MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 00889 7948

LIBRARY
Michigan State
University

This is to certify that the

dissertation entitled

On the Perception and Production of Mandarin Tones

by Adult English-Speaking Learners

presented by

Jya-Mei Lu

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Linguistics

Major professor

Date 8 May 1992

MSU is an Affirmative Action/Equal Opportunity Institution

O-12771

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
JUN 13 2001		
NOV 17 199?		
MAY 01 1998		
OCT 26 1998		
SEP 24 2005		

MSU is An Affirmative Action/Equal Opportunity Institution
c:\circ\datedue.pm3-p.1

ON THE PERCEPTION AND PRODUCTION OF MANDARIN TONES BY ADULT ENGLISH-SPEAKING LEARNERS

By

Jya-Mei Lu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Linguistics and Germanic,
Slavic, Asian and African Languages

1992

ABSTRACT

ON THE PERCEPTION AND PRODUCTION OF MANDARIN TONES
BY ADULT ENGLISH-SPEAKING LEARNERS

By

Jya-Mei Lu

The learning of tone is difficult for adult learners of non-tone languages learning a tone language. This study investigated how Mandarin tones are perceived and produced by adult English speakers in the first months of their learning of Mandarin.

American university students taking Elementary Chinese were tested (at the beginning and the end of the first term) on their perception and production of Mandarin tones in isolated syllables, 2-syllable words, and short sentences. The production data were judged by native Mandarin speakers. Scores of the tests were analyzed statistically.

Analyses of the results show that the perception and production of Mandarin tones were affected by tonal context and task. Tone perception and production were easier in isolated syllables than in 2-syllable words and short sentences. The production of tone was better in imitating than reading from a list.

Significant correlations between the perception of tone in isolated syllables and in sequences (2-syllable words and

short sentences) were found. However, the correlations between the production of tone in isolated syllables and in sequences were insignificant.

Significant correlations were found between the perception and the production (reading) of Mandarin tones in isolated syllables and 2-syllable words. However, the correlation between tone perception and production in short sentences familiar to the subjects was weak at the beginning and became insignificant at the end of a term.

Significant improvement in tone perception was found between the beginning of Mandarin learning and after a term of learning. However, no significant improvement in tone production was found.

Analyses of the tonal errors made by subjects show that the difficulty order of the four Mandarin tones as measured by error rates was different in different contexts at different times. The patterns of misperception and misproduction of Mandarin tones also differ depending on context.

Some suggestions of teaching Mandarin tones are proposed based on these findings.

To my parents and my husband, for their love and support.

ACKNOWLEDGMENTS

I wish to thank my guidance committee: Dr. Grover Hudson, Dr. Ruth M. Brend, Dr. Julia S. Falk, Prof. James P. Wang, and Dr. Yen-Hwei Lin, for their assistance and valuable comments. Thanks to Dr. Patricia Lunn for serving as the college reader and giving valuable comments. Special thanks to my committee chair Dr. Grover Hudson for his guidance and assistance throughout my doctoral program.

I am also indebted to Dr. Carl Lee of the Department of Mathematics at Central Michigan University for his valuable advice and assistance in the statistical analyses, to Dr. Hiou-Ming Wei and T. S. Liu for spending long hours serving as judges, and to Mrs. Betty Briggs for her assistance in recording test materials and editing production data with the language laboratory facilities.

Lastly, I am grateful to Dr. Yen-Hwei Lin and her Chinese teaching assistants for helping in conducting the tone tests in their classes, to the students of Chinese who agreed to have their tests studied, and to the Chinese speakers who served as controls.

TABLE OF CONTENTS

LIST OF TABLES xii

LIST OF FIGURES xiii

1. INTRODUCTION 1

1.1 Introduction 1

1.2 Scope of investigation 2

1.3 Outline of thesis 3

2. TONE AND INTONATION IN MANDARIN AND ENGLISH 6

2.1 Tone in Mandarin 6

2.1.1 Tone in isolated syllables (citation forms) 6

2.1.2 Tone in 2-syllable words 11

2.1.3 Tone in connected speech 13

2.2 Word stress and sentence intonation in English 30

2.2.1 English word stress 31

2.2.2 English intonation 33

2.3 Comparison of the two systems 39

3. PERCEPTION AND PRODUCTION OF SOME ASPECTS OF L2 PHONOLOGY 45
3.1 Perception and production of L2 phonology by adults 45
3.1.1 Variables affecting L2 phonology learning 47
3.1.1.1 Phonological context 49
3.1.1.2 Task 51
3.1.1.3 Sex 54
3.1.1.4 L2 experience 56
3.1.2 Relationship between the perception and production of L2 sounds in isolation and in sequence 57
3.1.2.1 Perception 57
3.1.2.2 Production 62
3.1.3 Relationship between perception and production of L2 phonology 64
3.1.4 Acquisition of L2 phonology over time . 73
3.1.4.1 Perception 73
3.1.4.2 Production 77
3.2 Perception and production of Mandarin tones by adults L2 learners 85
3.2.1 Perception 85
3.2.2 Production 89

4. ISSUES AND HYPOTHESES 95
4.1 Effects of some variables on L2 tone perception and production 98
4.1.1 Tonal context 99
4.1.2 Task 100
4.1.3 Sex 101
4.1.4 L2 experience 102
4.2 Relationship between the perception and production of tone in isolated syllables and in sequences 103
4.3 Relationship between L2 tone perception and tone production 105
4.4 Improvement of tone perception and production 107
4.5 Order of difficulty among the Mandarin tones 110
4.6 Error patterns 111

5. METHODS AND PROCEDURE 113
5.1 Tone Perception Test 113
5.2 Tone Production Test 116
5.3 Test Material 118
5.4 Subjects 121
5.5 Judges 122
5.6 Procedures 122
5.6.1 Testing 123
5.6.2 Evaluation 126
5.7 Computing the scores 129

