THE RELATIONSHIP BETWEEN THE PERCEPTION AND PRODUCTION OF MANDARIN TONES: AN EXPLORATORY STUDY

CHRISTINE E. ELLIOTT
University of Hawai'i

This study looks at the relationship between the perception and production of Mandarin tones by non-native speakers studying Mandarin. This issue is first examined in terms of the relationship between the perception and production of speech, in general, and then in terms of Mandarin tone *perception* and Mandarin tone *production*.

The subjects (N=33) were students studying Mandarin Chinese at the University of Hawai'i and were both native and non-native English speakers who neither spoke or heard Mandarin regularly outside of the classroom. Subjects were administered a three part "Mandarin Tone Test" focusing on a) tone production, b) tone perception, and c) subject perception of their own tone production.

85% of the subjects perceived tones better than they produced the same tones. Further, subject perception errors appeared to be of two primary types: a) mis-identification of a tone and b) identification of a tone according to incorrect tone features/boundaries. Overall, students showed global awareness of their tone production errors based upon subject tone "ratings". Finally, subject production, perception, and whole test mean scores increased corresponding to subject course-level (100-, 200-, 400-).

> The extreme difficulty English speakers experience in learning to speak tonal languages is widely attested. Any student who has struggled through three years of Chinese has probably been heard to say, at one time or another, that he [or she] would never learn the tones.
>
> (White, 1981, p. 27)

As such a student I have often felt equally perplexed by this "dilemma" of tones and have questioned my seeming inability both to perceive and produce tones accurately. I have been further frustrated by my seeming inability to produce certain tones to the satisfaction of my native-speaking teachers even while "consciously knowing" what tone I was trying to produce. This would

University of Hawai'i Working Papers in ESL, Vol. 10, No. 2, Fall 1991, pp. 177–204.

appear to parallel Beebe's (1980) account of her interchange with "an exceptionally bright" doctoral student where Beebe heard this friend ask "What's a sarami?" when the graduate student both meant and thought she was saying "What's a salami?". Although these are simply two anecdotes they seem to reflect an experience familiar to both second language learners and teachers.

As put forth by Baldzikowski (1988), the relationship between production and perception needs to be investigated more closely in order to determine whether "there are perceptual bases for production difficulties in second-language learning and vice-versa" (p. 36). Several studies have looked at either the question of the production *or* the question of the perception of Mandarin tones. In this paper, however, I propose to look at the interrelationship (or independence) of both the perception and the production of Mandarin tones by adult non-native speakers learning Mandarin.

In terms of second language acquisition, the perception and production of speech segments and suprasegmentals is often discussed under the broad category of "interlanguage phonology"(IL). The term "interlanguage" was first proposed by Selinker (1969) to describe the separate linguistic system resulting from a learner's attempts to produce a target language (TL). Such an "interlanguage" has been further characterized both as an "approximative system" (Nemser, 1971) and a "dynamic continuum" (Corder, 1971) underlining its transitional nature (from Beebe, 1980).

But as pointed out by Flege (1988) an "implicit assumption behind the notion of interlanguage phonology is that the L2 pronunciation of learners with the same L1 will evolve in a similar fashion." However, as shown by Flege, in terms of perception and production of speech this is not always the case. Explanations for this fact include differences in quantity and quality of input or experience, non-linear development such as found in L1 acquisition, or simply that the concept of 'interlanguage phonology' does not apply to the production and perception of speech segments. Further research is needed in order to corroborate or dispute these claims.

In his discussion of "phonological matching," Klein (1986) proposes three general reasons why language learners do not produce sounds with native-like

accuracy: (a) the learners can not hear the sound (perception), (b) they are unable to articulate it (production), and (c) the two factors have a cumulative effect. (p. 157) While it is known that perception plays an important role in production, the exact nature of this relationship is still unknown in terms of L1 as well as L2 acquisition. It is generally accepted, however, that the production of a sound requires the "establishment of a central phonetic representation" which contains information concerning both the perceptual "target" as well as a motor plan specifying how that target is to be achieved in speech production (Flege, 1987). Perception is not only language specific, but learners "map-out" L2 sounds through "their own first language phonetic 'chart'" (Flege, 1985, p. 10).

Flege further proposes that an essential aspect of the acquisition of L2 speech is the ability to perceive distinctions between L2 categories accurately, and to implement those distinctions through "finely controlled motoric plans" (p. 166). Therefore L2 studies examining this relationship between the perception and the production of speech provide important insight into the complex auditory, neurological, and psycho-motor processes they involve and can hopefully provide better understanding of the language acquisition process. Although the majority of the studies in this area have looked at the perception and production of speech segments rather than suprasegmental or prosodic features such as tone, Brosnahan and Malmberg (1970) propose that "the functioning of suprasegmental or prosodic features may be taken to be similar [to the functioning of phonetic features]" (pp. 204–205), and therefore perhaps cautious predictions can be made about the relationship between the perception and production of tone based upon these studies. On the other hand, Gandour et al. (1988), based upon their study of "tone in aphasia", propose that the relationship between speech perception and production may vary depending upon what aspect of speech is being investigated. They conclude, however, that further research is needed to better define this relationship.

