DE GRUYTER MOUTON
DOI 10.1515/caslar-2012-0003 — CASLAR 2012; 1(1): 33–53

Bei Yang

The gap between the perception and production of tones by American learners of Mandarin – An intralingual perspective

Abstract: Linguists have predominantly maintained that perception precedes production (Dinnsen 1983), an assertion also accepted by those studying second language acquisition (Flege 1995). However, an observation of acquisition of tones in Chinese as a second language suggests that American learners make different tonal mistakes in perception and production. This study explores tonal perception and production referring to the sound system of Mandarin, since a tone has a close relationship with an initial that is an onset and a final that is a rhyme within a syllable in Mandarin. The research instrument has 84 monosyllables that are representative according to the relationship among initials, finals and tones. Twenty-five American learners of Chinese in second-semester Chinese class and 11 learners of Chinese in fourth-semester Chinese class participated in this study. A two-way mixed ANOVA is the main statistical method used to analyze the acquisition data. The results reveal that tonal production is better than tonal perception. The error distribution of perception is influenced not only by tonal features, but also by initial features and final structures. For production, however, initial and final features do not influence tones. Therefore, the paper argues that tones are perceived at the phonological level and produced at the phonetic level and it takes L2 learners longer time to acquire phonological features of tones.

Bei Yang: e-mail: byang25@wisc.ed

1 Introduction

1.1 Tones of Mandarin

Tones can be regarded as suprasegmentals whose features contrast with segmental features, that is, features of consonants or vowels. There are two main

perceptual aspects of tones: "first, the human ability to perceive the physical properties of frequency, duration, and intensity, and second, the psychological response to various acoustic stimuli" (Chun 2002: 10). "Tones are perceived principally as differences in pitch, . . ." (Norman 1988: 145). Chinese is a tonal language. Each syllable carries a tone capable of distinguishing meaning. For example: two identical syllables composed of the same consonants and vowels can have entirely different meanings when they differ only by tones.

In Mandarin, each syllable has an initial, a final, and a tone. Initials, also called onsets, are the consonants at the beginning of syllables. However, an initial is not obligatory in a Mandarin syllable. The syllables without initials are called zero initial syllables, e.g. ài (love) and é (goose). Finals, that are also called rhymes, are composed of the vowels and consonants at the end of syllables. Not every syllable has a final consonant. And the only final consonants in Mandarin are the nasals.

There are four tones in Mandarin which can be further described by terms such as pitch value, tone contour, and tone category (see Table 1). Pitch is "an auditory property that enables a listener to place it on a scale going from low to high, without considering its acoustic properties" (Ladefoged 2006: 23). Its physical attribute is fundamental frequency. Pitch value presents the changes of a tone's position, including high, low, rising and falling. It shows the actual way a tone is pronounced. Tones that have the same pitch value belong to the same tone category.

Tone contour is a shape of the pitch. The 5-scale notation system of pitch value used here was designed by Chao (1930), and is widely used in academic research and pedagogy. In this notation system, 5 is high pitch, 3 is mid pitch, and 1 is low pitch. The pitch value of each tone category is marked with double numbers, 55, 35, etc.

1.2 Problem areas in existing research

American English-speaking learners have considerable difficulty learning Mandarin Chinese due to the obstacle of the four tones. Helping learners acquire the tonal system is a key point in the pedagogy of pronunciation.

The traditional method of teaching tones is based on the concept that students begin to learn tonal production only after they perceive the tones. Teachers demonstrate the four tones in class, then ask students to reproduce them. When the students can produce the four tones, tonal instruction concludes. This method fits the dominant view that perception is better than production (Dinnsen 1983), i.e. perception leads to production. In second language acquisition, a

Traditional Chinese category	Label	Pitch value	Tone contour	Mandarin example	English meaning
Yinping	T1	55	Level	tāng	soup
Yangping	T2	35	Rising	táng	candy
Shangsheng	T3	214	Dipping	tǎng	lie down
Qusheng	T4	51	Falling	tàng	burning hot

Table 1: Four tones in Mandarin

basic tenet is that production is based on perception too (Flege 1995). A study of tones indicated that perceptual training can improve production (increases of 5%) even without special production training (Wang et al. 2003). During the past twenty years, however, some studies have shown that tonal perception and production are not parallel in Mandarin for non-native speakers (Leather 1990, Elliot 1991, Chen 1997, Sun 1997). The experience of some Chinese language teachers is that some students cannot perceive tones correctly even though their production is understandable. At the heart of this issue is the fact that there is currently no proof that production is indeed better than perception at the suprasegmental level. On the other hand, there is evidence that at the segmental level, production can be better than perception when it comes to L2 acquisition, such as native Japanese learners of English and native Arabic learners of English (Dissosway-Huff 1981, Sheldon and Strange 1982, Port and Mitleb 1980). The relationship between tonal perception and production remains a controversial conundrum.

2 Literature review

A tremendous number of studies have been carried out in the area of Chinese tonal acquisition. In particular, studies that seek to elucidate how second language learners acquire Mandarin tones have garnered attention, driving forward research on tonal perception and production. Some research has focused on the order in which tones are acquired (Kiriloff 1969, Miracle 1989, Shen 1989, Leather 1990, Elliot 1991, Chen 1997, Sun 1997). Other studies have been conducted from the perspective of phonetics (Shen 1989, Leather 1990). New technology has allowed researchers to carry out experiments from a psycholinguistic or neurophysiological perspective (Leather 1983, 1987, Stagray and Downs 1993, Halle et al. 2004, Sereno and Wang 2008).