5.7.1 Perception test 129
5.7.2 Production test 129
5.8 Statistical Analyses 131
5.8.1 Analyses of the effects of some variables on tone perception and tone production 132
5.8.2 Analysis of correlations between the perception and production of tone in isolated syllables and in sequences . 133
5.8.3 Analysis of correlations between tone perception and tone production 133
5.8.4 Analysis of improvement in tone perception and tone production 134
5.8.5 Analysis of order of difficulty among the Mandarin tones 134

6. RESULTS AND DISCUSSION 136
6.1 Test validity and reliability 136
6.2 Variables affecting tone perception and production 137
6.2.1 Tonal context 138
6.2.2 Task 141
6.2.3 Sex 143
6.2.4 L2 experience 146
6.3 Relationship between the perception and production of tone in isolated syllables and

in sequences. 150
6.3.1 Perception 150
6.3.2 Production 154
6.3.2.1 Reading task 154
6.3.2.2 Imitation task 155
6.4 Relationship between tone perception and tone production 156
6.4.1 Tone in isolated syllables 156
6.4.2 Tone in 2-syllable words 158
6.4.3 Tone in short sentences 158
6.5 Improvement over one term of learning . . . 161
6.5.1 Perception 161
6.5.2 Production 163
6.6 Order of difficulty of the four tones . . . 166
6.6.1 Perception 168
6.6.1.1 Tone in isolated syllables . . 168
6.6.1.2 Tone in 2-syllable words . . . 169
6.6.1.3 Tone in short sentences . . . 171
6.6.2 Reading 175
6.6.2.1 Tone in isolated syllables . . 175
6.6.2.2 Tone in 2-syllable words . . . 175
6.6.2.3 Tone in short sentences . . . 178
6.6.3 Imitation 182
6.7 Error Patterns 185
6.7.1 Perception 186
6.7.1.1 Tone in isolated syllables . . 186

6.7.1.2 Tone in 2-syllable words . . . 189
6.7.1.3 Tone in short sentences . . . 192
6.7.2 Reading 194
6.7.2.1 Tone in isolated syllables . . 194
6.7.2.2 Tone in 2-syllable words . . . 195
6.7.2.3 Tone in short sentences . . . 199
6.7.3 Imitation 204

7. CONCLUSIONS 207
7.1 Summary of findings 207
7.2 Suggestions for teaching Mandarin tones . . 211

BIBLIOGRAPHY 215

APPENDIX A 228
Tone Perception 228
Pronunciation 229

APPENDIX B 230
Perception Scores 230
Reading Scores 231
Imitation Scores 232

LIST OF TABLES

Table 1. Correlation between actual and predicted tonal contours 23

Table 2. Group Voice Onset Time Means for Production . 81

LIST OF FIGURES

Figure 1. Tonal variations in sentence final position . 21

Figure 2. Basic tunes of Mandarin 22

Figure 3. Median scores of two pronunciation tasks

. 53

1. INTRODUCTION

1.1 Introduction

It is generally agreed that adults learning a second language rarely, if ever, achieve a native-like accent (Scovel 1988). This seems especially true for suprasegmentals since L2 learners who speak with only a slight accent are often found to have suprasegmental rather than segmental errors (Ioup and Tansomboon 1987:344). Even in the acquisition of non-tone languages, intonation contributes to the foreign accent by which a native speaker identifies the nationality of a non-native speaker (Van Els and De Bot 1987). Not surprisingly, in learning a tone language, non-tone language learners are likely to have problems with tone. One study on the acquisition of Thai (Ioup and Tansomboon, 1987) has shown that tone is one of the last aspects of Thai grammar to be mastered by adults.

For English speakers learning Mandarin as a second language, the difficulty of tone is well attested to by both students and teachers (Kiriloff 1969, Shen 1989). Incorrect production of tone is a very apparent feature of foreign accent to native Mandarin speakers. If a Mandarin speaker is asked to speak with a foreign accent, he/she is most likely to do it by making the tones wrong. Because tone is an integral part of almost every morpheme in Chinese, incorrect production of tone is more than just a problem of

accent. Mispronunciation and misperception of tone very often cause confusion and failure in communication (Chiang 1979, W.C.J. Lin 1985).

As stated by Ioup and Weinberger (1987:319), "Research on the acquisition of suprasegmentals is a recent development in studies of L2 phonology". Up until now, there have been relatively few studies on the acquisition of tones or other suprasegmentals by L2 learners. Thus, a study on the perception and production of Mandarin tones by L2 learners can definitely contribute to our understanding of L2 phonology.

1.2 Scope of investigation

The purpose of this study is to see how Mandarin tones are perceived and produced by English-speaking adult learners (university students) in a foreign language environment (Chinese class in the United States) in the first months of their learning. More specifically, this study will try to answer the following questions:

(1) Is tone perception and production affected by variables of tonal context, task, sex, and L2 learning experience?

(2) Is there a relationship between the perception of tone in isolated syllables and in sequences, and the production of tone in isolated syllables and in sequences?

(3) Is there a relationship between tone perception and tone production?

(4) Do tone perception and tone production improve over one term of learning?

(5) Among the four Mandarin tones, is there a difficulty order for English speaking learners? If there is, what can account for the order of difficulty?

(6) Are there general patterns in errors of tone made by English-speaking adults learning Mandarin? If such patterns can be found, what accounts for them?

Because of their important implications for theories of L2 acquisition and for L2 teaching, similar questions have been asked in other areas of L2 phonology concerning L2 learners of other languages. A more detailed account of these questions and hypotheses concerning some of the questions will be given in Chapter 4.

1.3 Outline of thesis

The following is a brief description of the rest of the chapters in this dissertation:

Chapter 2 gives descriptions of the Mandarin tone system. Studies on acoustic analysis and native speakers perception of the Mandarin tones are reviewed. The use of pitch in words and sentence in English are also briefly reviewed. Differences in the use of pitch in the two languages are compared.