In order to better understand the relationship between the perception and production of Mandarin tones I will now look at a) conclusions drawn from L2 studies looking at the relationship between the production and perception of segmentals (as no known studies exist focusing on the question of the

production and perception of Mandarin tone), b) studies focusing on the *perception* of Mandarin tones by adult L2 learners, and c) studies focusing on the *production* of Mandarin tone by adult L2 learners.

Several studies have been undertaken examining the perception and production of the /r/ versus /l/ contrast in English by native Japanese speakers (see Goto, 1971, Sheldon & Strange, 1982) and examining voice-onset time (VOT) of stop consonants among languages (e.g. Williams, 1979, Gass, 1984, and Flege, 1986). While the VOT studies suggest that production and perception evolve together, the /r/ versus /l/ studies have revealed a certain independence of perception and production (Flege,1988). Flege, however, cautions against drawing any conclusions concerning causality based upon these results until further research is conducted.

Even so, it is interesting to note that in both types of studies accurate L2 production was found to precede L2 perception in many cases. This parallels Ladefoged's (1967) claim for L1 development that acoustic differences in consonants or vowels can not be readily perceived until these speech segments can be articulated correctly. He continues that the reader may be surprised that people cannot generally hear differences between sounds until they have learned to make these differences, but cites a "growing body of evidence" which indicates that accurate perception does not necessarily precede production (p. 167). This general proposal, however, is contrary to many theories of child L1 acquisition where perception often does appear to precede production.

Based upon her 1982 study, Gass (1984) concludes that learner perception is "fundamentally different" from native language perception while L2 production appears more native-like though exhibiting a "lesser to greater to lesser" (U-shaped) degree of variability. However, much additional research in this area is required either to corroborate or dispute such findings. In addition, it is still unclear, if and how these results might apply to the perception and production of a suprasegmental language feature such as tone.

Gandour (1978) identifies "tone" as the "particular way in which pitch is utilized in languages." All (most) languages of the world use pitch in some fashion (Brosnahan and Malmberg, 1970). Pitch is often discussed in terms of

fundamental frequency (F_0), a measurement of the rate of vibration of the vocal cords during voice production. Mandarin Chinese is considered a tone language because pitch is primarily used phonemically to contrast individual lexical items. However, in English, an atonal language, pitch is used to convey syntactic, pragmatic, and emotional information in addition to the propositional meaning of an utterance, usually at the phrasal or sentence level. This second use of pitch is most often referred to as intonation. In both cases pitch contours are often described as *level, falling, rising,* or *dipping* and pitch register, the level at which tone production begins, as *high, mid,* or *low.*

Mandarin Chinese has four contrastive tones: Tone 1 (¯) a high level tone, Tone 2 (´) a high rising tone, Tone 3 (ˇ) a low dipping tone, and Tone 4 (`) a high falling tone. (Chao, 1968) This contrast is illustrated in the following well-known example:

mā	'mother'
má	'hemp'
mǎ	'horse'
mà	'to scold'

While it has been shown that native Mandarin speakers can "readily identify" these tones in real-speech monosyllabic words in isolation (Chuang, Hiki, Sone, and Nimura, 1972), for non-native speakers these contrasts are not necessarily apparent.

Unless a person suffers from physical deafness he is, as a general rule, capable of distinguishing between sounds uttered with different pitch configuration (Kiriloff, 1969). Broselow et al. (1983), suggest that non-native speakers' inability to distinguish Mandarin tones is due to the fact that they perceive Mandarin tones in terms of their L1. In the case of English speakers, for example, Mandarin Tone 4 is "acoustically quite similar....to the unmarked pattern for declaratives,...single word utterances, and...the final item in a series"(p. 353). And as pointed out by Lin (1985) based upon three years of observation, especially beginning learners are "perceptually biased" toward English intonation patterns and will be listening for similar intonation patterns in Mandarin utterances. This, in turn, places a heavy "learning load" on

speakers of atonal languages who must "learn to perceive relative pitches which differentiate meaning"(p. 32).

Further, it has been proposed that certain tones are more difficult to perceive than others both for non-native speakers and native-speaking children acquiring Mandarin as their L1. Li and Thompson (1977) found that Mandarin-speaking children confused Tone 2 and Tone 3 into "the later stages of the tone acquisition process," and Kiriloff (1969) found that adult L2 learners of Mandarin made "considerably more" errors identifying Tone 2 and Tone 3 than Tone 1 and Tone 4. It is hypothesized that Tone 2 and Tone 3 are more difficult to perceive and produce due primarily to their similarities in pitch shape (both are predominantly rising but differ in pitch register). Research continues to be conducted in order to better define distinguishing characteristics of these tones. (see, for example, Blicher, Diehl, and Cohen, 1990)

Based on his review of pertinent studies, Gandour (1978) concluded that preliminary findings indicate that perception of tone by native speakers is categorical (i.e. small steps along an acoustic continuum produce perceptible differences when they occur between phonetic categories but not when they occur within a phonetic category) as opposed to a continuum where discrimination is independent of category assignment. Gandour cautions, however, that numerous problems remain in the interpretation of categorical perception experiments on tone.

