productions were collected using a solid state recorder (PMD660/U3B) with a professional microphone (Audio-Techinca AT4041 Car). Before the real test, there was three warm-up stimuli, and the results from the three warm-up stimuli were not analyzed.

Judgment

Two native speakers of Mandarin with a Beijing accent transcribed the tones produced by participants based on their perception. Answer sheets were provided. The produced stimuli were printed with no tonal diacritics. Judges needed to provide a tonal

diacritic corresponding to the tone they heard (Wang 2003). The program PRAAT was also used to generate pitch contours to avoid the misperception of the speakers by their categorical perception. It was observed during the rating that the Mandarin speakers tended to perceive all the American learners' low tones in initial syllable position, whether low flat tone or low falling tone, as Tone 3.

Results and Analysis

In this section, the results will be presented based on two measures of learner knowledge: accuracy rate and error type.

Accuracy Rate

Three factors were considered in the study: language learning experience (inexperienced vs. experienced), syllable position (initial position vs. final position), and tonal environment (isolated tonal condition vs. coarticulated tonal condition, identical vs. nonidentical tonal sequence, compatible vs. conflicting tonal context).

Effects of language learning experience

The overall results in Figure 5-1 show that the experienced learners of Mandarin performed better than inexperienced learners on producing coarticulated Mandarin tones (54.81% inexperienced vs. 80.38% experienced). This difference is highly statistically significant [t (1, 538) = -7.876, p<.001].

As shown in Figure 5-2, overall, tonal production by experienced American learners of Mandarin were judged to be more accurate by native listeners than that of inexperienced learners (79.17% vs. 52.92%), main effect of Group [F (1,106) = 54.93, p<.001]. In addition, production accuracy varied significantly across the four tones, main effect of Tone [F (3,104) = 44.999, p<.001], Post hoc analyses (Bonferroni adjusted p<.05) showed that Tone 1 production was significantly more accurate than all other

tones. In addition, Tone 2 and Tone 4 were significantly more accurate than Tone 3. Production accuracy of Tone 2 and Tone 4 were, however, comparable.

A significant interaction between Group and Tone [F (1,104) = 3.702, p=.014] was also obtained. This was due mainly to the fact that experienced learners were more accurate than inexperienced learners in their production of Tone 1 (p<.001), Tone 3 (p<.001) and Tone 4 (p<.001), but not Tone 2 (p=.210).

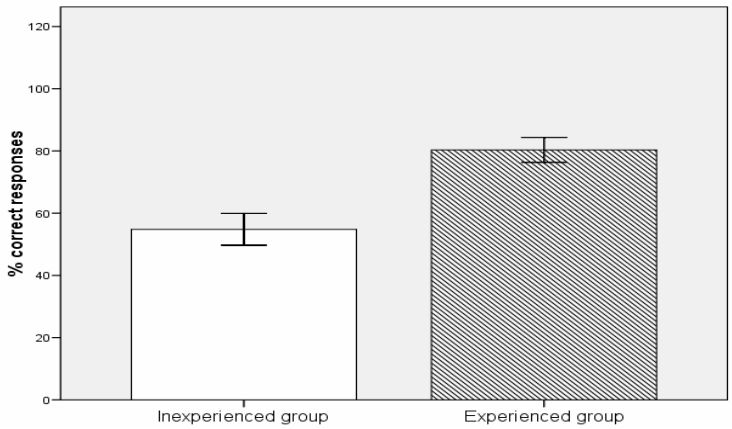


Figure 5-1. Mean percent correct production of coarticulated Mandarin tones by inexperienced and experienced learners of Mandarin

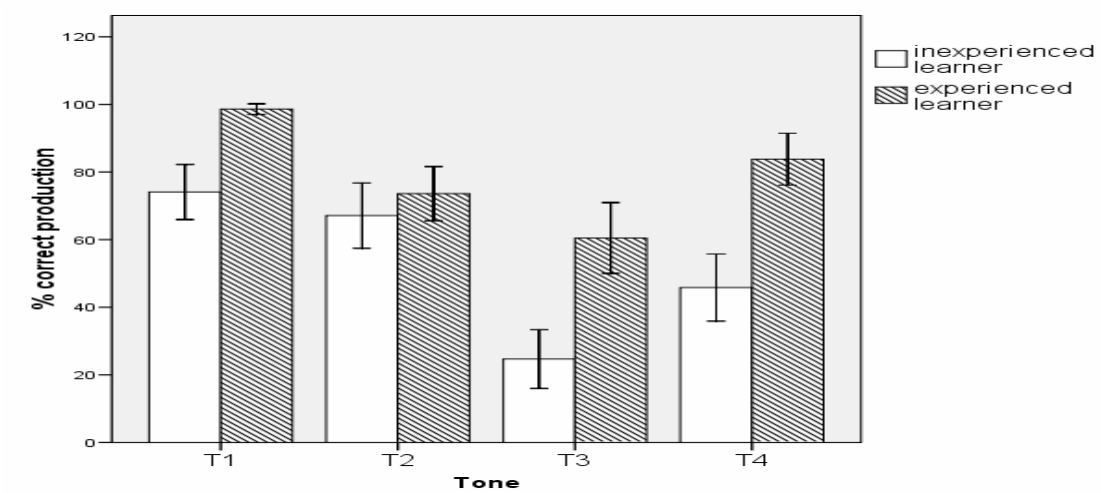
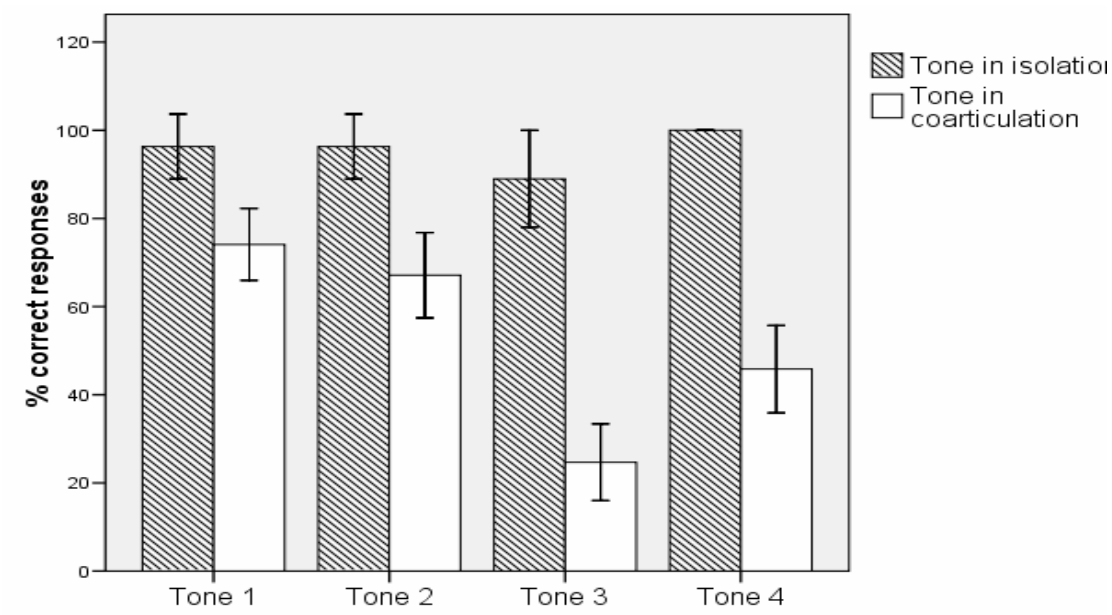


Figure 5-2. Mean percent correct production of coarticulated tones by inexperienced and experienced American learners of Mandarin. Error bars indicate standard error.

Effects of tonal condition

A)



B)

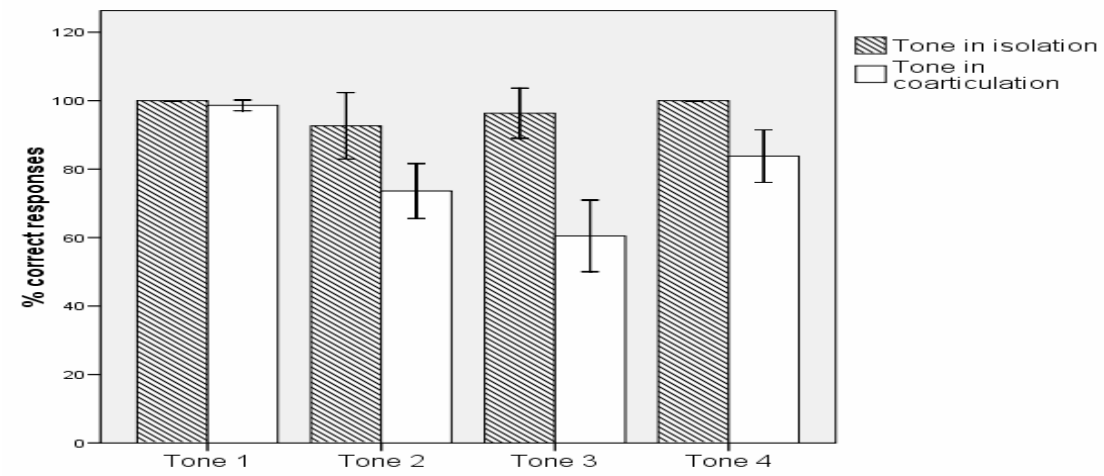


Figure 5-3. Mean percent correct production of coarticulated Mandarin tones in different tonal conditions by (A) inexperienced learners and (B) experienced learners. Error bars indicate standard error.

Figures 5-3 A and B show how inexperienced and experienced groups of American learners identify Mandarin tones under tonal condition. Overall, the isolated tonal condition allowed greater accuracy of identification (96.333% accurate responses)

for listeners than did the coarticulated tonal condition (65.722% accurate responses). A three-way repeated measures ANOVA, with Tonal condition (isolation and coarticulation) and Tone (Tone 1, Tone 2, Tone 3, Tone 4) as the within-subject factor and Group (inexperienced, experienced) as the between-subject factor, yielded a main effect of Tonal condition [F (1, 16) = 51.185, p< .001], Tone [F (3, 14) = 58.605, p< .001], but not main effect of Group [F (1, 16) = 2.899, p= .108].