Chapter 3 reviews studies on the perception and production of L2 phonology by adult L2 learners in general and also the perception and production of Mandarin tones by English speakers. The former part is focused on some aspects of L2 phonological learning that are of interest for the learning of tones in this study. These include the effect of some variables (phonetic context, task, sex and L2 learning experience) on L2 phonological learning, the relationship between the learning of L2 phonology in isolated words and in connected speech, the relationship between the perception and production of L2 phonology, and the improvement of L2 phonology over time. The latter part concerns the specific difficulties encountered and the errors made by English speakers learning Mandarin tones.

Based on the literature reviewed in Chapter 3, Chapter 4 addresses each question raised in this chapter in more detail and proposes specific hypotheses in the form of expected answers to most of the questions.

Chapter 5 describes the methods and procedures of collecting the data on the hypotheses and how the data are to be analyzed.

Chapter 6 discusses the results in regard to the hypotheses proposed. The results are also compared to the learning of other aspects of L2 phonology and to other studies on the learning of Mandarin tones.

Chapter 7 summarizes the results discussed in Chapter 6 and provides suggestions for teaching Mandarin tones to L2 learners.

2. TONE AND INTONATION IN MANDARIN AND ENGLISH

2.1 Tone in Mandarin

The term Mandarin in this study is used in a narrow sense to mean the standard Mandarin Chinese based on the Beijing (Peiking) dialect. It is the language most likely to be taught in a Chinese class in a foreign language environment. In this section the tonal system of Mandarin will be described. The Pinyin romanization system is used to spell Mandarin words.

2.1.1 Tone in isolated syllables (citation forms)

The basic tone-carrying unit in Mandarin is the syllable. Almost every syllable in Mandarin (as written in the form of a character) is a morpheme/word and can be read in isolation (citation form) with a distinctive tone. Syllables consisting of the same consonant and vowel but differing in tones will be perceived by a native Mandarin speaker as different morphemes/words. (There are also many homophones in Mandarin Chinese that are written in different characters.)

There are four distinctive tones in Mandarin. They are usually numbered as Tones 1, 2, 3, and 4 (or 1st, 2nd, 3rd, and 4th tones). The characteristic pitch features of the four tones in citation forms are: Tone 1 starts high and stays high, Tone 2 rises from mid to high, Tone 3 starts

at mid-low, dips to low and rises to mid-high, and Tone 4 falls from high to low. This can be summarized in the following table using Chao's five-point tone-letter system (Chao 1930, 1948):

Tone	Description	Pitch	Tone-letters
Tone 1	high-level	55	˥
Tone 2	high-rising	35	˧˥
Tone 3	low-dipping	214	˨˩˧
Tone 4	high-falling	51	˥˩

The tone-letters without the vertical reference line (-, /, ˇ, and \) are used as tone marks when writing Mandarin in some romanization systems, including Pinyin. They are placed above the syllabic vowel of a syllable, e.g., *tān*, *tán*, *tǎn*, and *tàn*. There are a limited number of morphemes (e.g., sentence particles, grammatical morphemes, and suffixes) which do not have a distinctive tone and are pronounced in a neutral tone whose pitch is generally determined by the preceding tone (see 2.1.2 below). These morphemes are never spoken in isolation in natural speech.

The actual pitch range and interval of the tonal reference points described by Chao are relative rather than absolute, according to the sex and voice of the individual, and the mood of the moment (Chao, 1948:25). Since the

description and the notation first used by Chao were based on auditory analysis, subsequent acoustic studies (e.g., Rumjancev 1972 (according to Lyovin 1978), Howie 1976, Ho 1976a, and Tseng 1981) have tried to find the acoustic correlates of the four tones in citation forms. The acoustic studies usually measure the absolute values of the sound waves (for example, the fundamental frequency (F0) correlates to the absolute pitch values). Not surprisingly, different figures were obtained from different individuals. Variation has also been found for different syllable structures (Howie 1976, Ho 1976a). One of the variations found is a dip located at the beginning (about 15% of the duration of the syllable nucleus) of some Tone 2 syllables (Ho 1976a). Generally, however, the acoustic studies confirm the characteristic pitch contours of the four tones observed by Chao.

Experiments on tone perception in citation forms by native speakers usually result in near perfect identification. For example, in Tseng's (1981) study the syllable /i/ spoken with the four Mandarin tones and produced by a native speaker as citation forms (along with tokens obtained by editing these recorded syllables) were presented four times each in random order to four native speakers. They were correctly identified 100% of the time. Experiments on the dimensions of tone perception (e.g., vowel quality, duration, loudness) by native speakers have

also been performed by manipulating the acoustical signals. For example, in a perception test conducted by Howie (1976), the stimulus sets consisted of synthetic tone contours (based on average fundamental frequency contours obtained from native speakers) imposed on the real-speech (level) Tone 1 syllable *bao*, yielding a set of four tonally differentiated words. Chinese listeners identified the tones correctly (in agreement with the synthetic tonal contour) 95% of the time. This experiment and others have shown that pitch is the primary factor in tonal perception; other concomitant features of tone, such as vowel quality, duration, and loudness are only of marginal importance. In another test conducted by Howie, the original fundamental frequency contours in syllables differing only in tone (*yīng*, *yíng*, *yǐng*, *yìng*) were replaced with a constant fundamental frequency of 128 Hz. He found that Chinese listeners did poorly in identifying the original tones under this condition.

Tseng's (1981) study investigated the role that fundamental frequency contour and duration play in production and perception of Mandarin tones. She found that fundamental frequency contour plays the primary role in both tone production and perception. Duration is not a primary phonetic parameter since the intrinsic vowel duration in the citation form is not maintained in spontaneous speech, and the tone of a syllable (/i/ in her experiment) can be

correctly identified with 25%, 50%, even 75% edited out from the offset if the basic tone feature is left intact (e.g., Tone 1, high level tone). She concluded that, when misidentification of the tones occur as a result of cutting part of the syllable, it is "due to the absence of the particular fundamental frequency contour, not purely due to the vowel length" (p.47).