Though several studies have examined the *production* of Mandarin tones, it is necessary to remain cautious in interpreting their results as the conclusions have been based upon a small number of subjects (4 to 10). An interesting L2 study by Neufeld (1978), however, (based upon data collected from 100 subjects) found that native English speakers (with no experience studying Mandarin) were able to acquire "native" or "near-native" production proficiency in a laboratory setting where subjects were asked to mimic a native-speaker and could monitor their progress.

While it therefore seems possible for NNSs to produce accurate tones in a highly controlled environment where the focus is solely on tones, research examining the study of the Mandarin language in a classroom environment

indicates that students are much less successful achieving "native-like" accuracy. In fact, Miracle (1989) concluded, "...students of Chinese need desperately to pay more attention to their tone production"(p. 56). Although data on NNS acquisition of tones are still inconclusive, Chiang (1979), Shen (1989), and White (1981) all propose that errors in the production of Mandarin tones by native English speakers are due to interference from English pronunciation either through disregard of the phonemic use of tone at the syllable level or transfer of L1 intonation patterns into L2.

Shen (1989) asserts that transfer plays an important role and further concludes that the misproduction of tones "generally lies in pitch register [low, mid, high] and not in pitch shape [rising, falling, dipping, level]"(p. 40) as Mandarin has a significantly wider pitch range than English. In addition, Shen cites differences in the "mechanisms of stress" between the two languages as a probable source of difficulty. According to White (1981), stress in English is associated with pitch height and because of this the English speaker will *hear* Tone 1 (¯) as stress and therefore re*produce* it in terms of English stress which is very similar to Mandarin Tone 4 (ˋ).

Finally, a factor which would appear to influence both the perception and production of segmentals and suprasegmentals is the learner's language experience. Although age of initial exposure to an L2 is considered crucial in relation to child L2 acquisition (see Long 1990), "experience" would also appear to play an important role, especially in exploring adult L2 acquisition. Unfortunately, at least in studies focusing on production and/or perception, neither the critical/sensitive period hypothesis nor "experience" has been clearly defined or controlled for.

In summary, the literature would indicate that while NNSs are physiologically able to produce Mandarin tones in a controlled environment focusing only on tone, in actual L2 learning situations the learner must deal not only with the problems of tone perception and production, but also with other suprasegmentals, segmentals, syntax, and (let us not forget) the conveyance of meaning. In addition, transfer of L1 suprasegmental patterns (at a relatively unconscious level) further complicates the perception and production tasks with which the learner is faced. While several studies looking at segmental aspects of speech have shown that production can, in fact, precede perception,

more research is needed in order to determine more precisely this relationship. Finally, it is proposed that learner awareness of perception and production (through self-monitoring and/or various types of external feedback) is a crucial component in the approximation and/or acquisition of target-language phonology and should be incorporated into future studies.

Therefore, as no known studies exist focusing specifically on the relationship between the production and perception of Mandarin tones by non-native speakers, the purpose of the present study is to initiate an exploratory investigation in order to better understand this particular relationship.

The major questions to be addressed are as follows:

Is there a significant correlation between the **perception** and **production** of Mandarin tones by non-native speakers?

> Are the frequency and distribution of tone **production** errors among the four tones the same as/similar to the frequency and distribution of tone **perception** errors?

To what degree are learners able to **monitor** and/or **reflect upon** their own perception and production of Mandarin tones?

Method

Subjects

Data were initially elicited from approximately 50 students studying Mandarin Chinese in classes at the University of Hawai'i. The data of 33 of these subjects were found suitable for analysis. These 33 subjects did not regularity speak or hear Mandarin outside of the classroom and completed all parts of the data collection process according to the directions given. The subjects were students in 100-, 200-, and 400-level Mandarin Chinese courses and were selected to participate in this study based upon the cooperation of

their instructor. Both male and female, graduate and undergraduate students participated in the study. All students had completed at least one semester of Mandarin prior to their participation in this study, and the amount of previous, explicit "tone training" that students had received in the classroom varied greatly depending course level and individual instructors. While this study initially intended to focus solely on native-English speakers learning Mandarin, due in part to the ethnic diversity of Hawai'i, participating students came from a variety of language backgrounds (See Table 1).

Table 1: Native Language

Language		# of students
(atonal)	English	12
	Japanese	6
	Marshallese	1
		19
(tonal)	Cantonese	7
	Other Chinese dialects	6
	Vietnamese	1
		14

All students, however, were non-native speakers of Mandarin and were placed in language classes at the University of Hawai'i based on the Department of East Asian Languages Chinese placement test.[1]

Materials

In order to gather data from the subjects pertaining both to their perception and production of Mandarin tones, two primary instruments were utilized: a) a questionnaire, and b) a "test" consisting of three parts: i) production of tones, ii) perception of tones as produced by a native Mandarin speaker, and iii) subjects' perception of their own tone production.

[1] The Department of East Asian Languages Chinese Placement Test has yet to be statistically validated. Students have the option of taking the department placement test or may present evidence of prior study and/or proficiency in Mandarin Chinese for use in placement.