The studies summarized in Table 2 present the overall order of tonal acquisition. It shows that on the whole, T2 is the most difficult to perceive. This could be

explained from the phonological perspective. There are two essential phonological features that are used to describe tones: the register that denotes pitch height and the contour that denotes pitch movement. Among the four tones, only T1 is a level tone that does not have any falling or rising contours. It is highly distinguishable from the other three tones that are contour tones. T2 mainly has a rising contour with a very short fall at the beginning. The length of the short fall could be used to distinguish T2 from T3 (Shen and Lin 1991). This makes T3 share both falling and rising contours with T2 without context. Meanwhile, both T1 and T2, including the beginning and the ending points, are in the high register. The secondary features of T2 (short fall at the beginning) let T2 share a falling contour with T4, and the beginning points of T2 and T4 are in the high register. T3 has completed and clear falling and rising contours, distinguishing it from other tones. T4 shares the falling contour with T3, and T1 and T4's beginning points are in the high register. It is clear that T2 shares the most features with other tone categories. Therefore, T2 is the hardest tone to acquire.

Table 2 also shows that T4 is perceived most easily compared to other tones in all but one of the studies. The reason could be that a falling contour is a default intonation in most languages.

Sun's study (1997) suggested that the order of T1, T3 and T4 is much less clear cut than other research indicates. It is important to note, however, that Sun's study is the only one that used inferential statistics, an aspect of the research that renders the results more valid.

With regards to production, five of the eight studies found that T1 is produced correctly more often than other tones while no single tone stands out as the most difficult to produce.

According to Table 2, one underlying result is that the acquisition order of tones is different in perception and production.

Research has been carried out on the relationship between perception and production (Leather 1990, Elliot 1991). Leather (1990) reported that the error patterns of production correlate with those in perceptual tests, and claimed that the perception and the production of tones are interrelated. Elliot (1991) discovered a moderate correlation between perception and production. Elliot also found that the relationship between perception and production is not close, especially for T3 and T4.

Research on tonal perception and production in Mandarin has produced conflicting results regarding language acquisition. Studies of acquisition order indicate there is a weak relationship between perception and production since the orders of tonal perception and tonal production are different. Studies on the relationship between perception and production, however, demonstrate that there is

Study	Mode	Order
Kiriloff (1969)	Perception	4 < 1 < 3 < 2
Elliot (1991)	Perception	4 < 3 < 1 = 2
	Perception: self	4 < 1 < 3 < 2
Sun (1997)	Perception: TIDT (stimulus)	4 < 1 < 3 < 2
	Perception: TIDT (response)	1 < 4 ≤ 3 < 2
Miracle (1989)	Production	1 < 4 < 3 < 2
Shen (1989)	Production	2 < 3 < 1 < 4
Leather (1990)	Production	1 < 4 < 2 = 3
Elliot (1991)	Production	1 < 4 < 2 < 3
Chen (1997)	Perception/production	1 < 4 < 2 < 3
Sun (1997)	Production: REPT	3 ≤ 4 < 1 ≤ 2
	Production: RDGT	4 < 1 ≤ 3 < 2
	Production: TRAT	1 < 2 ≤ 3 < 4

Note: TIDT is the acronym for tone identification task. REPT is the acronym for repetition task. RDGT is the acronym for reading aloud task. TRAT is the acronym for oral translation task. Stimulus refers to stimulus matches: the proportion of the tones in stimulus that match subjects' tone identifications. Response means response matches: the proportion of subjects' tone identifications that match the tones in stimulus.

Table 2: Relative difficulty of the four tones reported in previous studies *(adapted from Sun 1997: 196)*

a moderate relationship between the two. It is thus reasonable to conclude there is relationship between perception and production, and that they develop differently in the case of L2 acquisition.

When the problem was raised in a pedagogical context (see 1.2), it challenged the dominant view that perception is the foundation upon which production is built. Research findings from other languages provide a new perspective on the notion that perception and production can be at least partially independent at the segmental level. For example, native speakers of Japanese learning English can produce the /r/ and /l/ distinction better than they perceive it (Dissosway-Huff 1981; Sheldon and Strange 1982). Similarly, other researchers found that the production of the /p/ and /b/ distinction is better than its perception by native Arabic learners of English (Port and Mitleb 1980). Waldman, Singh, and Hayden (1978) and Paliwal, Lindsay, and Ainsworth (1983) have discussed such phenomena and supported the view that perception and production can be partially independent. However, whether there is a partial independence of tonal perception and production has not been studied. Most studies on tonal perception and production in Mandarin focused on acquisition order, lacking further explanation about why the orders of perception and production are different. Although

some studies try to explain tonal errors via acoustic analysis (Miracle 1989, Shen 1989), such analysis can only reveal the processing of production and neglect auditory performance. Few researchers have observed that tonal production may be achieved prior to tonal perception, and this phenomenon warrants further investigation.