Rumjancev (according to Lyovin 1978) also conducted perception experiments which involved erasing part of a syllable produced in isolation. Although his major concern was to find which part of the syllable carries the tonal features, he also found that when the final nonsyllabic elements (e.g., nasal in *yan* or *o* in *yao*) were erased, the tones were still correctly perceived by native listeners, provided that the total duration of the nonerased portion of the syllable was between 180 and 200 ms (p.132).

Massaro, et al. (1985) investigated the evaluation and integration of pitch height and pitch contour in the perception of Mandarin tones in isolated syllables. They found that "[b]oth the F0 contour and the F0 height influenced the perceptual recognition of tone. The cues were about equally effective. As in the perception of many other contrasts, lexical tone appears to be recognized on the basis of, at least, two cues, and the influence of one cue is the largest when the other cue is relatively ambiguous" (p.267).

2.1.2 Tone in 2-syllable words

Historically, most 2-syllable words in Mandarin have evolved from the combination of two monosyllabic morphemes or words. However, 2-syllable words now occur almost as frequently as 1-syllable words in Mandarin vocabulary (M-Y. Chen, 1981). One and two syllable words are the basic units in Mandarin speech. It has been proposed that, phonologically speaking, tones in all longer words and utterances are based on the combination of tones of one and two syllable words (Wu 1982).

Besides the 16 tone combinations that can occur in two syllable words (4×4), the second syllable of a 2-syllable word can also be a neutral tone. The neutral tone occurs in a completely unstressed syllable, where the distinctive tone features disappear and the pitch is generally determined by the tone of the preceding syllable. The neutral tone is not marked in Pinyin romanization and is represented as a dot to the left of the reference line in Chao's five-point tone letter system. Although the actual pitch contours of neutral tone can be seen from a spectrogram (falling when following Tones 1, 2, and 4, and level or rising when following Tone 3), they are usually too short to be perceived clearly. The average pitch height of the neutral tone, according to a recent acoustic study (Gao, 1980), is shown below:

after Tone 1 ·|3

after Tone 2 ·|3

after Tone 3 ˙|4

after Tone 4 .|2

There are some minimal pairs of 2-syllable words (or phrases) which have exactly the same phonological structure and tones except for the stress on the second syllable: in one the second syllable is normally stressed and bears a basic tone and in the other the second syllable is unstressed and bears a neutral tone (e.g., *dōngxī* "East and West" and *dōngxi* "things". See Lin and Yan, 1980). However, except for these cases and a small number of suffixes, which are lexically marked neutral tone (toneless), if the second syllable of a 2-syllable word can be pronounced unstressed with a neutral tone, it can also be pronounced with normal stress with no difference in meaning.

When two syllables form a 2-syllable word (or close-knit phrase), some tone sandhi rules apply. The most important one is that Tone 3 is realized as Tone 2 when followed by another Tone 3. This rule makes the tone sequence 3-3 phonetically indistinguishable from 2-3 (Wang and Li, 1967), and therefore only 15 tonal patterns exist for 2-syllable words. Another rule of Tone 3 involves only a phonetic change: Tone 3 is realized as low falling or low level (without the rising part which occurs in citation

forms), sometimes called half Tone 3, when followed by a syllable bearing one of the other tones. This variant of Tone 3 in fact occurs more often (83%) than the citation form in actual speech (M-Y. Chen 1983).

Although there can be some other changes because of the shorter duration of the syllables in 2-syllable words and the anticipation of the following syllable in pronouncing the first syllable (e.g., when followed by another Tone 4, Tone 4 in the first syllable does not fall as much as in a citation form), the changes usually do not change the basic features of the tones. The pitch contours in 2-syllable words, except for cases where tone sandhi occurs, are basically a succession of the basic tones of the two syllables.

2.1.3 Tone in connected speech

It is generally agreed that "many of the functions of intonation in non-tone languages are fulfilled in Mandarin by other devices than intonation, such as the use of particles" (Chao 1933:130). According to Lyovin (1978:143), Rumjancev believes that in tone languages "the chief acoustic features of intonation are the amplitude and relative durations of various segments". Chao used pitch, duration (tempo), and loudness in describing the intonation patterns in Mandarin (Chao 1968:40-44).

However, among pitch, loudness, and duration, "pitch is the prosodic feature most centrally involved in intonation" (Cruttenden 1986:3). Since pitch is used to distinguish lexical meanings in a tone language, it can be inferred that the use of pitch to express other information would be restricted in order not to interfere with the distinctive function of lexical tones. Ross, et al. (1986) compared the affective functions of pitch movement in tone (Taiwanese, Mandarin, and Thai) and non-tone (English) languages and found that the local movement of pitch in tone languages is greatly restricted compared with English (p.283):

> The presence of tone in a language significantly inhibits the unrestricted manipulation of three acoustical measures of prosody which are involved in producing local pitch changes in the fundamental frequency contour during affective signalling.

However, the use of pitch in sentence intonation does exist in Mandarin. Chao wrote about the existence of expressive intonation in Chinese (1933:130):

> Owing to the fact that Western learners of Chinese tend to use intonation in a purely expressive role, writers on and teachers of Chinese have gone to the other extreme of regarding Chinese intonation as entirely fixed by word-tones. It is true that many of the functions of intonation in other languages are fulfilled in Chinese by other devices than intonation, such as the use of particles. But there still remains a good deal of pitch movement in Chinese speech which expresses moods and attitudes of the speaker in addition to the modulation of pitch due to the etymological word-tones and the more or less regular forms of neutral speech intonation when tones are joined together.

As to the question of how intonation and tone interact in terms of pitch movement, Chao compared syllabic tone and sentence intonation to small ripples riding on large waves. "The actual result is an algebraic sum of the two kinds of waves. Where two pluses concur, the result will be more plus, when a plus meets a minus, the algebraic addition will be an arithmetical subtraction" (Chao 1968:39). He also states that there are at least two types of intonation additions: simultaneous and successive (Chao 1933:131).

Simultaneous addition occurs when one of the following four features is added simultaneously to tones to form the resultant speech melodies.