Furthermore, it was found that the interaction of Tonal Condition X Tone X Group was marginally different [F (3,14) = 3.05, p=0.64] For inexperienced groups of listeners, there was a consistent trend to perceive tones more accurately in the compatible tonal context than in the conflicting tonal context (Tone 1, 96.33% vs. 74.10%; Tone 2, 96.33% vs 67.01%; Tone 3, 89.00% vs 24.69%; Tone 4, 100.00% vs. 45.83%). All t-tests indicated that the differences between the accurate perception rates of tones which occurred in different tonal contexts was significant (Tone 1, [*t*(1, 79) =1.906, *p*=.060]; Tone 2, [*t*(1, 79) =2.120, *p*=.037]; Tone 3, [*t*(1, 61) =5.888, *p*=.000); Tone 4, [*t*(1, 79) =3.841, *p*=.000).

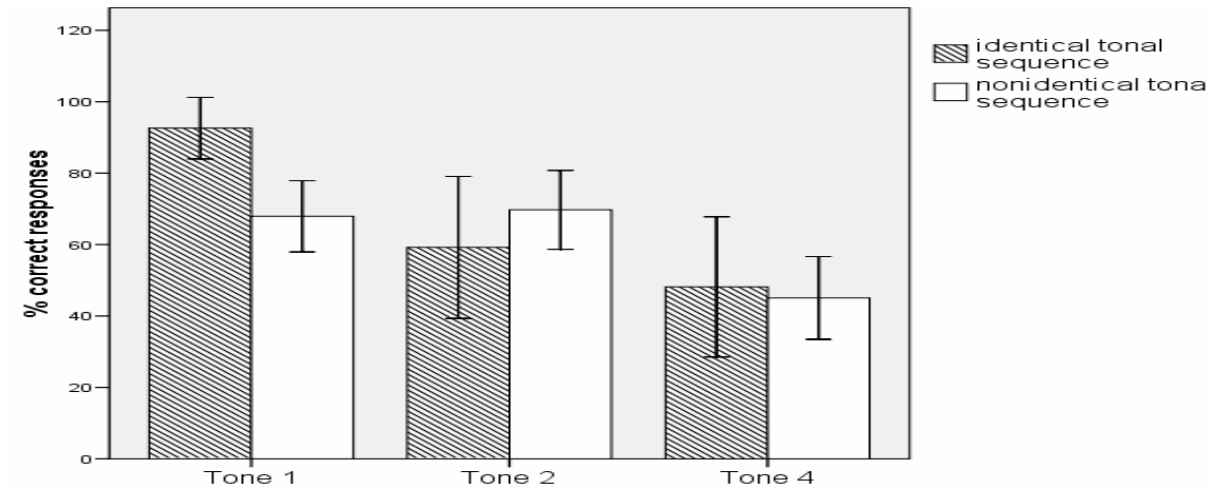
Although the experienced group (Figure 4-3 B) showed a tendency to perceive tones more accurately in isolated tonal condition than coaticulated tonal condition (Tone 1, 100.00% vs. 98.63%; Tone 2, 92.67% vs 73.61%; Tone 3, 95.33% vs

60.50%; Tone 4, 100.00% vs. 83.81%), only Tone 3s' (3's) difference in accuracy rate for the two tonal conditions was significant: Tone 3, [*t*(1, 61) =2.763, *p*=.008) whereas the difference in accuracy rate of Tone 2, [*t*(1, 79) =1.652, *p*=.102]; Tone 1, [*t*(1, 79) =.618, *p*=.539]; Tone 4, [*t*(1, 79) =1.481, *p*=.143) did not reach the significant level.

Effects of tonal sequence

Since Tone 3 only can be found in the nonidentical tone sequence, the data of Tone 3 was not analyzed in this section.

A)



B)

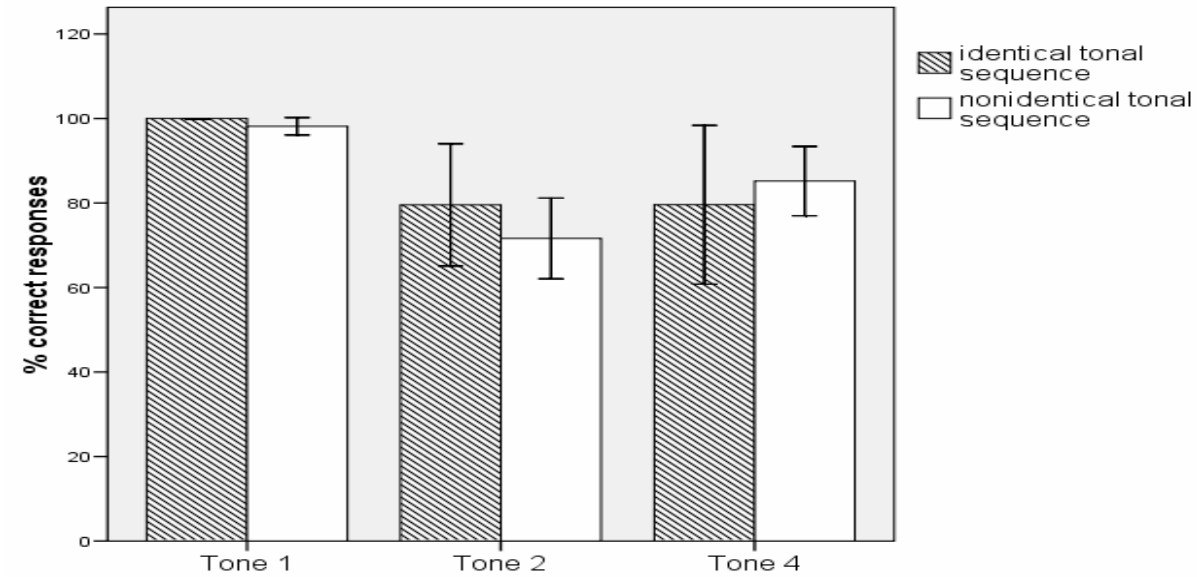


Figure 5-4. Mean percent correct production of coarticulated Mandarin tones in different tonal sequences by (A) inexperienced learners and (B) experienced learners. Error bars indicate standard error.

Figure 5-4 shows, overall, American speakers tend to produce tones better in identical tonal sequence (76.53%) than in nonidentical tonal sequence (68.52%). A three way repeated measures ANOVA, with Tonal sequence (identical and nonidentical) and Tone (Tone 1, Tone 2, Tone 3, Tone 4) as the within-subject factor and Group (inexperienced, experienced) as the between-subject factor, yielded significant main effects of Tonal Sequence [$F(1, 34) = 4.842$, $p<.05$], Group [$F(1, 34) = 11.692$, $p=.002$] and Tone [$F(2, 33) = 20.181$, $p<.001$]. Post hoc analyses (Bonferroni adjusted $p <.05$) showed that Tone 1 was produced more accurately than Tone 2 and Tone 4, but that production accuracy of Tone 2 and Tone 4 was not significantly different.

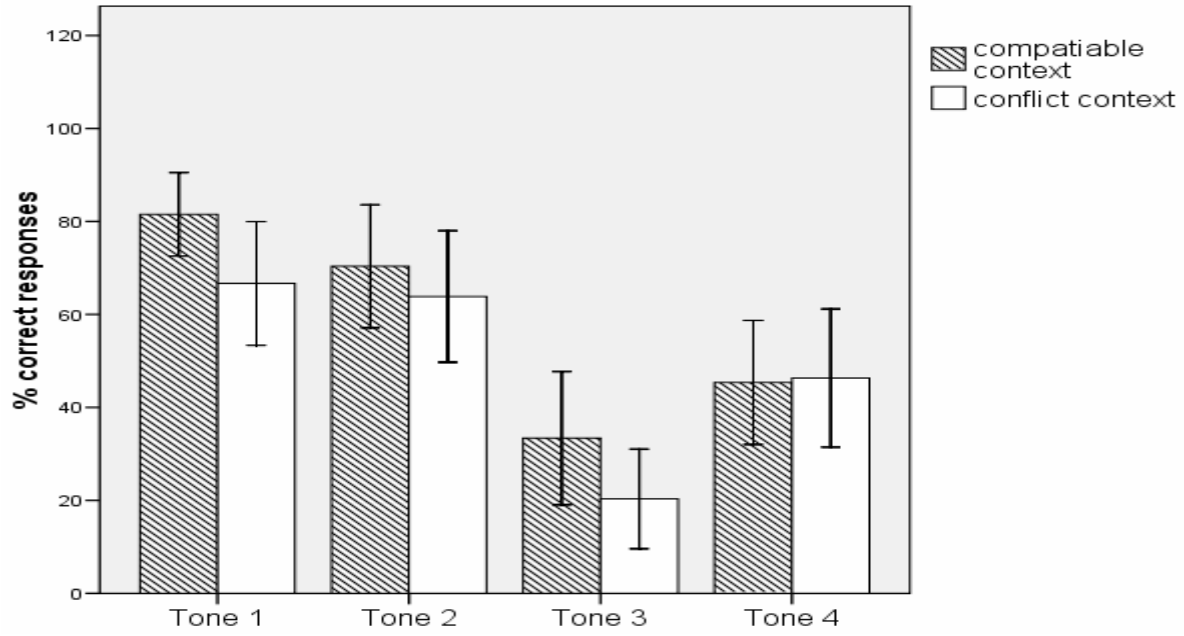
Additionally, a three-way interaction between Group, Tonal Sequence and Tone was significant [$F(2, 33) = 3.848$, $p< .05$]. Inexperienced learners' production of Tone 1 was found to be significantly more accurate in the identical environment than in the non-identical sequence [$t(1, 70) = 2.737$, $p<.01$]. In contrast, there was no significant difference in the production accuracy between the two tonal sequences for Tone 1 among experienced learners [$t (1, 70) =1.271$, $p=.208$].