The current research explores whether tonal production is better than perception, and if so, why. There are two possible ways to investigate the phenomenon: examining interlingual factors and examining intralingual factors. Many studies have explored the interlingual factors that may affect tonal acquisition, such as interference from English prosodic features. Some studies revealed the L1 boundary effect during tonal perception (Broselow, et al. 1987). However, researchers found that L1 does not play an important role during acquisition and it cannot help learners improve their tonal ability (such as White 1981), since the counterpart of a tone in English is intonation that is at the lexical or phrasal level. Chinese tones, by contrast, are at the syllabic level.

Some researchers claimed that the Chinese tone "may violate some universal phonetic constraint in the syllable" (White 1981: 32). Little research has examined whether the interaction between segments and superasegments within a syllable could influence tonal acquisition. In fact, many studies indicated that tones have a close relationship with segments (Hu 1987, Wu and Lin 1988, Yip 1990).

First, tones are carried by finals. Some researchers who study tonal acquisition by American learners claimed that tones are not affected by segments, as they are independent (Chen 1997). Despite this, tones are directly carried by finals. If a learner cannot pronounce a final, it is unlikely for that learner to be able to pronounce the tone that the final carries. Meanwhile, if a tone is carried by a schwa sound, generally, the tone is neutral.

Second, the relationship between segments and tones can be viewed from yet another perspective: "it is widely agreed that Chinese (and many other languages) underwent a process by that the voicing distinction on initial consonants was transformed into a tonal distinction" (Yip 1990: 46). Proto-Chinese did not have tones. In the history of the creation of tones, known as tono-genesis, tones developed from the higher-frequency sounds of voiceless initial consonants, then spread to each syllable. The development of some tones was created due to pronunciation of initial consonants. Even in modern Mandarin, few T2 syllables appear which combine the consonants b, d, z, zh, j, g, with nasal finals (Hu 1987).

All of these indicate that initials and finals have a close relationship with tones within syllables. Because of this, tonal acquisition should be observed in reference to the tonal system as well as the systems of initials and finals. As a re-

sult, this study will focus on intralingual factors including initials, finals and tones.

3 Research methods

3.1 Research questions

Three research questions are raised in this study.

1. Do perception and production develop in parallel? If so, does the degree of parallelism differ according to learners' proficiency?
2. Do initial and final variations affect tonal perception according to tonal variations? If so, how do they affect perception?
3. Do initial and final variations affect tonal production according to tonal variations? If so, how do they affect production?

3.2 Factors studied

There are five factors (variables) in this research: linguistic task, proficiency level, tone, initial and final. There are two main linguistic tasks: a perception task (tone-identification) and a production task (reading aloud). There are two proficiency levels in this research: level 1 – the students learning second-semester Chinese, and level 2 – the students learning fourth-semester Chinese. There are four tonal variations: level tone (T1), rising tone (T2), dipping tone (T3), and falling tone (T4) (Table 1).

Since the purpose of this study is to observe intralingual factors that may affect tonal perception and production, it is important to design syllables that are representative of the relationship among initials, finals, and tones. Thus, the initials and the finals varied systematically in the research design.

Mandarin has twenty-two initials. Traditionally, these initials are divided into seven categories in terms of the place of articulation (Table 3).

Mandarin has thirty-nine finals. Traditionally, Mandarin finals are divided into three categories: mono-vowel finals, multi-vowel finals and nasal finals. A multi-vowel final has at least two vowels. Nasal finals are the only finals composed of both vowels and consonants (Table 4).

When the tonal tasks were designed, monosyllables were selected according to the place of articulation and the types of vowels.

Place	Manner				
	Label	Unaspirated stops/ affricates	Aspirated stops/ affricates	Fricatives	Voiced
Labial	LAB	b	p	f	m
Alveolar	ALV	d	t		n, l
Dental	DEN	z	c	s	
Retroflex	RET	zh	ch	sh	r
Palatal	PAL	j	q	x	
Velar	VEL	g	k	h	
Zero Initial	ZI				ø

Table 3: Initial variations

Final category	Label	Examples
Mono-vowel	MONO	a, e, o, i, u, ü
Multi-vowel	MULTI	ai, u(e)i, ao, üe
Nasal	NASAL	an, un, eng, ong

Table 4: Final variations

3.3 Subjects

Thirty-six participants were recruited from the Chinese program at a Midwest university. They were grouped by proficiency level. Twenty-five participants were second-semester students of Chinese, and 11 participants were fourth–semester students of Chinese. All the learners had identified English as their native language and Mandarin Chinese was their target language. None of them had learned tonal languages before and none of them were heritage learners.

In this study, we used native speakers' perception and production as a standard input and output form. During perception, tones produced by native speakers (NS) were recorded, and were used as stimuli for non-native speakers (NNS) to perceive. The perception results indicate how NNS perceive tones. However, during production, tones were produced by NNS and perceived by two raters who are NS. The NS' perception results reflected the accuracy of NNS' tonal production. Some research (Yang, 2010) indicated that NS perceive four tonal categories as distinct from one another. Meanwhile, five NS were asked to complete the tone-identification task (the task is described in 3.4) for this study. Their accuracy was 100%. Since the current study explores the intralingual factors, that is, observes

NNS tonal processing in terms of the tonal system as well as the systems of finals and initials, this study does not have a control group of NS.