(a) general raised level of pitch,

(b) general lowered level of pitch,

(c) widening of range,

(d) narrowing of range.

All of these may affect either the whole or a part of the intonation group. (The general pitch level has also been referred to as register by some authors; see Cruttenden 1986:129-130 for a discussion.) Successive addition is the intonational rising or falling endings added after the pitch contours of the final lexical tone. Chao later (1968) treated these two intonational endings of phrases and sentences as particles. The following table shows the modification of tones (adapted from Chao 1933:133 and Chao 1968:812, 814; numbers represent pitch height).

	rising endings	falling endings
Tone 1	55 = 556	55 = 552
Tone 2	35 = 356	35 = 352
Tone 3	214 = 2145	214 = 2141
Tone 4	51 = 513	51 = 5121

In contrast to "expressive intonation", there is neutral or normal intonation in "plain matter of fact statements", which is the regular succession of tones (with their regular tonal changes) in connected speech (Chao 1933:127). Chao observed that for normal intonation "in short sentences of three or four syllables, there is no special intonational modification. In longer sentences, there is a slight tendency for the pitch to trail off to a lower key toward the end" (Chao 1968:40).

After Chao's studies, there have also been some acoustic studies on Mandarin tones in connected speech. Rumjancev's study (1972, as reviewed by Lyovin 1978) found that "syllable duration is generally shorter in connected speech than in isolated citation forms. Also, the longer the duration of a given syllable in connected speech, the greater the likelihood that the rise or fall of F0 will be more pronounced" (Lyovin 1978:134).

Regarding the interaction of Mandarin tones and intonation,

> Rumjancev briefly enumerates various factors which affect various parameters of tone in

> connected speech (phrase intonation, logical emphasis, emotional nuances, syntagmatic and actual segmentation of sentences, and speech tempo) and then discusses some of them in detail.
>
> In discussing phrase intonation and its effects on tone he mentions what is already known from previous studies done by others (mainly his Soviet colleagues); e.g., he mentions what happens when the pitch features of intonation clash with the pitch features of lexical tones: for the most part, the basic contour features of lexical tones are left intact, but register features are significantly affected. In this connection, he also discusses what he calls 'supplementary post-tonal stretches of phonation'--a phenomenon which has been observed by some earlier investigators. Thus, for example, Tone 4 syllables falling within a zone of rising pitch intonation sometimes exhibit their usual falling pitch contour, plus a terminal rise in pitch. (Using Y.R. Chao's notation, we can say that in such cases Tone 4 is realized as 513, and that the 13 stretch is a 'supplementary post-tonal stretch of phonation'.) (Lyovin 1978:143-144)

What Rumjancev calls "supplementary post-tonal stretches of phonation" appears to be equivalent to Chao's "successive addition".

> According to Rumjancev's findings, these 'extra' stretches are entirely conditioned by intonational factors, but are not always present where they may be expected. Even in those cases where they are present (and their presence seems to be conditioned by emotional factors), the basic intonational contrast is signalled primarily by the pitch register of what precedes such stretches. In other words, such stretches are largely redundant, except perhaps in those cases where intonational contrasts involve emotional nuances. (Lyovin 1978:144)

Rumjancev also says that intonational factors have a very significant effect on various acoustic parameters of neutral tone syllables (Lyovin 1978:145). In some cases the

pitch level of neutral tone (e.g., sentence-final particles) is determined by intonation, not the context of the preceding lexical tone.

Rumjancev examined the acoustic features which signal the contrast between interrogative and declarative intonation:

> According to Rumjancev's findings, the primary differentiation of interrogative and declarative intonation is carried out by means of pitch features....The basic contour features of lexical tones are not significantly deformed by intonation; primarily it is the register features, the steepness of the rise and fall of F0, that signal the intonational contrast. (Lyovin 1978:148)

However, the most important intonational differences are in the last word (Lyovin 1978:149):

> To be sure, intonational differences are found as a rule in *all* parts of the sentence from its beginning to its end (though the acoustic parameters involved may differ from section to section). However, the critical intonational differences are localized only in the last word. Elsewhere, all intonational differences are secondary and redundant. This is the conclusion reached by Rumjancev on the basis of experiments which involved the 'transplanting' of the final word from a recorded interrogative sentence into the context of a recorded declarative sentence from which the final word had been erased and vice versa: the resulting sentences were judged by the native auditors to be either declarative or interrogative solely on the basis of the transplanted last word in spite of the fact that intonational cues in the rest of the sentence clashed with the cues in the last word. Moreover, such 'prepared' sentences were not felt to be in the least 'unnatural'.

Ho (1976a) investigated the acoustic correlates of the four Mandarin tones in isolation, and variations in these parameters in relation to six different sentence environments: (1) before a question marker *ma* in a sentence with rising intonation (X__Q#); (2) at the end of a sentence with falling intonation (X__#); (3) at the end of a sentence with falling intonation with a preceding emphatic modifier (emphX__#); (4) at the end of a clause in a sentence with falling intonation (X__,); (5) immediately following a sentence-initial demonstrative in a sentence with falling intonation (dem__X), and (6) in the initial position of a sentence with falling intonation (#__X). She found that sentence environment affects the realization of the tone, and the influence of sentence environment is greater than that of the vowel and the initial consonant of the syllable. However, the four tones maintain their basic shapes.

Ho (1976b) found that in declarative sentences, tone contours in sentence-initial position are higher on the frequency scale than those in sentence medial position (at the end of a clause), and tone contours in sentence-medial position are higher on the frequency scale than those in sentence-final position. She also compared the influence of falling and rising intonations on tone contours in sentence-final position (actually, in rising intonation it is the position before the question marker *ma*): all four tones start and end in higher frequency in environment X__Q# than

in environment X__# except that Tone 3 ends higher in the latter environment than in the former environment (possibly due to the occurrence of the following question marker *ma*). She also observed the influence of an emphatic modifier or a demonstrative on tone contours. Again, the four tones maintain their basic shapes.