There was no significant difference on accuracy rate of Tone 2 or Tone 4 production between the two tonal sequences for either the inexperienced learners [Tone 2: $t(1, 70) = -.943$, $p=.349$; Tone 4: $t(1, 70) = 0.270$, $p=.788$] or the experienced learners [Tone 2: $t(1, 70) =.610$, $p=.544$; Tone 4: [$t(1, 70) =-1.497$, $p=.139$].

Effects of tonal context

Figure 5-5 shows that, overall, American speakers' production of Mandarin coarticulated in both compatible tonal environment (66.67%) and conflict tonal environment (62.03%) was equally accurate. A three way repeated measures ANOVA, with Tonal context (compatible context and conflict context) and Tone (Tone 1, Tone 2,

A)



B)

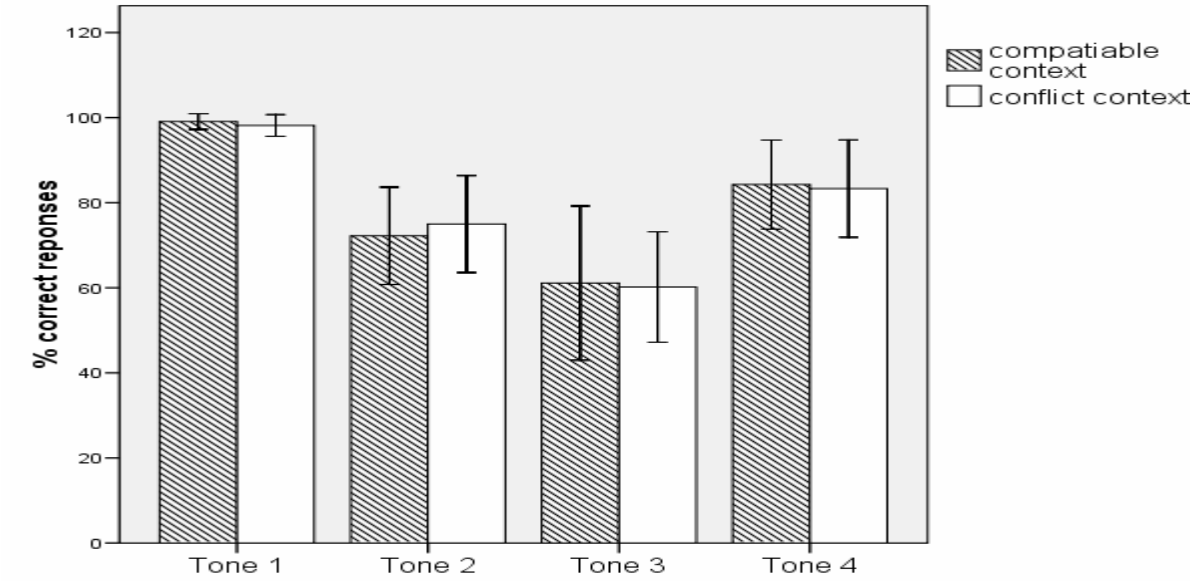


Figure 5-5. Mean percent correct production of coarticulated Mandarin tones in different tonal contexts by (A) inexperienced learners and (B) experienced learners. Error bars indicate standard error.

Tone 3, Tone 4) as the within-subject factor and Group (inexperienced, experienced) as the between-subject factor, yielded a significant main effect of Tone [$F(3, 32) = 50.928$, $p<.001$] and Group [$F(1, 32)= 26.263$, $p<.001$]. However, The effect of Tonal Context

[F(1, 34) = 2.635, *p*=.114] did not reach significance. Tone was found to interact marginally significantly with Tonal Context [F(3, 34) =2.686, *p*=.063].

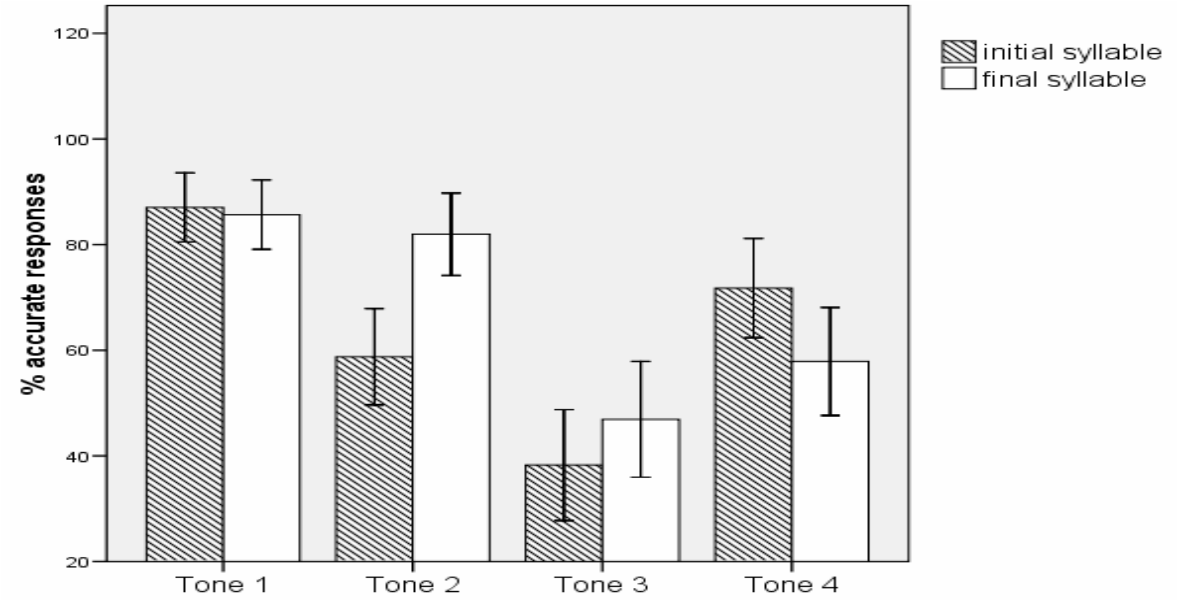
Follow-up tests showed that the production of Tone 1 was nearly significantly more accurate in compatible context than in conflicting context [t(1, 142) = 1.719, p=.088]. On the contrary, no significant difference was observed for the production of Tone 2 [t(1, 142) = 0.295, p=.768], Tone 3 [t(1, 106) = 0.866, p=.389] and Tone 4 [t(1, 142) = 0.000, p=1].

Further investigation into the effects of tonal context revealed that, for inexperienced learners, Tone 1 was produced with a higher accuracy rate in the compatible context than in the conflicting context, and this difference is marginally statistically significant [t(1, 70) = 1.852, p=.068]. For experienced learners, there was virtually no difference in accuracy rate in the two tonal contexts [t(1, 70) = .583, p=.562].

Effects of syllable position

Figure 5-6 shows that, overall, American speakers produce tones more accurately in final syllable position (68.53%) than in initial syllable position (62.95%). A three way repeated measures ANOVA, with Syllable position (initial syllable position and final syllable position) and Tone (Tone 1, Tone 2, Tone 3, Tone 4) as the within-subject factor and Group (inexperienced, experienced) as the between-subject factor, yielded a significant main effect of Syllable position [F(1, 52) = 4.435, p< .04], Group [F(1, 52) = 42.715, p<.001] and Tone [F(3, 50) = 45.033, p<.001]. Tone was found to interact significantly with Syllable position [F(3, 50) = 45.033, p<.001] and with Group [F(3, 50) = 3.636, p<.02].

A)



B)

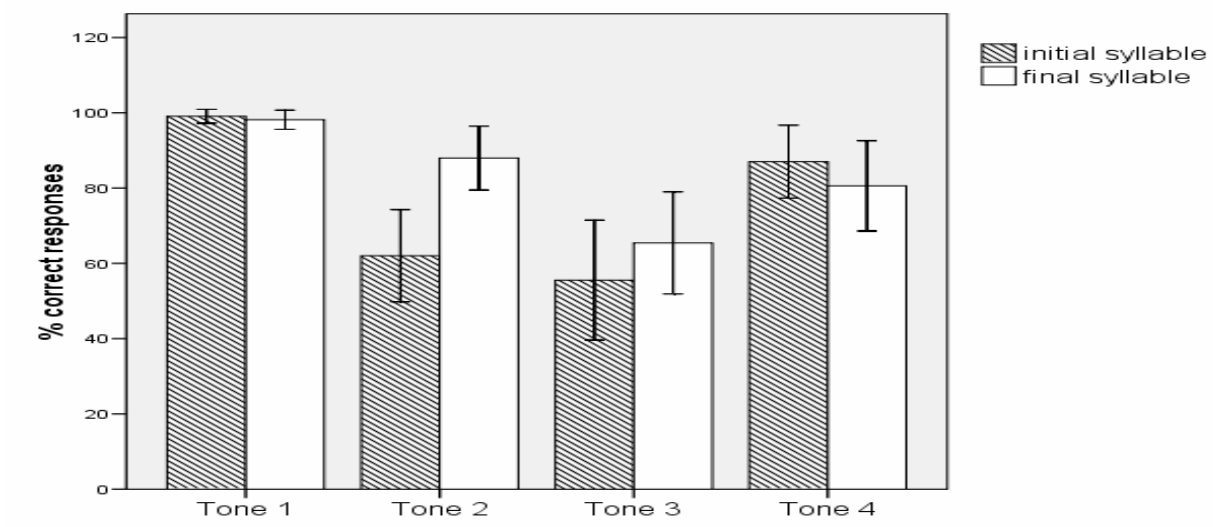


Figure 5-6. Mean percent correct production of coarticulated Mandarin tones in different syllable positions by (A) inexperienced learners and (B) experienced learners. Error bars indicate standard error.