Questionnaire. The questionnaire (See Appendix A) was designed to elicit information from the subjects concerning their language background as well as their own opinion of their production skills. Subjects were asked to rank the four tones according to how difficult they felt it was to consistently produce each tone accurately in relationship to the other tones.

Production Stimulus Material. The production portion (Part I) of the "Mandarin Tone Test" (MTT) required subjects to read a list of 20 items. These items consisted of four phonetic representations of one-syllable words ([ma], [man], [tʃa], [tʃan]) which naturally occur in Mandarin Chinese in all four tones. Thus, 16 different lexical items representing four examples of each tone were generated. All items contained the same vowel sound /a/ in order to eliminate the possibility that a change in vowel sound be mistaken by the subjects for a change in tone. Further, one additional token of each tone was chosen from this group; two of these tokens were placed at the beginning of the production stimulus list and two at its end in order to insulate the data to be analyzed and hopefully reduce error caused by situational and/or cognitive factors. These 20 items constitute the production stimulus list (PSL). (See Appendix B)

Perception Stimulus Material Part II (Perception of a NS) and Part III (Self-Perception) of the MTT required two different types of perception stimulus material. The stimulus material for Part II consisted of a master tape recording of a female native-speaker of Mandarin from Beijing, China. She was recorded reading the list of 20 items described above arranged in reverse order to minimize the chance of the subjects' using order as a memory cue. (See Appendix C) The perception stimulus material for Part III consisted of the students' individual recordings of their own production (based on the production stimulus list of 20 items described above).

Students perception results were captured on two answer sheets, one for Part II (See Appendix D) and one for Part III (See Appendix E). Students were asked to listen to each item and then circle the choice they felt was correct.

Students were also asked to make a "confidence rating" (See Appendix D and E) based upon how confident they felt the answer they had chosen was correct. It was soon determined that most students did not have enough time to accurately complete this section of the answer sheet, and therefore these data are not analyzed.

Procedure

Data were elicited from students in a language laboratory setting during their regularly scheduled Mandarin Chinese class. Students were familiar with the language laboratory as they were required to use it as part of their regular language course. The MTT was administered several times over a period of two months to groups ranging from approximately six to sixteen students.[2] The average time for administering the entire MTT including questionnaire averaged just under 30 minutes.

Part I (Production) of the MTT was administered first. Students were seated in carrels in the language laboratory and copies of the production stimulus list were distributed. The subjects were asked to read the stimulus list aloud, pausing briefly (3–4 seconds) between items. Their production was recorded on a normal bias audio cassette compatible with the language laboratory equipment. Upon completion of Part I the production stimulus list was collected.

Subjects were then asked to fill out the questionnaire, enabling the researcher to make technical preparations for the perception portion of the test as well as to increase the amount of time which would pass before the students would listen to their own production tapes.

Part II (Perception of Native-Speaker) of the MTT was then administered. Answer sheets were distributed, and all students listened to the master tape of the native Mandarin speaker simultaneously. The answer sheet for Part II was collected and the answer sheet for Part III (Subjects' Perception of their own Production) was distributed. All students listened to their own production tape and marked their answer sheets accordingly.

[2] Logistical constraints prevented the administering of the MTT to equal-sized groups over a limited period of time. Briefly, these constraints included limited access to language laboratory facilities, scheduling conflicts, unequal numbers of students at each instructional level, and the need for the co-operation of Mandarin language instructors.

Results

The master tape as well as individual student production tapes were rated by three native speakers of Mandarin who were students in the ESL Department at the University of Hawai'i. All raters were female and had had varying degrees of instruction in linguistic analysis. Inter-rater reliability was found by calculating Pearson product-moment correlation coefficients for the data and then adjusting for having three sets of ratings using the Spearman-Brown prophecy formula. (See Table 2)

Table 2: Inter-Rater Reliability

	R2	R3
R1	.95	.90
R2	—	.93
R3	—	—

Table 3 further shows that the majority of rater disagreement was between Tone 2 and Tone 3.

Table 3: Tones most often disagreed upon by raters (% of total disagreement)

Tone	R1-R2	R1-R3	R2-R3
2-3	88	88	86
1-2	2	8	6
1-3	4	2	3
1-4	4	2	3
3-4	2	0	2

The raters were in 100% agreement (Pearson product-moment correlation=1.0) in identifying the tones produced by the native speaker on the master tape.

Student perception tests were scored according to the number of correct responses divided by the total number of items. Correct answers were determined by native-speaker raters, both for the master tape and individual subject tapes. As can be seen in Table 4, Part I (Prod) and Part III (PerSelf) both exhibit a peaked distribution with very similar central tendencies, while Part II (PerNS) is more negatively skewed. Viewed as a whole, the MTT approaches normal distribution.

Table 4: MTT Descriptive Statistics

N=33	k	$\bar{X}$	mode	med	mdpnt	range	sd
Part I (Prod.)	16	.71	.75	.75	.63	.63	.14
Part II (PerNS)	16	.85	1.00	.88	.72	.57	.17
Part III(PerSelf)	16	.71	.75	.75	.75	.70	.14
MTT (Total)	48	.76	.75	.79	.69	.47	.12

Split-half and K-R20 reliability for the MTT were calculated at R_{xx}=.75. The standard error of measure is .06.