Ho (1977) investigated the acoustic correlates of intonation for declarative, interrogative and exclamatory expressions in a frame sentence of five words: *Zhèi ge zì shì niàn* ___. ("This word is pronounced ___.") Results showed that interrogative and exclamatory intonation have a higher fundamental frequency throughout the whole sentence and also have a rising contour in the sentence-final position, while declarative intonation has a distinctive lower fundamental frequency and a falling contour in sentence-final position. Ho also studied and compared the influence of intonation and word position on some acoustical parameters (duration, amplitude, and fundamental frequency) of individual syllables (Ho 1977:454):

> To summarize the findings of this study, the duration of syllable nuclei in a sentence varies according to word position rather than intonation, except that in the sentence-final position. Both intonation and word position have influenced the fundamental frequency, tone range, and peak amplitude of syllable nuclei. As for the four tones in sentence-final position, the duration, fundamental frequency and peak amplitude are modified by intonation. However, the basic characteristic of the shape of the four tones remains very close to that of tones in isolated words.

According to Ho's statistics:

> ...word position has more influence on duration, fundamental frequency at the beginning and middle points, peak amplitude and tone range, but intonation has more influence on fundamental frequency at the end point. Furthermore, it has been found that intonation accounted for more variation on fundamental frequency at the end point in words 5 and 6. (Ho 1977:454)

The duration and fundamental frequency of sentence-final syllables by tone and intonation are shown in Figure 1 from Ho (1977:453).

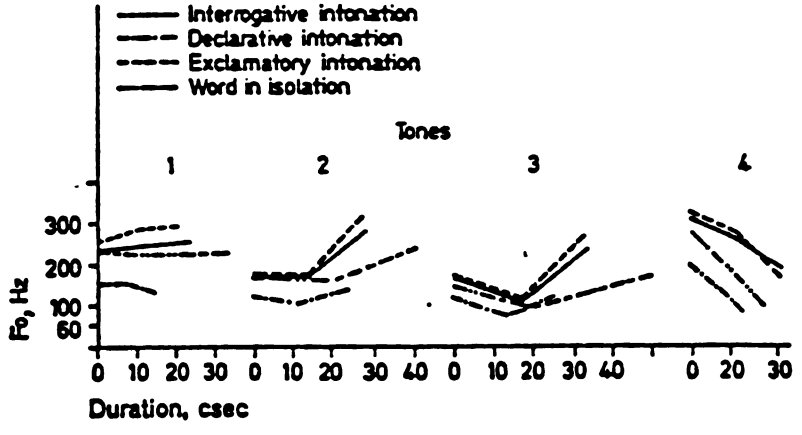


Figure 1. Tonal variations in sentence final position (Ho 1977:453)

Shen (1990) studied sentences and other utterances consisting of syllables of the same tone (e.g., all syllables in a sentence are Tone 1) spoken in different prompted contexts. She tried to find the basic intonation

patterns of Mandarin and the results from her instrumental analysis lead to the following conclusion (p.78):

> The intonation minimum...consists of three tunes: Tune I, Tune II, and Tune III. They function basically as statement (Tune I) and question (Tunes II and III). The clue to differentiating them involves their starting point: Tune II and Tune III start higher than Tune I regardless of their ending point, which may be in the high key (Tune II) or the low key (Tune III). Thus, the intonations are layered: an upper layer for interrogative and a lower layer for assertive.

The three tunes are shown in Figure 2 from Shen (1990:26).

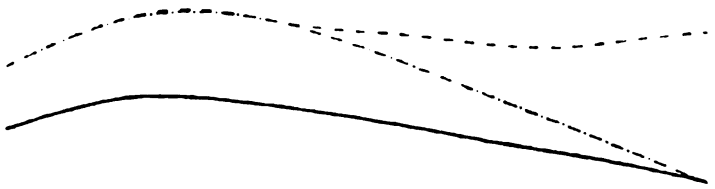


Figure 2. Basic tunes of Mandarin (Shen 1990:26)

Like Rumjancev and Ho, Shen found that tone contours are affected by intonation, especially for Tone 1, which is transformed from level into a slope. However, "[r]egardless of these modifications, the basic features of the four tones are left intact. The perception test carried out by Tseng

(1981) revealed that the shallow slope of Tone 1 singled out by careful instrumental analysis is still perceived by listeners as a level tone" (p.38).

Tseng (1981), on the other hand, has quite different findings analyzing data from spontaneous speech. She found that only 36.13% of the tones in her spontaneous speech data correspond to tonal contours predicted by phonological rules. "[T]he predictions were made on the basis of word and phrase levels only, i.e., on the rules that would predict word stress, modification of neutral tone suffixes, as well as neutralization of certain tones" (p.67). With a further break-down of the syllables into unreduced, reduced and neutral tone according to phonological prediction, Tseng reported the following percentages of correlation between actual and predicted tonal contours (p.73):

Table 1. Correlation between actual and predicted tonal contours

phonological prediction	percentage of correlation
syllable whose lexical tones should be present in production	40.17%
syllables that are neutral-tone suffixes	29.55%
syllables whose lexical tones are reduced to neutral tone	16.67%

The low correlation is in sharp contrast to the findings of Rumjancev, Ho and Shen, who found that tonal contours are not greatly affected in connected speech. This

suggests that read or monitored speech is quite different from spontaneous natural speech in respect to tone production. In spontaneous running speech, "the speakers' production is the consequence of a complicated interaction among lower level phonetics, phonology, and higher levels, syntax, semantics, and moreover their knowledge of the language as well as intention toward such an activity. ...This can be one interpretation of the low phonological-phonetic correlation of the production study" (Tseng 1981: 108).

Nevertheless, in accordance with Chao's observation of normal intonation in sentences of three or four syllables, Tseng found "...cases where the sentence is rather short in duration, about 700-900 msec. The lexical tones are clearly produced, leaving no trace of intonation....In this case, each lexical tone is clearly identifiable..." (p.162).