Follow-up tests on the significant interaction between Tone and Syllable Position showed that the production of Tone 2 was significantly more accurate in final syllable position than in initial syllable position [$t(1, 142) = -3.870$, $p<.001$]. On the contrary,

Tone 4 production was significantly more accurate in initial syllable position than in final syllable position [$t(1, 142) = 2.001$, $p<.05$]. However, Tone 1 and Tone 3 productions were equally accurate in both syllable positions, [Tone 1, $t(1, 142) = .300$, $p=.765$; Tone 3, $t(1, 104) = -1.239$, $p=.218$].

Further investigation into the effects of syllable position showed that Tone 2 was produced significantly more accurately in final syllable position by both inexperienced learners [$t (1, 70) = -2.49$, $p=.015$] and experienced learners [$t(1, 70) = -3.05$, $p=.001$]. Inexperienced learners produced Tone 4 significantly better in initial syllable position [$t(1, 70) = 2.21$, $p=.030$], but syllable position did not affect experienced learners' production accuracy of this tone [$t(1, 70) = .84$, $p=.41$].

Tone 3 in tone sandhi

When Tone 3 occurs before another Tone 3, it is changed to a rising tone which perceptually sounds like Tone 2. The format of this phonological change is:

Tone 3 → Tone 2] / __ Tone 3

In the present study, the production of Tone 3 + Tone 3 by two groups of learners was tested. It was found that no learners in the inexperienced group produced the tonal combination as Tone 2 + Tone 3 and only one participant once produced the tonal combination as Tone 3 + Tone 3, and six participants produced the tonal combination as Tone 2 + Tone 2 with different frequencies. In the group of experienced learners, only one participant once produced the tonal combination Tone 3 + Tone 3 as Tone 2 + Tone 3, and the same participant produced the tonal combination twice as Tone 3 + Tone 3. The other eight participants produced the tonal combination as Tone 2 + Tone 2 with various frequencies.

Obviously, initial syllable Tone 3 and final syllable Tone 3 tonal combination Tone 3 + Tone 3 were not acquired by either inexperienced learners or experienced learners. Only one participant in each group produced Tone 3 accurately in the tonal combination. Most participants exaggerated the rising part of Tone 3 on both initial syllable Tone 3 and final syllable Tone 3. Therefore the tone combination Tone 2 + Tone 2 is the most common substitution tonal sequence for the intended tone combination Tone 3 +Tone 3.

Error Type

Confusion matrices for production of coarticulated tones in different syllable positions are shown in Tables 5-2 through 5-5. The data was based on 108 responses (12 stimuli * 9 participants) for Tone 1, Tone 2 and Tone 4, and 81 responses (9 stimuli * 9 participants) for Tone 3. The results of Tone 3 in the tonal combination of Tone 3 +Tone 3 were not included.

Table 5-1 Confusion matrix for tones in initial position produced by inexperienced learners, with the percentages in parentheses

Tone	Produced as					
	Tone 1	Tone 2	Tone 3	Tone 4	Low flat	Low falling
Tone 1	82 (76.6)	1(9)	0 (0)	14 (13.1)	10 (9.3)	0
Tone 2	14 (13)	60 (55.6)	0 (0)	12 (11.1)	20 (18.5)	2 (1.9)
Tone 3	3 (3.7)	49 (60.5)	17 (8.6)	12 (14.8)	0	0
Tone 4	36 (33.3)	2 (1.9)	0 (0)	61 (56.5)	0	9 (8.3)

Table 5-2 Confusion matrix for tones in initial position produced by experienced learners, with the percentages in parentheses

Tone	Produced as					
	Tone 1	Tone 2	Tone 3	Tone 4	Low flat	Low falling
Tone 1	107 (99.1)	0 (0)	0 (0)	0 (0)	1 (0.9)	0
Tone 2	9 (8.3)	67 (62.1)	0 (0)	3 (2.8)	24 (22.2)	5 (4.6)
Tone 3	2 (2.5)	29 (35.8)	45 (55.5)	5 (6.2)	0 (0)	0 (0)
Tone 4	7 (6.5)	0 (0)	0 (0)	94 (87)	0 (0)	7 (6.5)

Table 5-3 Confusion matrix for tones in final position produced by inexperienced learners, with the percentages in parentheses

Tone	Produced as				
	Tone 1	Tone 2	Tone 3	Tone 4	Toneless
Tone 1	79 (73.1)	18 (16.7)	5 (4.6)	3 (2.8)	3 (2.8)
Tone 2	12 (11.1)	84 (77.8)	9 (8.3)	3 (2.8)	0 (0)
Tone 3	2 (2.5)	55 (67.9)	23 (28.4)	1 (1.2)	0
Tone 4	15 (13.9)	21 (19.4)	3 (2.8)	38 (35.2)	31 (28.7)

Table 5-4 Confusion matrix for tones in final position produced by experienced learners, with the percentages in parentheses

Tone	Produced as				
	Tone 1	Tone 2	Tone 3	Tone 4	Toneless
Tone 1	106 (98.1)	2 (0.9)	0 (0)	0 (0)	0 (0)
Tone 2	5 (4.6)	92 (85.2)	10 (9.3)	0 (0)	0 (0)
Tone 3	0 (0)	27 (33.3)	53 (65.4)	1 (1.2)	0 (0)
Tone 4	0 (0)	2 (1.9)	0 (0)	86 (79.6)	20 (18.5)

As shown in the above tables, three kinds of error types were found: tonal direction confusion, tonal height confusion, and simultaneous direction and height confusion. The definition of each error type is as following:

- **Tonal direction confusion**: the direction of tonal category of the target tone is different from the produced tone.
- **Tonal height confusion**: the height of tonal category of the target coarticulated tone is the same as the identified tone, but the tonal height is different.
- **Tonal direction plus height confusion**: the direction and height of tonal category of the target tone are different from the produced tone.

The tonal confusions are divided according to error type in Table 5-5. In the following paragraph, the explanation of Table 5-5 is provided:

1) Initial syllable T1 is a flat tone, whereas T2 is a rising tone, T3 is a falling-rising tone and T4 is a falling tone. When Initial syllable T1 was produced as T2, T3 or T4, the error was labeled as direction confusion.

When initial syllable T1 was produced as a low flat tone, the error was treated as height confusion.

2) Initial syllable Tone 2 is a rising tone, whereas T1 is a flat tone, and T4 is a falling tone. When initial syllable T2 is produced as T1 or T4, the error was labeled as direction confusion.

Initial syllable T3 phonetically shows the same direction as T2. However, T3 starts lower and ends lower than T2, and turns later than T2. Therefore, T2 produced as T3 was labeled height confusion.

When initial syllable T2 was produced as a low flat tone or a low falling tone, tonal direction was not rising and tonal height was not high, so the error was labeled as direction plus height confusion.

3) Initial syllable Tone 3 is a falling-rising tone, whereas T1 is a flat tone, and T4 is a falling tone. When Initial syllable T3 was produced as T1, T2 or T4, the error was labled as the error direction confusion.

Initial syllable T2 is a rising tone, which has the same direction as initial syllable T3. Therefore, when initial syllable T3 was produced as T2, the error was labled as the error height confusion.

4) Initial syllable Tone 4 is a falling tone, whereas T1 is a flat tone, T2 is a rising tone and T3 is a falling-rising tone. When initial syllable T4 was produced as T1, T2 or T3, the error was direction confusion.

When initial syllable T4 was produced as a low falling tone, the error was treated as height confusion.

5) When Final syllableT1 was produced as T2, T3 or T4 by American learners, the confusion was categorized as tonal direction confusion.

When final syllable T1 was produced as a tone with lower pitch level, truncated duration of pitch and smaller amplitude of pitch, the confusion was categorized into tonal direction confusion.

6) Final syllable T2 is a rising tone, whereas T1 is flat and T4 is a falling tone. Therefore, T2 misproduced as T1 or T4 was labeled a direction error.

Final syllable T2 is a rising tone, as is final syllable T3. However, T3 rises lower than T2. Therefore, final syllable T2 produced as T3 was labeled a tonal height error.

7) Final syllableT3 rises, whereas T1 and T4 are not rising tones. Therefore, T3 produced as T1 or T4 was labeled a tonal direction error.

Final syllable T3 shows a rising direction as T2 does, but final syllable T3 falls first and rises lower than T2. Therefore, T3 produced as T2 was labeled a tonal height error.

8) Final syllable T4 is a falling tone, unlike T1, T2 and T3. Therefore, T4 produced as T1, T2 or T3 was labeled a tonal directional error.

When T4 was produced as a tone with lower pitch level, truncated duration of pitch and smaller amplitude of pitch, the confusion was categorized as tonal direction confusion

Figure 5-7 shows that inexperienced learners tended to produce more direction errors than experienced learners for final syllable Tone 1 [χ^2=23.635, p=.000], initial syllable Tone 2 [χ^2=7.855, p=.005], syllable final Tone 2 [χ^2=5.510, p=.019], initial syllable Tone 4 [χ^2=26.975, p=.000], and final syllable Tone 4 [χ^2=41.213, p=.000]. When producing initial syllable Tone 1, experienced learners did not make a single mistake on the tonal direction, whereas inexperienced learners still made 15 errors on the flat direction of initial syllable Tone 1. However, the difference in numbers of errors

made by inexperienced learners and experienced learners was only marginally significant for initial syllable Tone 3 [χ^2=3.366, p=.066] and was non significant for final syllable Tone 3 [χ^2=1.035, p=.311].

Figure 5-8 shows that inexperienced learners tended to produce more height errors than experienced learners for initial syllable Tone 1 [χ^2=7.759, p=.005], initial syllable Tone 3 [χ^2=9.890, p=.002], and final syllable Tone 3 [χ^2=19.361, p=.000]. When producing final syllable Tone 1, experienced learners did not even make a single mistake on the tonal height, whereas inexperienced learners still made 3 errors on final syllable Tone 1. However, the difference between inexperienced learners and experienced learners was only marginally significant for final syllable Tone 4 [χ^2=3.747, p=.053] and non significant for syllable final Tone 2 [χ^2=.058, p=.810] and initial syllable Tone 4 [χ^2=.270, p=.603].