3.4 Procedure

The study assigned participants two main tasks, one involving a tone-identification task, and the other reading aloud. The two tasks were counterbalanced: half of the subjects began with the tone-identification task, while the remainder began with reading aloud.

An instrument containing eighty-four syllables was created to elicit data for this study (see Appendix 1). The design of the eighty-four single monosyllables considered all the variations of initials and finals defined in 3.2. First, seven initial variations depending on the place of the articulation were combined with three kinds of finals, resulting in 21 kinds of syllables. Four tones were then distributed to each kind of syllables for a total of eighty-four target syllables.

Before the tone-identification task, 84 monosyllables written in *Pinyin* were read by a native Mandarin speaker, and the pronunciations were recorded. This audio recording of the stimulus syllables was used as an elicitation instrument in the tone-identification task, as were answer sheets for participants to write down the tones of eighty-four syllables they heard.

The reading task presented participants with the 84 syllables written in *Pinyin* Romanization, in an order different from that of the tone-identification task. Subjects were asked to read the syllables aloud, their pronunciations were audio recorded, and then the tones were transcribed in *Pinyin*. Two native Mandarin speakers listened to the tape and transcribed the recording. 93% of results from the two raters are identical. The remaining 7% was rated by another native speaker before attaining final scores.

Why were subjects only asked to mark tones, without writing down *Pinyin* symbols of initials and finals? Kiriloff (1969) conducted three tests in his study. In the first and second tests, students were asked to transcribe sounds with *Pinyin* symbols as well as identify tones. In the third test, students only needed to identify tones. The scores in the third test were better than those in the previous tests. Since the current study focuses on tonal perception, it was important to avoid disturbance from other factors, such as the task of recalling and transcribing *Pinyin* symbols. Therefore, subjects were only asked to mark tones during the identification task.

Meanwhile, if tonal errors are caused by the initial variations or final variations, the distribution of tonal perception errors in terms of initial and final

variations will be not even, i.e., tones carried by some types of initials or finals will be perceived more inaccurately than others. If a tone is perceived incorrectly, it does not matter whether the initial or final in this syllable is perceived correctly. In other words, whether a segment is perceived correctly or not will not change the fact that a segment in the syllable affects the tonal perception. The same thing happens to production.

3.5 Statistical analysis

In the current research, the tasks sought to identify the factors that relate to tonal perception and production by American learners of Mandarin. An analysis of variance (ANOVA) was the main statistical method employed. The study also utilized a Pearson Correlation.

4 Results

4.1 Results for RQ1: Perception and production are correlated, but proficiency level makes no difference

According to both the identification task and the read-aloud task, the descriptive data indicates that tonal perception and production are strongly correlated (Table 5). The mean perception and production scores for each participant are calculated by Pearson Correlation at different levels. The Pearson Correlation between production and perception at level 1 is 87%, and the Pearson at level 2 is 76%. This indicates that perception and production are strongly correlated based on each learner's performance (Figure 1).

The results from ANOVA (Table 6) indicate that the level main effect does not differ significantly, which means that the mean score of subjects at level 1 is al-

	Perception		Production		Correlation	
	Mean	SD	Mean	SD	Pearson	*p*-value
Level 1 (n = 25)	59.52	16.46	70.96	13.63	0.87	0.00
Level 2 (n = 11)	57.36	12.62	69.10	13.87	0.76	0.01
All levels Total (n = 36)	58.86	15.24	70.39	13.53	0.84	0.00

Table 5: Mean scores for task * level (in %)

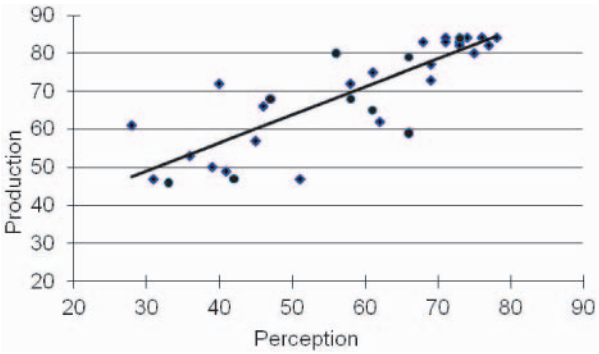

Fig. 1: Scatter plot of production and perception by individual subject

most the same as the mean score of subjects at level 2. The perception/production main effect differs significantly, which means that the perception mean and the production mean differ significantly. The level and perception/production interaction does not differ significantly, which means that the perception/production patterns of magnitude difference are similar regardless whether a participant is level 1 or level 2.

		SS (Type III)	df	MS	F	*p*-value
Within-subjects	Perception/Production	2049.98	1	2049.98	56.07	0.000
	Perception/ Production*Level	0.32	1	0.32	0.01	0.93
	Error	1243.17	34	36.56		
Between-subjects	Level	61.89	1	61.89	0.16	0.69
	Error	13233.48	34	389.22		

Table 6: ANOVA summary table for task * level

The results indicate that mean scores of perception and production differ. Although there is no difference in perception or in production according to the proficiency level, the mean scores of perception and production differ at each level (Figure 2). The production mean score is significantly larger than the perception mean score. The tonal production performance is better than the tonal perception performance.

Since the results show that there are no significant differences between level 1 and level 2, the data of level 1 and level 2 are combined to conduct further analyses.