When grouped by instructional level the subjects' mean MTT scores increase correspondingly (See Table 5). It can also be seen that mean perception scores peaked as subject course-level increased while mean production scores remained equidistantly dispersed. This could indicate that perception improves markedly after only 100-level instruction whereas production continues to improve but at a more gradual rate. These results should be interpreted cautiously, however, due to the uneven number of subjects at each level.

Table 5: Distribution of mean scores among course levels

Course	N	MTTPart I	(Prod)	Part II(Per)
100	7	.69	.62	.73
200	21	.76	.70	.87
400	5	.81	.79	.91

Pearson Product-Moment correlation coefficients were calculated comparing overall correlation between Parts I, II, and III of the MTT (See Table 6). No comparisons between the groups of students with tonal and atonal language backgrounds could be made, however, due to the uneven distribution of these groups with respect to instructional level.

Table 6: Correlation between MTT test parts I, II, and III

		PerNS (II)	PerSelf (III)
All subjects			
(I)	Prod	.59*	.53*
(II)	PerNS	—	.28
(III)	PerSelf	—	—

*$p < .01$ (non-directional)

The number of errors made in each tone category was tabulated and converted to percentages of the overall number of tonal errors. This process was repeated identifying the pairs of tones which were confused. As can be seen in Table 7, the distribution of errors among the tones reveals that 66% of the production errors involved Tone 3, while perception of the NS errors were fairly evenly distributed among Tone 1, Tone 2, and Tone 3. The majority of self-perception errors (62%) were of Tone 2.

Table 7: Distribution of Errors Among Tones(% of total errors)

TONE	1	2	3	4	No Answer
PartI (Prod)	7	20	66	6	0
PartII (PerNS)	30	30	21	19	0
PartIII (PerSelf)	19	62	5	5	10

Looking at these errors in more detail, however, it can be seen in Table 8 that the large majority of production errors (93%), perception errors (71%), and self-perception errors (82%) involved Tone 2, especially confusion between Tone 2 and Tone 3, and between Tone 2 and Tone 1. In addition, there was some confusion in perception between Tone 1 and Tone 4.

Table 8: Tones Confused (% of total errors)

Tones:	1-2	1-3	1-4	2-3	2-4	3-4	No Answer
PartI (Prod)	16	2	4	68	2	1	8
PartII(PerNS)	23	5	21	36	12	2	0
PartIII (PerSelf)	13	3	11	56	2	2	13

Tables 9, 10, and 11 show the distribution of errors among the tones for each of the test parts. While perception errors were more symmetrically distributed (for example Tone 1 is perceived as Tone 4 about as often as Tone 4 is perceived as Tone 1), production and self-perception errors were more asymmetrically distributed (for example much confusion between Tone 3 and Tone 2 but not between Tone 2 and Tone 3).

Table 9: Distribution of Production Errors Among Tones (% of total errors)

Produced y instead of x

Tone	1	2	3	4
1	—	12	1	3
2	4	—	65	1
3	1	3	—	0
4	1	1	1	—

Table 10: Distribution of Perception Errors Among Tones (% of total errors)

Perceived y instead of x

Tone	1	2	3	4
1	—	5	4	9
2	18	—	14	8
3	1	22	—	1
4	12	4	1	—

Table 11: Distribution of Self-Perception Errors Among Tones (% of total errors)