Table 5-5 Tonal confusions categorized into error types.

Target tone	Tonal direction confusion	Tonal height confusion	Tonal height and direction confusion
1) Initial high flat T1	T2 T3 T4	Low flat tone	
2) Initial rising T2	T1 T4	T3	Low flat tone Low falling tone
3) Initial low falling T3	T1 T4	T2	
4) Initial high falling T4	T1 T2	Low falling tone	
5) Final flat T1	T2 T3 T4	Toneless	
6) Final rising T2	T1 T4	T3	
7) Final falling-rising T3	T1 T4	T2	
8) Final falling T4	T1 T2, T3	Toneless	

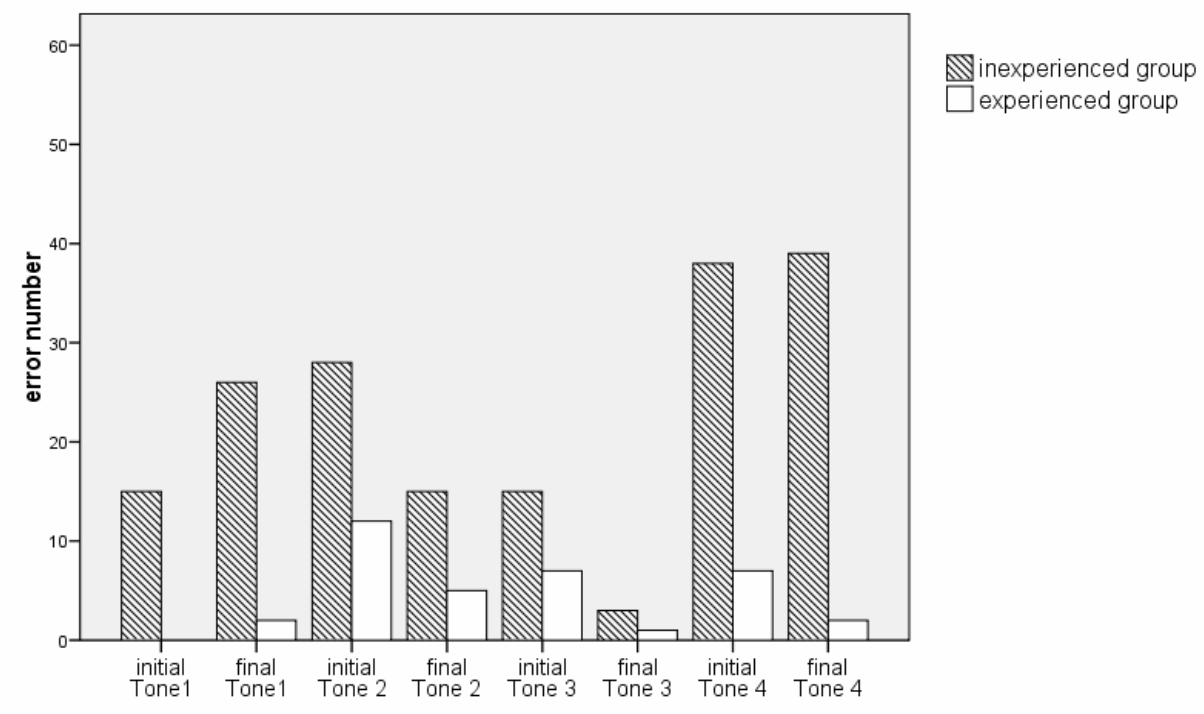


Figure 5-7. Frequency of tonal direction confusion by two groups of American learners

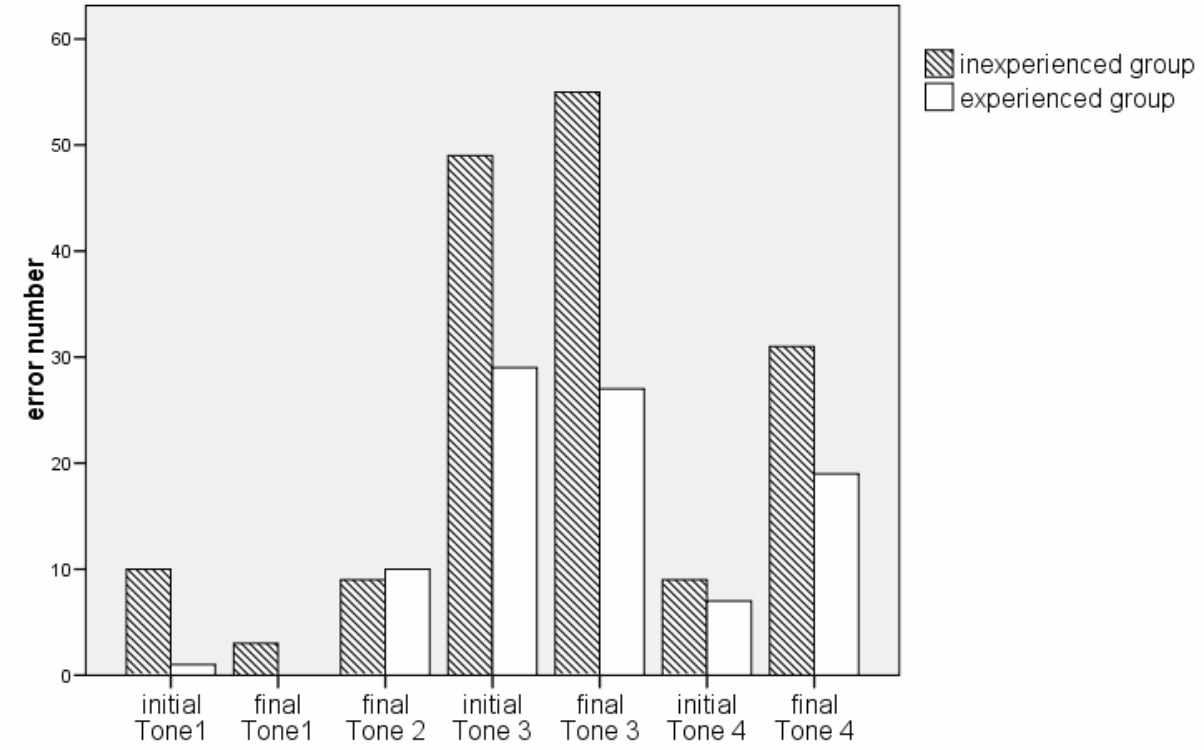


Figure 5-8. Frequency of tonal height confusion by two groups of American learners

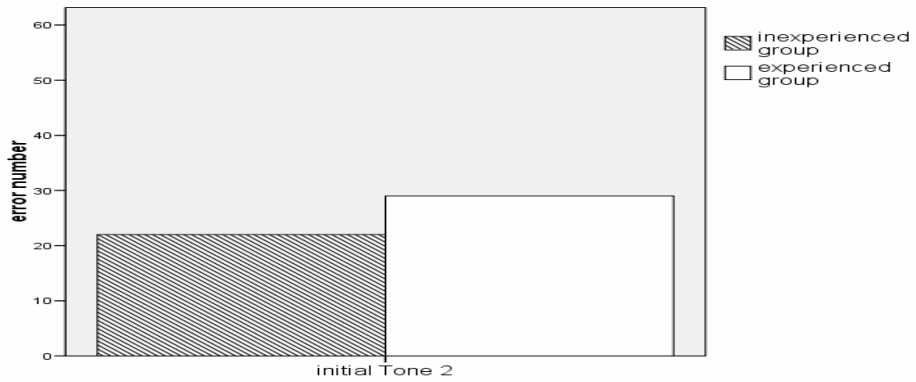


Figure 5-9. Frequency of tonal direction plus height confusion by two groups of American learners

There was no significant difference in the number of simultaneous height and direction errors by inexperienced and experienced learners for initial syllable Tone 2 [χ^2=.961, p=.327].

Relationship between Perception and Production of Coarticulated Tones

In this section, the relationship between perception and production of coarticulated Mandarin tones is analyzed based on the data from the two whole groups and individual learners in those two groups.

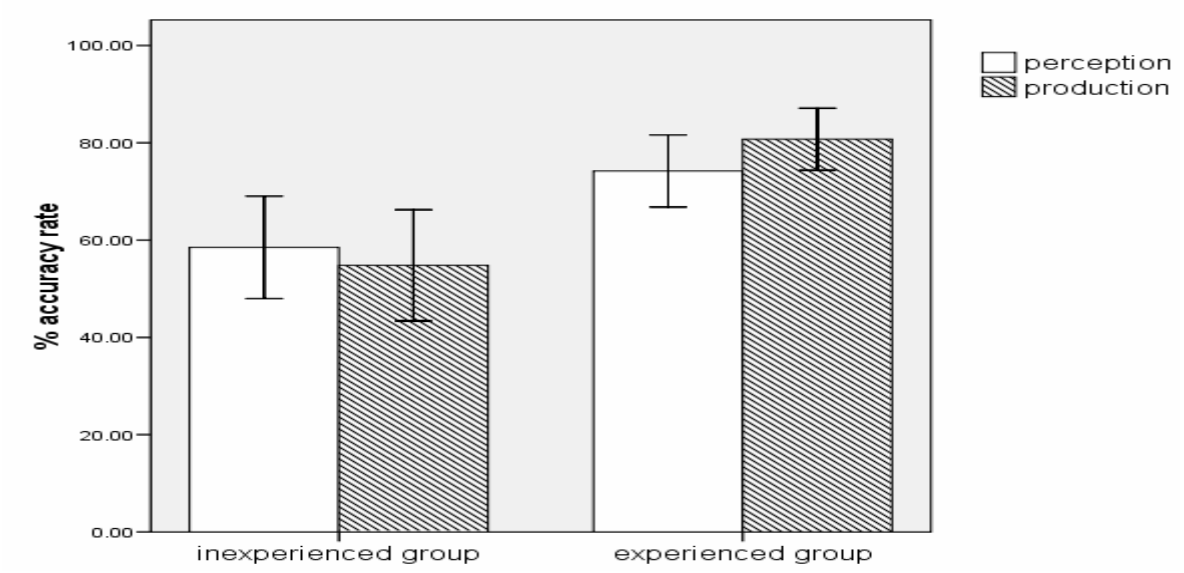


Figure 5-10. Mean percentage of accuracy of perception and production of coarticulated Mandarin tones by two groups. Error bars indicate standard error.