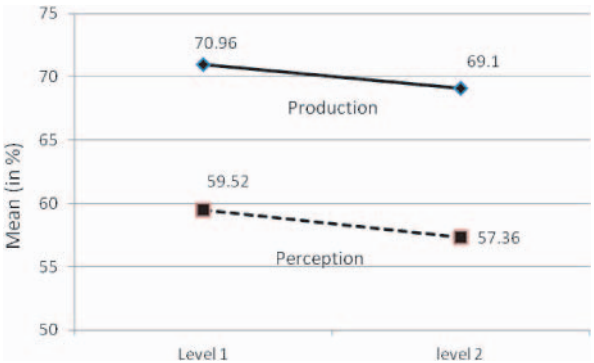

Fig. 2: Mean scores of perception and production (in %)

4.2 Results for RQ2: Initial variations and final variations affect tonal perception

The results of the tone-identification task indicate that initial variations and final variations do affect tonal perception according to tonal variations.

(a) Do initial variations affect tonal perception according to tonal variations?

Learners' perception of tones differs significantly by initial variation, learners' perception of tones differs significantly by tonal variation, and the effects of tonal variation differ significantly according to initial variation.

Student no. = 36	Tone1		Tone2		Tone3		Tone4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
LAB	2.33	.16	1.56	.16	2.50	.14	2.36	.16
ALV	2.06	.16	1.42	.18	2.36	.12	2.06	.16
DEN	2.28	.15	1.33	.17	1.92	.16	2.33	.15
RET	2.50	.16	1.81	.14	2.22	.14	2.31	.16
PAL	2.56	.13	1.61	.17	2.33	.12	2.02	.19
VEL	2.47	.14	1.61	.17	2.08	.14	2.14	.18
ZI	2.50	.14	1.53	.17	2.14	.14	2.53	.14

Table 7: Perception mean scores for tonal * initial variations (mean scale from 0 to 3)

Table 7 shows the descriptive results for perception according to each initial type. Every mean in the table indicates the number of correctly perceived tones within a specific type of initials. For example, the second column shows the mean number of correctly perceived T1, and the third row shows the means of correctly perceived tones in which the syllable has labial initials. The results indicate that T2, regardless of initial types, has the least perception accuracy.

A two-dimensional mixed ANOVA analysis of the tonal perception scores was conducted. In the 'Tonal Variation X Initial Variation' design, the tonal variation and initial variation serve as the within-subjects factors. The results indicate that the tonal variation effect, $F(3, 102) = 21.024$, $p < 0.0001$, the initial variation effect, $F(6, 210) = 3.709$, $p = 0.002$, and Tonal Variation X Initial Variation, $F(18, 630) = 2.312$, $p = 0.002$, are significant.

The follow-up analysis is the pairwise comparison. It shows that the tonal perception mean of T2 is significantly lower than other tones at the 0.05 level. The perception mean of alveolars is significantly lower than that of labials at the 0.05 level. The perception mean of alveolars is significantly lower than that of retroflexes at the 0.05 level.

The results specify that T2 is much harder for English-speaking students of Chinese to perceive than other tones. As for the initials, the tones in the syllables that have alveolar consonants are more challenging for students to perceive, while the tones in the syllables that have labial consonants and retroflex consonants are easier for students to perceive. The perceptual difficulty for the tones in which the syllables have other initials is similar or in the middle point.

(b) Do final variations affect tonal perception according to tonal variations?

Learners' perception of tones differs significantly by final variation, learners' perception of tones differ significantly by tonal variation, and the effects of tonal variation differ significantly according to final variation.

Student no. = 36	Tone1		Tone2		Tone3		Tone4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
MONO	6.17	.21	2.83	.34	6.06	.25	5.75	.24
MULTI	5.03	.34	3.22	.32	5.19	.26	5.31	.36
NASAL	5.50	.33	4.75	.34	4.17	.32	4.89	.38

Table 8: Perception mean scores for tonal * final variations (mean scale from 0 to 7)

Table 8 shows the descriptive results for perception according to each final type. Every mean in the table indicates the correct perception number of each tone that is carried by a specific final. For example, the second column shows the

T1 perception means, and the third row shows the perception means of tones carried by mono-vowel finals. The results illustrate that T2 carried by any type of finals has the least perception accuracy.

A two-dimensional mixed ANOVA analysis of the tonal perception scores was conducted. In the 'Tonal Variation X Final Variation' design, the tonal variation and final variation serve as the within-subjects factors. The results indicate that the tonal variation effect, F (3, 105) = 22.647, $p < 0.0001$, the final variation effect, F (2, 70) = 5.856, $p = 0.004$, and Tonal Variation X Final Variation, F (6, 210) = 19.846, $p < 0.0001$, are significant.

The follow-up analysis is the pairwise comparison. The results demonstrate that the tonal perception mean of T2 is significantly lower than others at the 0.05 level. The tonal perception mean of mono-vowel finals is significantly higher than that of multi-vowel finals at the 0.05 level.

From these results, we can conclude that T2 is much harder to perceive than other tones. As for the finals, the tones carried by mono-vowel finals are easier for learners to perceive while the tones carried by multi-vowel finals are harder for learners to perceive.