Perceived y instead of x

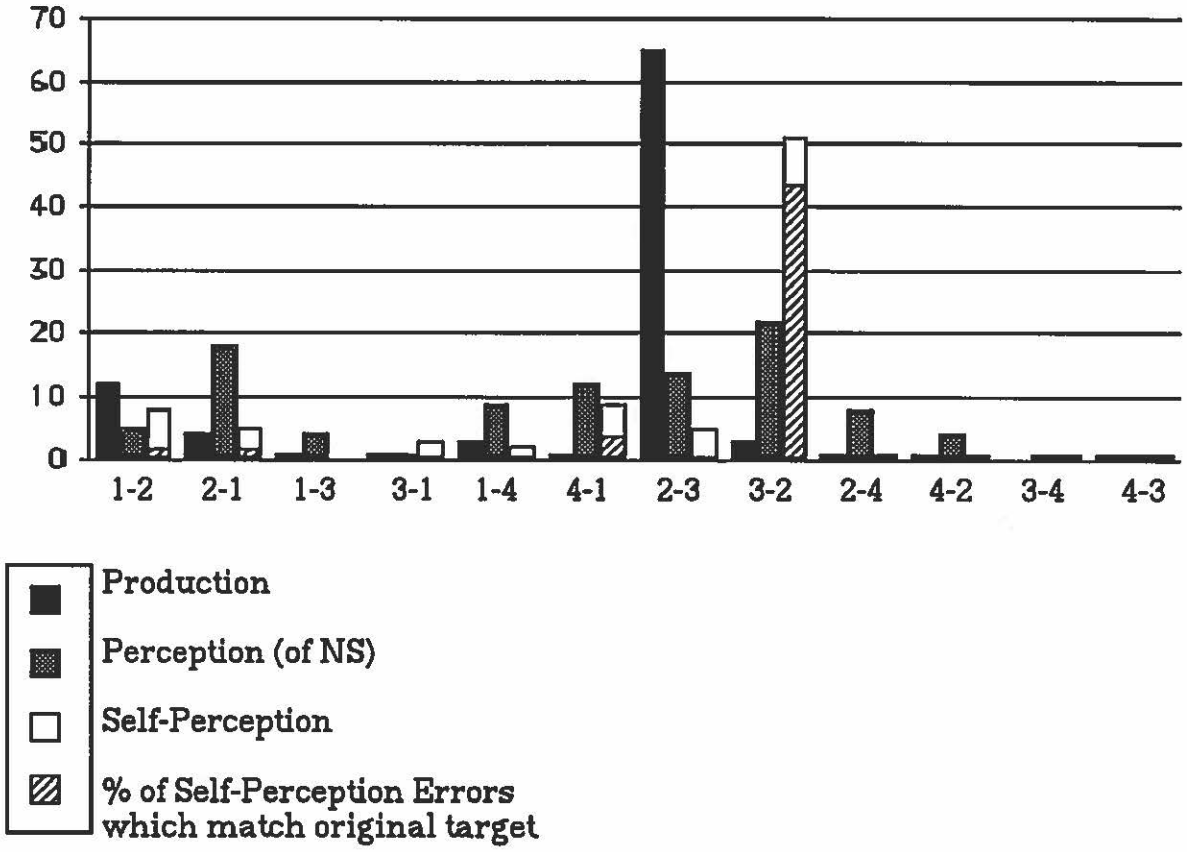
Tone	1	2	3	4
1	—	8	0	2
2	5	—	5	1
3	3	51	—	1
4	9	1	1	—

Figure 1 summarizes this information and indicates what percentage of the self-perception errors matched the original production stimulus list (PSL)

target but did not agree with NS perception as indicated through NS ratings.

Figure 1

Distribution of Tone Errors (% of total errors for each test part)



Finally, the percentage of subjects rating each of the four tones according to a four degree scale of difficulty was calculated (See Table 12). These results were then condensed into two broader categories "more difficult" and "easier." (See Table 13) Based upon subject ratings, Tone 1 and Tone 4 were judged "easier" to consistently produce accurately while Tone 2 was overwhelmingly judged "more difficult". Subjects, however, are fairly equally divided in their opinion of Tone 3 with slightly more subjects rating Tone 3 as "more difficult".

Table 12: Subject Rating of Tone Production Difficulty (% of subjects)

N=31 Tone	Most difficult	Difficult	Easier	Easiest
1	3	23	35	39
2	55	35	7	3
3	29	29	23	19
4	13	12	36	39

Table 13: Condensed Subject Rating of Tone Production Difficulty (% of subjects)

N=31	More Difficult	Easier
Tone 1	26%	74%
2	90%	10%
3	58%	42%
4	25%	75%

Discussion

The results from Part I and Part II of the MTT show that 28 out of 33 subjects perceived tones more accurately than they produced tones for the same items under identical testing conditions. Therefore, it could cautiously be concluded that non-native speakers of Mandarin have more difficulty producing tones correctly than perceiving tones correctly. This is also reflected in the fact that many 200- and especially 400- level students perceived tones with 100% accuracy while no subject produced tones with 100% accuracy. Experience also appears to play an important role with mean production and perception scores increasing with instructional level. However, there was much individual variation between production and perception scores and in fact production scores were higher than perception scores for certain subjects.

No significant correlation was found between Part I(Prod) and Part II(PerSelf) but no conclusions can be drawn based upon these results due to the skewed distribution of Part II. Further, no comparisons can be made between subjects with tonal versus atonal language backgrounds because of the uneven distribution of subjects among each course level between the two groups. This, however, is an important area for further study. Do speakers of tonal languages develop set patterns of production and perception of tone in their L1? Do L1 tonal categories/targets interfere with learning an L2 with different tone categories/targets?

Analysis of individual tone production and perception errors indicates that the majority of errors in both cases (as well as for the NS raters) involved Tone 2 (especially being confused with Tone 3). These results agree with those obtained in other L2 as well as L1 studies looking at Mandarin tone. While this similarity in kinds of errors may indicate a close relationship between production and perception errors (i.e. tones that are difficult to produce are also difficult to perceive and/or vice versa) in other cases more errors were made perceiving certain tones than producing them and vice-versa. Therefore, while certain problem areas overlap between production and perception (for example, confusion between Tone 2 and Tone 3) other areas are predominantly a perception or a production problem (For example confusion between Tone 1 and Tone 4).