As shown in Figure 5-10, the accuracy rate of experienced learners producing Mandarin coarticulated tones is better than the accuracy rate of their perception of the tones [*t* (538)=1.974, *p*=-2.154], whereas inexperienced learners performed slightly worse at producing the coarticulated tones than at perceiving them [*t* (538)= 1.056, *p*=0.292]. However, neither group's difference in accuracy rates of perception and production reached significance.

Table 5-6 Individual inexperienced American learners' perception and production accuracy rate. These data are across all coarticulated lexical tones in different syllable positions.

Participant	Perception	Production	Difference
1	80	70	-10
2	38	38	0
3	42	31	-11
4	53	60	7
5	80	89	9
6	46	56	10
7	67	50	-17
8	53	48	-5
9	68	52	-16

Table 5-7 Individual experienced American learners' perception and production accuracy rate. These data are across all coarticulated lexical tones in different syllable positions

Participant	Perception	Production	Difference
1	80	83	3
2	84	79	-5
3	71	88	17
4	53	83	30
5	68	60	-8
6	87	88	1
7	80	80	0
8	81	92	11
9	63	73	10

Table 5-6 displays the accuracy rate of perception and production of coarticulated Mandarin lexical tones by inexperienced American learners. According to a Pearson Correlation test, perception and production are not correlated (r=.802, p<.01).

Table 3-6 shows that accuracy rates of perception and production of coarticulated Mandarin tones by experienced American learners are also not significantly correlated (r=.352, p=.352).

Discussion

This study was designed to investigate how coarticulated Mandarin Chinese tones in disyllabic words are produced by American learners with different amounts of language experience (inexperienced learners vs. experienced learners). The purpose of the production experiment was to test whether American learners' pronunciation of Mandarin Chinese coarticulated tones improves with language learning experience, whether the syllable position and tonal context affect identification accuracy, and what strategies were used by American students to modify Mandarin coarticulated tones in disyllabic words.

Effects of Learning Experience

The results showed that experienced learners produced coarticulated tones with a higher accuracy rate than inexperienced learners. Experienced learners performed better not only overall, but also on each individual tone. Therefore, we can conclude that learning experience helps American learners produce Mandarin tones more accurately when the tones occur in a coarticulation environment. This result confirmed the findings from previous studies that experienced learners produced L2 features more accurately than inexperienced learners.

Language learning experience also made learners' production of coarticulated tones less subject to contextual effects. In the present study, syllable position affected inexperienced learners' but not experienced learners' production of coarticulated Tone 1. In terms of effect of linguistic factors, both tonal sequence and tonal context influence

the accuracy of production of Tone 1 by inexperienced learners, but not experienced learners. Since experienced learners have had more time to study the target language, their tone production is more native like, which means the experienced learners are less influenced by linguistic factors that do not affect the production of tones by native speakers of Mandarin.

Regarding the substitution of tones in production, there is more variation in the substitution of errors made by the inexperienced group than by the experienced learners. This phenomenon is concluded from the comparison between final syllable tonal errors made by inexperienced learners and experienced learners. The distribution of tonal errors of the target tone was across other three tones by inexperienced learners, whereas only one tone was chosen to substitute the target tone by experienced learners.

Effects of Tonal Condition

Tonal condition affected how accurately Tone 3 was produced by both experienced and inexperienced American learners. However, the effect of tonal condition on Tone 1, Tone 2 and Tone 4 is only found on inexperienced learners.

Effects of Tonal Sequence and Tonal Context

There was no effect of tonal sequence (identical vs. nonidentical sequence) or tonal context (compatible vs. conflicting context) on Tone 2, Tone 3, or Tone 4 production accuracy by either group of learners. However, there was an effect of tonal sequence and tonal context on the accuracy rate of Tone 1 production by inexperienced learners, but not by experienced learners. Inexperienced learners produced Tone 1 with a higher accuracy rate in the identical tonal sequence than in the non-identical tonal

sequence. In addition, their production of Tone 1 was more accurate in compatible tonal context than in conflicting tonal context.

Two modification strategies have been found to be used by American learners to produce Mandarin coarticulated tones. One is no pitch gap in disyllable and another is no double stressed disyllable. The most frequent combination of modifying tones in conflicting tonal context is Tone 4 +Tone 2. The falling tone and the rising tone are connected by the low F0 offset of the falling tone and the low F0 onset of the rising tone. Therefore, there is no pitch gap within a disyllabic word. This kind of avoidance of pitch break was also found in a previous study on production of Vietnamese tones by American speakers (Nguyen & Macken, 2008). This modification strategy might be driven by the American learners' L1 background. It is rare to have a dramatic pitch shift from one nuclear tone to another nuclear tone in one sentence (Cruttenden, 1997). Meanwhile, the other two most frequent combinations of modifying tones are low tone +high tone for the tonal combinations Tone 2 + Tone 1 and high tone+ toneless for in the tonal combinations Tone 4 + Tone 4. As with the first type of strategy, the second type of strategy can be also attributed to L1 transfer. Disyllabic English words do not allow both syllables to be stressed. A stressed syllable is associated with high pitch, whereas an unstressed syllable is associated with low pitch. American students transfer this pitch constriction from their L1 to their Mandarin learning and modify a tone with high pitch to a low pitch. This strategy was employed on modifying initial Tone 2 and final Tone 4.

Effects of Syllable Position

Syllable position affected how accurately Tone 2 was produced by both experienced learners and inexperienced American learners. Both groups showed a

higher production accuracy rate for Tone 2 in the final syllable position. This result may be attributed to the rising intonation used to form questions in English. On the other hand, the factor of syllable position only affected inexperienced learners' production of Tone 4. For inexperienced learners, the accuracy rate of Tone 4 production on the initial syllable was higher than that on the final syllable. No clear explanation can be offered to explain why Tone 4 was produced more accurately in initial syllable position. Syllable position, however, did not affect production accuracy of Tone 1 or Tone 3 production.

Development of Production of Coarticulated Tones

Three types of errors were found in both groups: tonal direction confusion, tonal height confusion, and tonal direction plus height confusion. There was no difference between inexperienced and experienced learners' frequency of directional confusion, height confusion, and tonal direction plus height confusion errors. It was found that the ability to produce tones in terms of both tonal direction and tonal height quickly improved.

Relationship between Perception and Production of Coarticulated Tones

If we take the groups of learners as a whole, we can say that the ability to perceive coarticulated tones developed in parallel with the ability to produce coarticulated tones for both inexperienced and experienced American learners.

Considering individual performance on perception and production of coarticulated tones, there is a strong positive correlation between perception performance and production performance by inexperienced American learners, but not experienced American learners. In other words, the proficiency of inexperienced learners' perception is the predictor of the proficiency of their production of Mandarin coarticulated tones.

Accurate perception leads to accurate production for American learners with less learning experience.

Conclusion

This study revealed that American learners with more learning experience were more accurate in producing coarticulated Mandarin tones in disyllabic words. Furthermore, it was found that with increased experience, production of certain coarticulated tones becomes more resistant to such phonological and phonetic factors as syllable position and tonal environment. It was also found that tonal sequence and tonal context only affected tone 1 production; and syllable position affected tone 2 and tone 4 production accuracy. A strong positive correlation exists between perception performance and production performance for inexperienced American learners, but not experienced American learners.

CHAPTER 6
GENERAL DISCUSSION AND CONCLUSION

In this chapter, the findings from the study are first summarized. Second, general discussion based on the findings is provided. Third, I propose future research on the acquisition of Mandarin tones by L2 speakers based on the results and limitations of the present study. Finally, I make some pedagogical suggestions regarding teaching Chinese tones.

Summary of Results

The experiments in this dissertation set out to test whether amount of learning experience affects adult American learners' perception and production of Mandarin tones. One of the goals of the study was to investigate whether syllable position and tonal environment affect English speakers' perception and production of Mandarin tones, and another goal was to study the relationship between perception and production of Mandarin tones in second language acquisition. The first two experiments, presented in Chapter three, investigated the perception and production of isolated Mandarin tones. The third experiment, presented in Chapter four, examined the perception of coarticulated Mandarin tones in disyllabic words. The fourth experiment, presented in Chapter five, studied the production of coarticulated Mandarin tones in disyllabic words. The main hypothesis tested across all experiments in this study was that L2 learners with more learning experience outperform those with less learning experience when perceiving and producing Mandarin tones. Results from the four experiments supported this hypothesis.

Experiments 1 and 2

Results from the first perception experiment suggest that experienced American learners perceive isolated Mandarin tones more accurately than inexperienced American learners. However, the production experiment showed no significant difference between experienced and inexperienced learners on producing isolated Mandarin tones. The error types in perception and production experiments are the same for both groups: Tone 1 and Tone 4 are frequently confused with each other, as are Tone 2 and Tone 3. However, although American learners ignore the pitch contour differences between these pairs of tones in production, they do tend to produce *duration* differences to differentiate Tone 1 from Tone 4 and Tone 2 from Tone 3. Generally speaking, inexperienced learners' perception and production abilities developed equally. As their language learning experience increased, experienced learners' perception ability became better than their production ability. (However, there is no correlation between perception and production of isolated Mandarin tones for either group of learners.)