4.3 Results for RQ3: Initial variations and final variations do not affect tonal production

The results of the read-aloud task indicate that initial variations and final variations do not affect tonal production according to tonal variations.

(a) Do initial variations affect tonal production according to tonal variations?

Learners' production of tones does not differ significantly by initial variation, learners' production of tones differs significantly by tonal variation. But the effects of tonal variation do not differ significantly according to initial variation.

Student no. = 36	Tone1		Tone2		Tone3		Tone4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
LAB	2.61	.12	2.44	.16	2.83	.09	2.30	.16
ALV	2.67	.13	2.56	.15	2.89	.05	2.33	.14
DEN	2.61	.12	1.83	.18	2.64	.12	2.28	.18
RET	2.78	.08	2.72	.09	2.81	.08	2.00	.18
PAL	2.58	.11	2.42	.15	2.78	.08	2.31	.17
VEL	2.58	.13	2.19	.15	2.86	.06	2.25	.17
ZI	2.61	.13	2.64	.10	2.72	.11	2.14	.19

Table 9: Production mean scores for tonal * initial variations (mean scale from 0 to 3)

Table 9 shows the descriptive results for production according to each initial type. Every mean in the table indicates the number of correct productions of each tone following a specific initial type.

A two-dimensional mixed ANOVA analysis of the tonal production scores was conducted. In the 'Tonal Variation X Initial Variation' design, the tonal variation and initial variation serve as the within-subjects factors. The results indicate that the tonal variation effect, F (3, 105) = 10.371, $p < 0.0001$. There is no significant production mean difference among initials at the 0.05 level.

The follow-up pairwise comparison of tones elucidate that the tonal production mean of T1 is significantly higher than that of T4 at the 0.05 level. The tonal production mean of T3 is significantly higher than that of T2 at the 0.05 level. The production mean of T3 is significantly higher than that of T4 at the 0.05 level. The results indicate that the mean scores of producing T2 and T4 are lower than those of T1 and T3.

(b) Do final variations affect tonal production according to tonal variations?

Learners' production of tones does not differ significantly by initial variation, learners' production of tones differs significantly by the tonal variation, and the effects of tonal variation do not differ significantly according to final variation.

Student no. = 36	Tone1		Tone2		Tone3		Tone4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
MONO	6.19	.23	5.56	.31	6.56	.14	5.17	.38
MULTI	5.94	.27	5.64	.28	6.58	.18	5.56	.30
NASAL	6.31	.22	5.61	.27	6.39	.21	4.89	.40

Table 10: Production mean scores for tonal * final variations (mean scale from 0 to 7)

Table 10 shows the descriptive results for production according to each final category. Every mean in the table indicates the correct number of the production of each tone that is carried by a specific final type.

A two-dimensional mixed ANOVA analysis of the tonal production scores was conducted. In the 'Tonal Variation X Final Variation' design, the tonal variation and final variation serve as the within-subjects factors. The results indicate that the tonal variation effect F (3, 105) = 10.371, $p < 0.0001$ is significant. There is no significant production mean difference among finals at the 0.05 level.

The follow-up analysis is the pairwise comparison of tonal variation. The results are identical to those discussed in the section 4.3a regarding initial variations in production.

4.4 Summary of results

Tonal perception and production are strongly correlated. However, the performance of tonal production is significantly better than perception. Unexpectedly, there is no significant difference between proficiency level 1 and level 2.

Both perception and production are affected by tonal variation. In the perception task, the perception of T2 is significantly lower than other tones. In the production task, the production of T1 is significant higher than that of T4; the production of T3 is significantly higher than that of T2; the production of T3 is significantly higher than that of T4.

Learners' perception of tones differs significantly by initial variations and by final variations. These results indicate that learners can perceive the tones that are carried by some types of initials or finals better than others. They also suggest that learners' tonal perception is affected by the initial system and the final system. However, learners' tonal production is not seriously impacted by initial variations and by final variations.

5 Discussion

Some previous research concluded that there is a moderate correlation between perception and production (Leather 1990, Elliot 1991). The current study finds a strong correlation. Nevertheless, the outcomes of this study demonstrate that tonal production is much better than tonal perception. The proficiency level makes no significant difference in either perception or production. In other words, what the second semester students do is similar to what the fourth semester students do in terms of the perception and production. A possible reason is that tones are taught at the beginning of the first semester, and little tonal instruction is conducted after that. Although students continue to produce and perceive tones when they study Chinese, it is possible that they are not receiving sufficient tonal inputs and outputs to improve their tonal perception and production significantly. Some near-native NNS reported that they noticed a dramatic shift in their ability to distinguish the four tones after they have learned Chinese for four or five years (informal conversation with near-native NNS). Further research is very much needed as to how much input and output are required to produce and perceive four distinct tones.

The strong correlation also reveals that the development of the tonal perception and production are interrelated. At the same time, variations affect perception and production differently. Why do the tonal variations, the initial variations and the final variations affect NNS' tonal perception and production differently?