Wu (1982) reached the following conclusions concerning Mandarin intonation, based on his acoustic study of intonation in (prepared) statements, questions and emphatic sentences, in both normal and accelerated speed:

(1) Statements tend to have falling contour at sentence-final position.

(2) Questions without the question particle can have rising intonation at the end; questions with the question particle usually have the same ending as statements.

(3) In emphatic sentences, generally the pitch range is raised or the duration lengthened. Questions can also be emphasized and modified the same way.

(4) In accelerated statements, the pitch ranges of "non-critical" morphemes or words tend to be narrowed or even leveled. In very fast speech, these morphemes or words become the transition between the preceding and following tones and lose their original tone contours.

Besides doing acoustic analyses on the production of tone in connected speech, some researchers have also performed experiments on the perception of tone extracted from connected speech. Rumjancev (Lyovin 1978) has found that there exists a "duration threshold" for tone perception: when syllables shorter than 100-120 ms are taken out of the context of connected speech, the tone characteristics are leveled or neutralized and native listeners have difficulties identifying the tone correctly. According to Lyovin, Rumjancev found that

> ...in syllables whose tone contours are fully manifested, the said contours are sufficient for correct discrimination among the tones on the part of the auditors, and the latter do not need to pay much attention to register, which is a highly relative component of tone (since it is dependent on the individual speaker's voice). On the other hand, in syllables whose duration is close to minimal (as in allegro speech), contour features tend to be neutralized (leveled), and tones therefore are recognized mostly on the basis of register features, especially the register of their starting point. In such cases, difficulties arise since two...tones begin in the high register (Tone 1 and Tone 4) and the neutral tone may begin in any register (depending on the tone of the

> preceding syllable). This means that out of context these three tones tend to be confused by the auditors. (Lyovin 1978:136-137)

However, Rumjancev's results indicate a high percentage of correct identification of tones out of context (Lyovin 1978:140):

> Tone perception experiments involving syllables removed from their original contexts in connected speech indicate that even in the case of allegro speech the percentage of correct tone identification solely on the basis of the acoustic properties of tone is still relatively high (about 65%). In normal tempo this percentage, is, of course, significantly higher (80-85%).

On the other hand, Tseng's (1981) perception test showed that selected syllables taken out of the context of spontaneous running speech are correctly identified only 40.5% of the time by native speakers. For unreduced tokens, 55.66% were correctly identified. For the reduced tokens, 25.4% were correctly identified. If scoring is based on actual phonetic output, i.e., the tone of a syllable is determined by its actual fundamental frequency pattern rather than by the predicted tone according to lexical and phonological rules, the identification rate was not much different: 58.58% for unreduced tokens and 24.83% for reduced tokens. (The correlation of phonological and phonetic tone was 90% for unreduced tokens, 53% for reduced tokens). This identification rate is much lower than that found in Rumjancev's study. The reason for this low identification rate is not clear. The duration threshold

for tone perception found in Rumjancev's study may account for some of the difficulties since in Tseng's study the duration of each syllable varies greatly (108-618 ms for unreduced tokens p.92). The greatly modified pitch register in long running speech may be another reason for the low identification out of the context. The fact that the listeners were given six choices, including neutral tone and nothing recognizable in addition to the four lexical tones, may also have affected the results. However, data provided by Tseng were not sufficient for us to make a firm conclusion.

Leather (1983) studied speaker normalization in the perception of lexical tone. Two sets of synthetic Mandarin tone stimuli based on the syllable /y/ (lip-rounded high-front vowel) were created. They were then embedded in natural speech utterances of two native speakers with different but overlapping voice ranges, and presented to native listeners for lexical tone identification. All synthetic stimuli in the two sets had the same duration, formant frequencies, amplitudes and bandwidths, but the two sets differed in being adapted to the vowels of the two speakers. Two stimuli in each set had "good" Tone 1 and Tone 2 F0 contours, other stimuli had starting F0 or rate of rise between "good" Tone 1 and Tone 2, and four of them had F0 contours identical to both sets. While individual listeners differed somewhat in the acoustic criteria by

which they apparently decided lexical identity (Tone 1 /y/ "muddy" or Tone 2 /y/ "fish"), all listeners significantly assigned "ambiguous" stimuli with identical absolute F0 contours to different lexical categories depending on which of the speakers was heard to "produce" them. Leather concluded that the experiment provided support for the supposition that inferred F0 range of a speaker is a determining factor in the categorization of lexical tones. If the pitch range of a speaker can be significantly affected in spontaneous running speech, then listeners likewise need to "normalize" the pitch range in that context to perceive the tone correctly, especially when the contour features are not clear.

Connell, et al. (1983) conducted an experiment designed to determine how far the shape of a tone in Mandarin Chinese can be changed before it is consistently recognized as a different tone. The end point of pitch contour of naturally produced monosyllabic words was raised or lowered to simulate the effect of intonation on sentence-final syllables. These were then presented to a group of native Chinese speakers for tone identification. Results indicate changes in perception of tone. These changes correspond generally with their expectations based primarily on pitch contours. For example:

> [T]one 1 lowered by a slope of approximately 40 Hz or more is largely recognized as tone 4. On the other hand, when raised even by as much as 80 Hz, tone 1 was only very rarely recognized as tone

> 2...[which] lends some credance to Rumjancev's (1972) belief that the consonant to vowel transition contains sufficient information for tone recognition, since this portion of the stimuli was least affected by our pitch modifications. (Connell et al. 1983:349-350)

The recognition of tone appears to be stable over a wide range of pitch contour changes. The authors concluded (Connell et al. 1983:350):

> It is apparent that lexical tones...can withstand a large degree of perturbation before the risk of ambiguity arises. This range of stability is crucial, as it permits intonation to function in the same manner in Chinese as it does in non-tonal languages. Were the range of stability smaller, the use of pitch as the main cue for indicating such linguistic functions as interrogative and indicative, and paralinguistic functions such as attitude and emotion, would be severely restricted.