Experiment 3

The results from experiment 3, which tested perception of coarticulated Mandarin tones, showed that experienced learners outperformed inexperienced ones on perception of coarticulated tones; this holds for the pooled results across all four tones and for each tone individually. Therefore, learning experience does aid American learners' ability to perceive coarticulated Mandarin tones. Tonal condition, that is, whether a tone in isolation or in coarticulation, affected the perception of tones by both groups of learners. Tones in isolation were identified with higher accuracy rate than

tones in coarticulation. Tonal sequence, that is, whether the two syllables of the word had the same tone or two different tones, was found to affect the perception of Tone 1 for inexperienced learners. Tone 1 in identical tonal sequence was identified with higher accuracy rate than it was in nonidentical tonal sequence. Tonal context, that is, whether the pitch of the two tones at their intersection was compatible or conflicting, affected the perception of Tone 1 by both groups of learners. Tone 1 in compatible context was perceived better than it was in conflicting tonal context. In terms of the effect of syllable position, it was found that Tone 2, Tone 3 and Tone 4 of the final syllable were identified more accurately than those of the first syllable.

Two types of perceptual errors were found: tonal direction misperception and tonal height misperception. As learning experience increased, the frequency of tonal direction errors decreased, whereas the frequency tonal height errors did not change.

Experiment 4

The results from experiment 4, which tested production of coarticulated Mandarin tones, showed that experienced learners outperformed inexperienced ones, not only overall but also for each of the four tones individually. Tonal condition, affected the perception of tones by both groups of learners. Tones in isolation were produced with higher accuracy rate than tones in coarticulation. Tonal sequence and tonal context only affected tone 1 production. Tone 1 was produced better in identical tonal sequence and compatible tonal context than it was in nonidentical tonal sequence and conflicting tonal context. Syllable position affected Tone 2 and Tone 4 production accuracy. The accuracy rate of Tone 2 production on the final syllable was higher than that on the initial syllable, whereas, the accuracy rate of Tone 4 production on the initial syllable

was higher than that on the final syllable. With increased experience, production of coarticulated tone becomes less affected by such phonological and phonetic factors as syllable position and tonal sequence and tonal contexts.

Three types of errors were found: tonal direction confusion, tonal height confusion, and tonal direction plus height confusion. It was found that the ability to produce tones in terms of both tonal direction and tonal height quickly improved. A strong positive correlation exists between perception performance and production performance for inexperienced American learners, but not experienced American learners.

General Discussion

The findings confirmed the hypothesis that longer language learning experience makes American learners perceive Mandarin tones more accurately and partially confirmed the hypothesis that longer language learning experience helps American learners produce Mandarin tones more accurately. Experienced learners' ability to identify tones is significantly better than inexperienced learners' for both isolated and coarticulated tones, whereas experienced learners' ability to produce tones is only significantly better than inexperienced learners' for coarticulated tones. In order to identify tones accurately, American speakers could acquire an ability to perceive tones categorically, that is, to exaggerate the phonological differences between four lexical Mandarin tones and ignore the phonetic variability among tokens of the same tone. This ability to categorically perceive tones keeps developing throughout their language learning experience. On the other hand, American speakers might still detect the phonetic change of tones which is driven by tonal context. However, with more learning experience, experienced learners become more familiar with the allophonic variations of tones.

The fact that there was no significant difference in the production of isolated tones between two groups, despite a difference of nine months in learning experience suggested that, in comparison to perception ability, production ability may improve at a slower pace. This might be attributed to the absence of L1 negative transfer. The English stress pattern prohibits two high tones from occurring in one disyllabic word. However, this constraint does not keep native English speakers' from producing tones on one-syllable words correctly. Therefore, both inexperienced learners and experienced learners produced tones in isolation well. When the tones are coarticulated (in two-syllable words), experienced learners show an advantage in production and produced tones in coarticulation with significantly higher accuracy rate. Here, it is assumed that this L1 negative transfer affects inexperienced learners more than experienced learners.

The results from the four experiments suggest that coarticulated tones are harder to identify and produce than isolated ones. In other words, even if learners are able to perceive and produce an isolated tone accurately, they might not be to perceive or produce it with the same accuracy when the tone is coarticulated with another (in the current study, tones in disyllabic words).

Tonal sequence and tonal context affect American learners' perception and production of Tone 1. In terms of perception, the narrower pitch range triggered by the conflicting tonal context prevents accurate identification of coarticulated Tone 1. Moreover, a changed pitch direction triggered by the conflicting tonal context makes Tone 1 harder to identify. In terms of production, there is no pitch gap in the tonal compatible context, such as Tone 1 + Tone 1 and Tone 1 + Tone 4. American learners

perform well when producing Tone 1 in these Combinations. In contrast, the conflicting tonal combination, such as Tone 4 + Tone 1, leads American students to make more errors on production. The results suggest that American learners prefer producing Tone 1 in a tonal environment without a pitch shift between two tones.

Syllable position affects American learners' perception and production of Tone 2 and Tone 4. It was found that the accuracy rates of perception and production of Tone 2 were higher on final syllable position than initial syllable position. This result may be attributed to the rising intonation used to form questions in English. However, the accuracy rate of perception of Tone 4 was higher on final syllable position, whereas the accuracy rates of production of Tone 4 was higher on initial syllable position. No clear explanation can be offered to explain why Tone 4 was produced more accurately but was perceived less accurately in initial syllable position.

One interesting finding regarding the production of isolated tones is that American learners use duration rather than pitch to differentiate Tones 1 and 4 and Tones 2 and 3. English speakers in the present study acquired the duration difference among tones earlier than pitch contour.

For the perception of coarticulated tones, it seems that tonal height is harder to acquire than tonal direction, given that tonal direction errors decreased with increased learning experience, but tonal height errors remained. However, for the production of coarticulated tones, it was found that the ability to produce tones in terms of both tonal direction and tonal height quickly improved.

The relationship of perception and production of isolated tones for American learners is still not clear. Perception and production are not correlated for either group of

learners. Some learners perceive isolated tones better than they produce them, while some do the opposite. Perception of coarticulated tones predicts production for inexperienced but not experienced learners. This finding partially contradicts the Speech Learning Model's (SLM) (Flege, 1995) proposal that accurate perception is a 'must' for accurate production.

Limitations and Future Research

This present study is the first to systematically examine the perception and production of coarticulated Mandarin tones by American adult learners with different amount of learning experience. Because this is the first such study, acquisition of coarticulated Mandarin tones by L2 speakers needs to be studied further in the future.

This study investigated the effect of language learning experience on the perception and production of Mandarin tones. Language learning experience was defined as the amount of language learning time in the classroom. The inexperienced learners had learned Mandarin for three months, the experienced learners for twelve months. Therefore, the difference in learning duration is not as great as in previous studies (Jun and Cowie, 1994; Bohon and Flege, 1992; Flege et al, 1997) on the effect of language experience on perception and production of L2 sounds. Flege et al. (1997) categorized inexperienced and experienced group based on their length of residence in the US. The former group had lived there for only 0.7 years, the latter group for 27.3 years. In future research, we may need to recruit experienced learners from an advanced level who have longer learning experience and may perform more differently than beginning learners.

Many previous studies have determined other factors than language learning experience that may also affect L2 learners' perception and/or production of target segments and suprasegmental features. An interesting topic would be to investigate how L2 learners' linguistic background affects their perception and production of Mandarin tones. Due to practical limitations, the native language background of all participants is American English, which is a non tonal, stress-based language. In the future, studies should involve participants with various language backgrounds, such as Japanese speakers whose L1 is a pitch-accent language, Thai speakers whose L1 is a tonal language, and bilingual English-Cantonese speakers who can speak a non tonal stress-based language and a tonal rhythm-based language.

Another limitation of the study is that the stimuli are not as natural as real language in conversation. It is difficult to balance the elimination of possible research bias caused by uncontrolled stimuli in the experiments against stimulation of natural behavior in perception and production of Mandarin tones in a language lab. In the future, we need to find a better way to test the perception and production of coarticulated tones by L2 learners. Here I suggest extending the testing units from word level to sentence level. Participants need to be tested on the entire sentence, since sentences are more frequently used in real conversation than isolated words.

Also, the stimuli used in the production experiments are alphabetically presented, *na* for isolated tones and nana for coarticulated tones. Because the alphabetic form looks phonologically similar to English words, American learners may be influenced to transfer the stress or intonation pattern from English to their production of these words.

The order to eliminate the alphabetic confusion, we need to change the method in the future and use Chinese characters or pictures to illustrate the tested words.

Pedagogical Implications

From the results of the present study, we can tell that, for American learners, the ability to perceive and produce Mandarin tones in isolation does not guarantee the ability to perceive and produce them in coarticulation. In the current classroom, Chinese instructors focus on introducing the phonetic and phonological features of Mandarin tones in isolation. More specifically, the pitch direction of tones in isolation is emphasized.

The biggest problem with this tendency is that it misleads learners into thinking Tone 3 is always a falling-rising tone and putting more emphasis on the rising part. Therefore, initial syllable Tone 3 is frequently produced as Tone 2. Actually, Tone 3 is changed into a low falling tone when it occurs in initial syllable position. I suggest that Chinese instructors more emphasize the low pitch of Tone 3 to American students and less emphasize the rising part of Tone 3 when they perceive and produce it. I have three arguments to support this idea. First, Tone 3 is the only low tone in the Mandarin tonal inventory. Therefore, low pitch is the defining feature of Tone 3. Second, the contour of falling rising does not appear in English at the syllable level and, American students will expect it to be difficult to master the pitch track. However, low pitch does exist in English. Therefore, simplifying the pitch contour will help American students deal with the exotic tone. Third, due to the categorical perception of Mandarin tone, I observed that native speakers of mandarin categorize all the low tones as Tone 3 when the low tone occurs in the initial syllable of disyllabic words. Therefore, American

learners will not be misunderstood by native speakers if they simply produce initial syllable Tone 3 as a low tone.

The second suggestion for teaching coarticulated Mandarin tones to American students is to emphasize the pitch height as much as pitch direction. This is essential to call American students' attention to the importance of producing two high tones in disyllabic words, although this tonal pattern conflicts with the pitch pattern in English disyllabic words.