In the tonal system, as we discussed in section 2, T2 shares the most features with other tone categories. Thus, it is not surprising the perception experiment demonstrated that T2 is the hardest tone for NNS to perceive. This shows that learners have acquired some tonal features that could help them distinguish tones and have awareness about tone categories, since they make more errors with T2, which is the hardest tone to distinguish. The error numbers of the other three tones are almost the same. The result is similar to Sun's study (1997). The current study indicates that the tonal features affect the perception. However, the results of production indicate that the production of T1 and the production of T3 are significantly better than that of T4; and the production of T3 is significantly better than that of T2. For production, T2 is not the most error-prone. It is no more difficult to produce than T4. This shows that the phonological features of the tonal system influence production much less than perception.

The findings regarding the influences by the initials and the finals also support the explanation that the sound system affects tonal perception. Some mean scores of the tonal perception are significantly different in terms of the initial variations and the final variations. The distribution of tonal perception errors is uneven regarding initials and finals respectively, and this suggests that initials and finals cause NNS to perceive tones differently. Thus, we conclude that an initial or final system affects tonal perception. The initial system and the final system do not affect tonal production since the error distribution is even in production, which means that no specific initial or final variations influence tonal production differently.

The place features of initials influence the tonal perception. Some scholars (Lin 2005) treat the alveolar consonants and dental consonants as one initial type since the place of alveolar consonants and that of dental consonants are very close. Thus, it is hard for NNS to distinguish the combinations that contain alveolar consonants and the four tones from those of dental sounds. However, retroflex consonants have a salient feature, that is, the tongue curls back against the palate. The feature of curled tongue helps NNS distinguish the combinations that contain retroflex consonants and the four tones from others. Therefore, the tones associated with retroflex consonants are heard more accurately than those associated with other consonants.

The structure of a final also affects tonal perception. Multi-vowel finals have a more complicated architecture than mono-vowel final. Thus, it is more difficult for NNS to perceive tones carried by multi-vowel finals and easier for them to perceive tones carried by mono-vowel finals. Although nasal finals are composed of a vowel and a nasal, there are only two kinds of nasals in the finals. Therefore, combinations of nasal finals could be a little simpler than those in the multi-vowel finals.

Thus, a contradiction appears: learners' production is better than their perception, while their perception is affected by the tonal features, initial features and final structures more systematically. The results in the current research indicate that the awareness of tone categories is not shown well through production. Why is NNS' production better than their perception?

An explanation is that tones are perceived at the phonological level and produced at the phonetic level. NNS imitate a couple of tones from their teacher or a few native speakers and generalize some phonetic features to produce. This is similar to L1 acquisition. Research on child language development revealed that imitation is an important psychological process at the early stage (Irwin 1946, Mowrer 1952, McCarthy 1960, Tomasello 2000). In our study, the imitation of tones can be accepted by NS if the tones produced by NNS are within the NS' perceptual categories of tones. However, when NNS perceive tones, it is impossible for them to imitate because they lack an underlying phonological system to categorize the individual tonal variations of different speakers. Note that imitation is generally based on a single sound or the sounds that only have small differences.

Stagray and Down (1993) found that Mandarin speakers have poor differential sensitivity because they categorize sounds that have similar frequency together to perceive tonal phonemes. English learners of Chinese, by contrast, are more sensitive to small frequency differences. Their over-sensitivity to minor pitch variations resembles beginning readers, who are distracted by graphic differences between capital and small letters, different fonts, and print versus handwriting. This indicates that NS use phonological features to distinguish tone categories. However, NNS cannot perceive tones well because they are more sensitive to the phonetic feature, that is, the small frequency differences, but lack robust phonological categories. In other words, NNS start with imitating phonetic features, while it takes longer time for them to acquire the distinctive features at the phonological level. L2 learners of Chinese produced tones simply based on the imitation of phonetic features. The features they employed to facilitate production do not involve in phonological contrastive features. On the contrary, they are phonetic differences. However, when a learner perceives a tone, s/he has to employ different phonological features to compare it with other tones in their inventory of tone categories. This comparison has to be conducted with a phonological system, even though the learner may still be sensitive to phonetic difference. Therefore, some tonal features, initial features and final structures could affect the perception according to tone categories. Thus, the learner's production is a surface process while the perception is a deeper process that reflects a growing, contrastive phonological system. It is easier for NNS to produce tones better than they perceive them, even though NNS perceive tones more systematically.

References

Broselow, E., R. Hurtig, & C. Ringen. 1987. The perception of second language prosody. In G. Ioup & S. Weinberger (eds.), *Interlanguage phonology: the acquisition of a second language sound system,* 350–361. Cambridge, MA: Newbury House.

Chao, Y.R. 1930. A system of tone letters. *La Maitre Phonetique*, 45, 24–27.

Chen, Q.-H. 1997. Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers Association*, 32(1), 21–39.

Chun, D.M. 2002. *Discourse intonation in L2: from theory and research to practice*. Amsterdam: John Benjamins Publishing Company.

Dissosway-Huff, P. 1981. Japanese difficulty with English /1/ and /r/: Importance of both perception and production research in second-language acquisition. *Research in Phonetics* 2, 227–260.

Dinnsen, D. A. 1983. *On the characterization of phonological neutralization (booklet)*. Bloomington, IN: Indiana University Linguistics Club.

Elliot, C.E. 1991. The relationship between the perception and production of Mandarin tones: An exploratory study. *University of Hawai'i Working Papers in ESL*, 10(2), 177–204.