Self-perception errors were distributed similarly to perception of NS errors, though subjects confused Tone 2 and Tone 3 much more frequently. While native speakers heard subjects' production of Tone 3 as Tone 2, the subjects themselves were able to identify their own production as Tone 3. This would seem to indicate that certain acoustic properties that non-native speakers consider indicative of Tone 3 (at least in their own production) are irrelevant for native speakers in terms of the classification of Mandarin tones. This is also reflected in the fact that for 81% of all the Tone 3-Tone 2 self-perception errors subjects were, in fact, "hitting" their original PSL target. Therefore, the majority of self-perception errors were not a mis-identification of tone but a "correct" identification of tone according to incorrect tone boundaries/categories. In other words, it would appear that while some subjects simply have not yet developed consistently accurate

categories/targets for production and perception of tone, others have developed them inaccurately and/or are experiencing transfer errors from their L1. This phenomenon did occur with other tone contrasts but to a much lesser degree.

It is possible that subject rating of tone difficulty also reflects this problem. While almost half of the subjects felt Tone 3 was "easier" most of their production and self-perception errors involved this tone. Subjects' ratings for Tone 1, Tone 2, and Tone 4, however, accurately reflect the percentage of production errors they made. This seems to indicate a generally high level of global awareness for their tone production. Future research should look at this area of inquiry in more detail as a step in determining how much conscious awareness learners have of their individual tone production and perception errors.

In conclusion, this study has shown that while certain tones pose both perception and production problems for the learner, other tones are production or perception errors independent of one another. Subjects' self-perception data provides additional insight, but further research is needed in order to define the nature of this relationship more clearly. Self-perception results also imply a need to make students consciously aware of their tone errors in order to solve their tone production/perception problems. Other tone errors appear to be attributable to lack of experience with or exposure to the Mandarin tone paradigm and results of this study suggest that accuracy improves with experience. Further research concerning the effects of exposure/experience on the acquisition of tone is needed. Overall, learners are able to make correct judgements concerning the accuracy of their own tone production. Future research should focus on learners' ability to assess their own tone production and perception on a more "micro" level. This will hopefully provide useful information concerning learners conscious awareness of their errors and pedagogic (or other) steps which might be taken in order to reduce/eliminate these errors.

References

Baldzikowski, M. (1988). Second language speech perception: the flip side of the phonological coin. *Texas Papers in Foreign Language Education, 1*(1), 26–40.

Beebe, L. M. (1980). Myths about interlanguage phonology. In Ioup, G. & Weinberger, S.H. (Eds.) *Interlanguage Phonology*. (pp. 165–175). New York: Newbury House.

Blicher, D., Diehl, R., and Cohen, L. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics*. *18*, 37–49.

Broselow, E., Hurtig, R. R. & Ringen, C. (1983). The perception of second language prosody. In Ioup, G. & Weinberger, S.H. (Eds.) *Interlanguage Phonology*. (pp. 350–361). New York: Newbury House.

Brosnahan, L.F. & Malmberg, B. (1970). *Introduction to phonetics*. Cambridge: W. Heffer & Sons, LTD.

Chiang, T. (1979). Some interferences of English intonations with Chinese tones. IRAL, 17(3), 245–250.

Chao, Y.R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chuang, C. K., Hiki, S., Sone, T., & Nimura, T. (1972). The acoustical features and perceptual cues of the four tones of standard colloquial Chinese. In *Proceedings of the 7th International Congress of Acoustics* (Vol. 3). Budapest: Akademial Kiado. 297–300.

Corder, S. P. (1971). Language continua and the interlanguage hypothesis. In Corder, S. P. & Roulet, E. (Eds.), *The notions of simplification, interlanguages, and pidgins, and their relation to second language pedagogy*. Geneva: Librairie Droz.

Flege, J. E. (1986). A critical period for learning to pronounce foreign languages. 162–177.

Flege, J.E. & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*. 43, 155–171.

Flege, J. E. (1987). Effects of equivalence classification on the production of the foreign language speech sounds. In James & Leather (Eds.) *Sound patterns in second language acquisition*: Foris Publications.

Flege, J.E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.) *Human communication and its disorder.* (pp. 224–401). Norwood: Ablex Publishing Corporation.

Gandour, J. T. (1978). The perception of tone. In Fromkin, V. (Ed.) *Tone, a linguistic survey*. (pp. 41–76). New York: Academic Press, Inc.

Gandour, J. T., Soranee H. Petty, and Rochana Dardarananda. (1988). Perception and production of tone in aphasia. *Brain and Language, 35*(2), 201–240.

Gass, S. (1984). Development of speech perception and speech production abilities in adult second language learners. *Applied Psycholinguistics, 5*, 51–74.

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

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica, 20*, 63–67.

Klein, W. (1986). *Second language acquisition*. Cambridge: Cambridge University Press.

Ladefoged, P. (1967). Units in the perception and production of speech. In Ladefoged, P. (Ed.) *Three areas in experimental phonetics*. (pp. 143–172). London: Oxford University Press.

Li, C. N. & Thompson, S. A. (1978). The acquisition of tone. In Fromkin, V. (Ed.) *Tone, a linguistic survey.* (pp. 271–284). New York: Academic Press, Inc.

Lin, W. C. J. (1985). Teaching Mandarin tones to adult English speakers: analysis of difficulties with suggested remedies. *RELC Journal, 16*(2), 31–47.