The third suggestion is to point out there is a pitch gap between the two tones in words with the conflicting tonal context, although such a gap does not exist in the pitch contour of English disyllabic words. In the present study, it was found that American students tend to change the tonal direction or tonal height in order to fill the pitch gap. Chinese instructors need to realize that the errors of tonal direction and tonal height do not stem from their ability to produce the tone in isolation. The errors are triggered by the tonal context. Therefore, Chinese instructors must force their students to notice the pitch gap by providing more practice on the tonal combinations with pitch gap.

LIST OF REFERENCES

Best, C. T. (1995). A direct realist view of cross-language speech. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

Best, C. T., and Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34). Amsterdam: John Benjamins.

Blicher, D.L., Diehl, R.L., and Cohen, L.B. (1990). Effects of syllable duration on the perception of the Mandarin tone2/tone3 distinction: Evidence of auditory enhancement. *Journal of Phonetics.* 18, 37-49.

Bluhme, H. and Burr, R. (1971). An audio-visual display of pitch for teaching Chinese tone. *Studies in Linguistics,* 22, 51-57.

Borden, G., Gerber, A & Milsark, G. (1983). Production and Perception of the /r/-/l/ Contrast in Korean Adults Learning English. *Language Learning,* 33(3), 499-526.

Bohn, O.-S., and Flege, J. E. (1997). Perception and production of a new vowel category by adult second language learners. In J. Leather, & A. James (Eds.), Second language speech: structure and process (pp. 53–73). Berlin: Mouton de Gruyter.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (1997) Training Japanese Listeners to Identify English /r/ and /l/: Some Effects of Perceptual Learning on Speech Production*. The Journal of the Acoustical Society of America*, 101 2299-2310

Broselow, E., Hurtig, R.R., and Ringen, C. (1987). The perception of second language prosody. In G. Ioup and S.H. Weinberger (eds.), *Inter-language Phonology, The Acquisition of Second Language Sound System.* Cambridge: Newbury House Publishers. pp. 350-361.

Chao, Y. (1948). *Mandarin Primer.* Cambridge, MA: Harvard University Press.

Chen, G.T. (1974). The pitch range of English and Chinese speakers. *Journal of Chinese Linguistics,* 2, 159-171.

Cruttenden, A (1997). *Intonation*. Cambridge University Press.

Flege, J. E. and Eefting, W.(1987) "Cross-language switching in stop consonant perception and production by Dutch speakers of English", *Speech Communication.* 6, 185-202.

Flege, J. E. (1993) "Production and perception of a novel, second-language phonetic contrast", *Journal of the Acoustical Society of America*, 93 1589-1608.

Flege, J. (1984). The effect of experience on Arabs' perception of the English/s/vs./z/contrast, Folia Linguistica, 18, 117-138.

Flege, J. E. (1995). Second-Language Speech Learning: Theory, Findings and Problems. *in Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (Timonium, MD). pp. 233–273.

Flege, J.E. (1997). The role of phonetic category formation in second-language speech learning. In J. Leather and A. James (eds.): *New Sounds 97. Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech*, Amsterdam: University of Amsterdam, pp. 79-88.

Flege, J. E. Bohn, O, and Jang, S (1997) Effects of Experience on Non-Native Speakers' *Production* and *Perception* of English Vowels. *Journal of Phonetics*, 25, 437-470

Flege, J., MacKay, I. and Meador, D. (1999) Native Italian speakers' production and perception of English vowels. *The Journal of the Acoustical Society of America*, 106, 2973-987.

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

Fu, Q. J., Zeng, F. G., Shannon, R. V., and Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *The Journal of the Acoustical Society of America*, 104, 505–510.

Fu, Q., and Zeng, F. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing*, *5*, 45-57.

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

Gandour, J.T. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics,* 12, 235-261.

García Pérez, G. M. (2005). Perception of English vowels by native speakers of Spanish in a regular classroom setting. *Revista Virtual de Estudos da Linguagem, 3, 5*.

Gårding, E., Kratochvil, P., Svantesson, J.O., and Zhang, J. (1986). Tone 4 and Tone 3 discrimination in modern standard Chinese. *Language and Speech,* 29, 281-293.

Goto, H. (1971) Auditory perception by normal Japanese adults of the sounds "l" and "r", *Neuropsychologia,* 9, 317-323.

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

Gussenhoven, C. (2004), *The phonology of tone and intonation.* Cambridge, UK; New York : Cambridge University Press.

Halle′, P. A., Chang, Y. C., and Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs.French listeners. *Journal of Phonetics,* 32, 395–421.

He. Y (2004). The production of English /L/ by Mandarin speakers. MA Thesis: University of Victoria.

He, Y., Wang, Q, and Wiltshire, C. (2008). Production of English lexical stress by inexperienced and experienced learners of English. *Canadian Acoustics*, 36, 128-129.

Hewings, M. (1995). Tone choice in the English intonation of non-native speakers. *IRAL* 36, 251-265.

Ingram, J. and Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, 25, 343-370.

Jun, S. A.;and Cowie, I. (1994). Interference for 'new' and 'similar' vowels in Korean speakers of English. *Ohio State Univ. Working Papers 43:* 117–130.

Li, C. N. and Thompson, S. A. (1977). The acquisition of tone in Mandarin –speaking children. *Journal of Child Language,* 4: 185-199.

Liu, S.; Samuel, A. (2004). Perception of Mandarin Lexical Tones When F0 Information Is Neutralized. *Language and Speech*, 47, 2, 109-138.

Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In J. Leather and A. James (eds.): *New Sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second Language Speech*. Amsterdam: University of Amsterdam, pp.305-341.

Lee, Y., Vakoch, D. and Wurm, L. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 25(5), 527-544.

Liu, F. (1924). *Szu Sheng Shih Yen Lu*. Ch'un Yi Shanghai.

Liu, S., and Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language Speech*, *47*, 109–138.

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

Howie, J.M. (1976). *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge University Press, Cambridge.

MacKain, K. (1981) Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*. 2, 369-390.

Massaro, D. W., Cohen, M.M., and Tseng, C. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics,* 13, 267-290.

McAllister, R., Flege, J. E. and Piske, T. (2002) The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics,* 30, 229-258.

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

Moore, C. B., and Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America,* 102, 1864-1877.

Rochet, B. (1995) Perception and production of second-language speech sounds by adults. In *Speech perception and linguistic experience* (W. Strange, editor). Timonium, MD: York Press.

Schmidt, A., & Meyer, K. (1995). Traditional and phonological treatment for teaching English fricatives and affricates to Koreans. *Journal of Speech and Hearing Research,* 38,828-838.

Sheldon, A., and Strange, W. (1982). The acquisition of 11-1 and 111 by Japanese learners of English: evidence that speech production can precede speech perception. *Applied Psycholinguistics,* 3,243-261.

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

Shen, S. X-N. (1990). Tonal coarticulation in Mandarin. Journal of Phonetics, 18: 281-295.

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

Shih , C . (1986). *The phonetics of the Chinese tonal system..* Technical memorandum. AT & T Bell Laboratories.

Trofimovich, P., and Baker, W. (2006). Learning second-language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28,* 1-30.

Xing, J (2006). *Teaching and Learning Chinese as a Foreign Language. A Pedagogical Grammar.* Hong Kong: Hong Kong University Press.

Wang, H., Chang, B., Li, Y., Lin, L., Liu, J., Sun, Y., Wang. Z., Yu, Y., Zhang, J., and Li, D. (1986). Xiandai Hanyu Cipin Cidian [Frequency Dictionary of Modern Chinese] Beijing, China: Beijing Language Institute Press.

Wang, Y., Spence, M.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. (1995). *Error Analysis of Tone production by American Speakers and Pedagogical strategies in Teaching Tones.* Phonetics research and Teaching Chinese as a Second Language, 379-394.

Wang, Y., Sereno, J., Jongman, A., and Hirsch, J. (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of Cognitive Neuroscience,* 15, 7, 1019-1027.

Wang, Y., Jongman, A., and Sereno, J.A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America,* 113, 1033-1044.

Wang, Y. and Kuhl, P.K. (2003). Evaluating the “critical period” hypothesis: Perceptual learning of Mandarin tones in American adults and American children at 6, 10, and 14 years of age. *Proceedings of the 15th International Congress of Phonetic Sciences,* 1537-1540.

Wang, Y., Behne, D., Jongman, A., and Sereno, J.A. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Journal of Applied Psycholinguistics* 25, 449-466.

Wayland, R., Guion, S. (2003). Perceptual Discrimination of Thai Tones by Naive and Experienced Learners of Thai. *Applied Psycholinguistics,* 24, 1, 113-129.

Wayland, R., Guion, S. (2004). Training English and Chinese Listeners to Perceive Thai Tones: A Preliminary Report. Language Learning, 54, 4, 681-712.

Whalen, D. H. and Xu, Y., (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica,* 49, 25-47.

White, C.M. (1981). Tonal perception errors and interference from English intonation. *Journal of Chinese Language Teachers Association,* 16, 27-56.

Wiltshire, C and Moon, R. (2003). Phonetic Stress in Indian English vs. American English. *World Englishes*, 22, 291-303.

Wu, Z.J. (1986). *The Spectrographic Album of Mono-syllables of Standard Chinese*. Beijing: Social Science Press.

Xu, Y. (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics,* 5, 757-797.

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

Xu, Y. (1994). Production and perception of coarticulated tones, *Journal of the acoustical Society of America,* 95, 2240 – 2253.

Xu, Y. (1993). Contextual tonal variation in Mandarin Chinese. Ph.D. dissertation. The University of Connecticut.

BIOGRAPHICAL SKETCH

Yunjuan He was born in Changchun, China. She received her B.A. in Teaching Chinese as Second Language from Beijing Language and Culture University in 2001. She earned her M.A. in Linguistics from University of Victoria in 2004. She joined the University of Florida as a Ph.D. student in fall 2005