Flege, J. 1995. Second-language Speech Learning: Theory, Findings, and Problems. In W. Strange (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-language research*, 233–277. Timonium, MD: York Press.

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

Hu, Y. 1987. *Xiandai Hanyu. [Modern Chinese]*. Shanghai: Gaodeng Jiaoyu Chubanshe.

Irwin, O. C. 1946. Infant speech: vowel and consonant frequency. *Journal of Speech Disorders*, 11, 123–125.

Kiriloff, C. 1969. On the auditory perception of tones in Mandarin. *Phonetica*, 20, 63–64.

Ladefoged, P. 2006. *A course in phonetics*. Boston: Thomson Wadsworth.

Leather, J. 1983. Speaker normalization in perception of lexical tone. *Journal of Phonetics* 11, 373–382.

Leather, J. 1987. The interrelation of perceptual and productive learning in the initial acquisition of second-language tone. In A. James & J. Leather (eds.), *Second-language speech: Structure and process*, 75–101. Berlin: Mouton de Gruyter.

Leather, J. 1990. Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In J. Leather & A. James (eds.), *New sounds 90: proceedings of the 1990 Amsterdam symposium on the acquisition of second language speech*, 72–97. University of Amsterdam.

Lin, Y. 2007. *The sounds of Chinese*. Cambridge: Cambridge University Press.

McCarthy, D. 1960. Language development and language disorders: a compendium of lectures. *Monographs of the Society for Research in Child Development* 25(3), 5–14.

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

Mowrer, O. H. 1952. Speech development in the young child. *Journal of Speech Hearing Disorders* 17, 263–268.

Norman, J. 1988. *Chinese*. Cambridge: Cambridge University Press.

Paliwal, K., D. Lindsay & W.A. Ainsworth. 1983. Correlation between production and perception of English vowels. *Journal of Phonetics 11*, 77–83.

Port, R. & F. Mitleb. 1980. Phonetic and phonological manifestations of the voicing contrast in Arabic-accented English. *Research in Phonetics* 1, 137–162.
Sheldon, A. & W. Strange. 1982. The acquisition of /r/ and /1/ by Japanese learners of English: evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3, 243–261.
Shen, X.-N. S. 1989. Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association* 24(3), 27–47.
Shen, X.-N. S., & M. Lin. 1991. A perceptual study of Mandarin tones 2 and 3. *Language and Speech* 34, 145–156.
Sun, S. H. 1997. *The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese.* Honolulu, HI: University of Hawai'i Press.
Sereno, J. A., & Y. Wang. 2008. Behavioral and cortical effects of learning a second language: The acquisition of tone. In O. Bhon & M. Munro (eds.), *Language experience in second language speech learning: In honor of Manes Emil Flege*, 239–258. Amsterdam: John Benjamins.
Stagray, J. R., & D. Downs. 1993. Differential sensitivity for frequency among speakers of a tone and a non-tone language. *Journal of Chinese Linguistics* 21(1), 143–163.
Tomasello, M. 2000. Do young children have adult syntactic competence? *Cognition Linguistics* 74(3), 209–253.
Waldman, F. R., S. Singh, & M.E. Hayden. 1978. A comparison of speech-sound production and discrimination in children with functional articulation disorders. *Language and Speech* 21, 205–220.
Wang, Y., A. Jongman, & J. Sereno. 2003. Acoustic and perceptual evaluation of Mandarin tone production before and after perceptual training. *Journal of the Acoustical Society of America* 113(2), 1033–1043.
White, C. 1981. Tonal perception errors and interference from English intonation. *Journal of the Chinese Language Teachers Association* 16, 27–56.
Wu, Z. & M. Lin. 1989. *Shiyan Yuyinxue Gaiyao* [Introduction to experimental phonetics], Shanghai: Gaodeng Jiaoyu Chubanshe.
Yang, B. 2010. *A model of Mandarin tone categories – a study of perception and production.* Doctoral dissertation, the University of Iowa, Iowa City.
Yip, M. 1990. *The tonal phonology of Chinese.* New York: Garland Publishing.

Appendix: Eighty-Four Syllables

1 bāo	2 tuō	3 húi	4 jiǔ	5 tè	6 xióng
7 hǎn	8 dūn	9 qù	10 hā	11 bǐn	12 bà
13 rén	14 tǒng	15 kún	16 xiǎn	17 pō	18 pàn
19 mǐ	20 kuài	21 jùn	22 jiāo	23 shēng	24 xǐ
25 dié	26 xià	27 fá	28 quān	29 nǚ	30 kuā
31 cí	32 qué	33 hòng	34 néng	35 lā	36 măi
37 qí	38 chuàn	39 zhǎng	40 gú	41 kù	42 guāng
43 năi	44 fēng	45 cǔ	46 gě	47 zī	48 máng

49 fèi	50 zhǔ	51 jū	52 lüè	53 zǔn	54 póu
55 rì	56 dú	57 gǒu	58 lìng	59 ā	60 shá
61 yǐ	62 cán	63 zuǒ	64 zòng	65 chī	66 yú
67 sūn	68 yòng	69 āi	70 sè	71 chūi	72 èr
73 wēng	74 zhòu	75 wǎi	76 yuán	77 ráo	78 yǐn
79 shuǎi	80 yè	81 cài	82 yá	83 sūi	84 cóu