Long, Michael. (1990).

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.

Neufeld, G.G. (1978). On the acquisition of prosodic and articulatory features in adult language learning. In Ioup, G. & Weinberger, S.H. (Eds.) *Interlanguage Phonology*, (pp. 321–349). New York: Newbury House.

Nemser, W. (1971). Approximative systems of foreign language learners. *IRAL, 9*, 115–123.

Selinker, L. (1969). Language transfer. *General linguistics, 9*, 67–92.

Sheldon, A. & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: evidence that speech production can precede speech perception. *Applied Linguistics, 3*, 243–261.

Shen, X. S. (1989). Toward a register approach in teaching Chinese: the problems with rising tone. *Journal of the Chinese Language Teachers Association*. 24(3), 27–47.

White, C. M. (1981). Tonal pronunciation errors and interference from English intonation. *Journal of the Chinese Language Association Teachers, 16*(2), 27–56.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception and Psychophysics, 26*(2), 95–104.

Appendix A

Name _____ Seat # _____

(circle one) CHN 101 201 301 401

1. Does anyone at your home speak Mandarin?

   since you were how old? _____ until you were how old? _____

   Another Chinese dialect? Which one(s)?

   since you were how old? _____ until you were how old? _____

   A language other than Chinese or English?

2. Do you speak Mandarin regularly outside of the classroom?
   hear Mandarin regularly outside of the classroom?

3. Have you ever lived in a country where Chinese is spoken?

   Which one(s)?

   For how long?

Rank the 4 tones according to which one you think you produce most accurately, most often to the one you think you have most problems with.

a) _____ ---> "most accurately produced most often"
b) _____
c) _____
d) _____ ---> "most difficult to consistently produce accurately"

Would you like to receive your personal results from this study?
Would you like to receive the overall results of this study?

If yes, please provide your local address:

THANK YOU FOR YOUR PARTICIPATION!

Appendix B

(Production Stimulus List)

When instructed to begin please read the following words as naturally as possible. Pause for a few seconds after reading each word. Please read each word only once. If you make a mistake please continue on to the next word.

1. măn
2. chā
3. mā
4. chà
5. chā
6. mă
7. chăn
8. chán
9. mān
10. má
11. màn
12. mán
13. má
14. chān
15. chă
16. chān
17. măn
18. chá
19. chàn
20. mā

Appendix C

1. mā
2. chàn
3. chá
4. măn
5. chàn
6. chă
7. chān
8. má
9. mán
10. màn
11. mà
12. mān
13. chán
14. chăn
15. mă
16. chā
17. chà
18. mā
19. chā
20. măn

Appendix D

Name _____

(circle one)	CHN			101	201		301	401
				1=very sure	2=somewhat sure		3=guessing	
1.	mā	má	mǎ	mà	1	2	3	
2.	chān	chán	chǎn	chàn	1	2	3	
3.	chā	chá	chǎ	chà	1	2	3	
4.	mān	mán	mǎn	màn	1	2	3	
5.	chān	chán	chǎn	chàn	1	2	3	
6.	chā	chá	chǎ	chà	1	2	3	
7.	chān	chán	chǎn	chàn	1	2	3	
8.	mā	má	mǎ	mà	1	2	3	
9.	mān	mán	mǎn	màn	1	2	3	
10.	mān	mán	mǎn	màn	1	2	3	
11.	mā	má	mǎ	mà	1	2	3	
12.	mān	mán	mǎn	màn	1	2	3	
13.	chān	chán	chǎn	chàn	1	2	3	
14.	chān	chán	chǎn	chàn	1	2	3	
15.	mā	má	mǎ	mà	1	2	3	
16.	chā	chá	chǎ	chà	1	2	3	
17.	chā	chá	chǎ	chà	1	2	3	
18.	mā	má	mǎ	mà	1	2	3	
19.	chā	chá	chǎ	chà	1	2	3	
20.	mān	mán	mǎn	màn	1	2	3	

Appendix E

Name _____ slf

(circle one) CHN 101 201 301 401

1=very sure 2=somewhat sure 3=guessing

1.	mān	mán	măn	màn	1	2	3
2.	chā	chá	chă	chà	1	2	3
3.	mā	má	mă	mà	1	2	3
4.	chā	chá	chă	chà	1	2	3
5.	chā	chá	chă	chà	1	2	3
6.	mā	má	mă	mà	1	2	3
7.	chān	chán	chăn	chàn	1	2	3
8.	chān	chán	chăn	chàn	1	2	3
9.	mān	mán	măn	màn	1	2	3
10.	mā	má	mă	mà	1	2	3
11.	mān	mán	măn	màn	1	2	3
12.	mān	mán	măn	màn	1	2	3
13.	mā	má	mă	mà	1	2	3
14.	chān	chán	chăn	chàn	1	2	3
15.	chā	chá	chă	chà	1	2	3
16.	chān	chán	chăn	chàn	1	2	3
17.	mān	mán	măn	màn	1	2	3
18.	chā	chá	chă	chà	1	2	3
19.	chān	chán	chăn	chàn	1	2	3
20.	mā	má	mă	mà	1	2	3