Gårding, et al. (1986) showed that the production of Tone 4 (*mài* "sell") and Tone 3 (*mǎi* "buy"), both falling in non-final position, are realized in different ways by different speakers in different speech acts, but some features are constant. Tone 3 generally has a low pitch level throughout the second half of the vowel and Tone 4 with a gradual fall over the main part of the vocalic segment. These observations were tested in a series of manipulations of pitch movements over *mai* from Tone 4 to Tone 3 in the sentence *Sòng Yán mai niúròu* ("Song Yan sells/buys beef"). The manipulated sentences were presented in a test, in which listeners were asked if they heard *mǎi* ("buy") or *mài* ("sell"). The result confirmed the observed

constant features. In addition, when the pitch contours lacked a clear reference (e.g., had conflicting cues or had no distinctive tonal cues at all) the identification by native speakers did not show a clear preference for Tone 3 or Tone 4.

Although the multiple use of pitch in Mandarin speech has clear effects on the production and perception of Mandarin tones, especially in spontaneous running speech, it can be concluded that in stressed positions tonal features generally are clearly produced in connected speech. Although the identification by native speakers of tone extracted from connected speech is far from perfect, the degree of correct identification is far above chance, even in fast reading and spontaneous speech.

2.2 Word stress and sentence intonation in English

In non-tone languages such as English, pitch is not an integral part of words or morphemes, yet all utterances still must be spoken with pitch. "Every sentence, every word, every syllable, is given some pitch when it is spoken. Even a sound in isolation is produced by vibrations whose frequencies constitute its pitch" (K. Pike 1972:53). Although English is probably the most studied of all languages, its use of pitch and other related prosodic features is still far from being well understood. There are different theories, models and hypotheses in the literature;

none is agreed upon by most researchers. The purpose of this section is not to give a full description of English prosody but to give a brief account of what has been found about English prosody which is relevant to this study, namely what is relevant to the use of pitch in words and sentences.

2.2.1 English word stress

English is often described as a stress-timed language. Stress plays an important role in English rhythm: "a general rule of English rhythm is that we take an equal amount of time from one stressed syllable to the next" (Cruttenden 1986:24). Word-stress is defined by Lehiste (1970:150) in relation to sentence stress and intonation. "Word-level stress is the capacity of a syllable within a word to receive sentence stress when the word is realized as part of the sentence."

Word-stress in English is not predictable in a simple way as in some other languages (e.g., French) which have word-stress regularly in a certain position on almost all words. Nevertheless, word-stress in a majority of English words can be predicted by a set of complex morphological and phonological rules (Chomsky and Halle 1968). For our purpose, these rules of English word-stress will not be described. What concerns us here are the acoustic

parameters used by native English speakers in the production and perception of word-stress.

As we will see below, in acoustic studies investigating the relative importance of pitch (fundamental frequency F0), loudness (intensity) and duration in the perception and production of English word-stress, pitch has been found to be the primary correlate while loudness and duration are only secondary in the perception and production of stress.

Lieberman (1960) studied the acoustic correlates of word-stress in American English. Twenty-five verb-noun pairs (e.g., CONflict-conFLICT) were recorded by 16 speakers of American English. The test words were presented in context in lists of sentences but subjects read only the test words aloud. Only utterances whose stressed syllables can be clearly perceived were analyzed. He found that the stressed syllable had a higher fundamental frequency than the unstressed syllable of the same utterance in 90% of the cases, a higher peak envelope amplitude in 87% cases, and a longer duration in 66% instances.

Fry (1958), in his experiments in the perception of stress, tested three physical dimensions--intensity, duration, and fundamental frequency--as cues for the perception of stress. Synthesized English noun-verb pairs (e.g., SUBject-subJECT) varying systematically in certain dimensions were judged by listeners. Duration ratio was found to be more important than intensity ratio in the

judgement of stress, and the fundamental frequency cue may outweigh the duration cue. Moreover, change in fundamental frequency differs from change of duration and intensity in that it tends to produce an all-or-none effect. "...so long as the resulting pitch change is easily perceptible to the listener, he tends to judge a higher syllable as more stressed, but the magnitude of the pitch change makes little contribution to his judgement" (p.144).

Morton and Jassem (1965) conducted another study on the perception of word stress:

> Morton and Jassem (1965) used synthetic nonsense syllables of the form /sisi/, /sɔsɔ/ and /sasa/ ...The stimuli were expressly constructed to have the same vowels in both syllables in order to avoid the problem of different intrinsic intensities. English listeners showed a high degree of consistency in their responses. Variation in fundamental frequency produced far greater effects than variations in either intensity or duration; a syllable was marked stressed if it was different from the "context" fundamental, and a raised fundamental was more efficient than a lowered one. In general, the more intense and longer syllables were more likely to be marked as stressed. (Lehiste 1970:128)

2.2.2 English intonation

In intonation languages such as English, intonation "involves the occurrence of recurring pitch patterns, each of which is used with a set of relatively consistent meanings, either on single words or on groups of words of varying length" (Cruttenden 1986:9).

One important difference between Mandarin tones and English intonation is that English intonation can occur with

utterances of varying length. As stated by K. Pike (1972:59):

> An intonation contour is not limited to specific syllables or words, but may be spread out over as many syllables and words as are colored by the speaker's attitude. For example, an intonation contour which begins low and rises slowly could be spread over three syllables, as in *(He's) doing it?* or the entire rising contour may occur on a single syllable such as *Tom*?

When pitch patterns are associated with more syllables than contour points (those points in the contour crucial to the establishment of its characteristic rises and falls), "[t]hese extra syllables can be pronounced with intermediate pitches..., or with considerable variation in the amount of drop from syllable to syllable" (K. Pike 1972:62). Thus the same contour and the same words can be associated in different ways. The following example shows how the same pitch contour can be realized differently on the same words (From K. Pike 1972:63):

telephone number!

2 3 + 3 4 + 4	(2 = high,
2 3 3 3 4	3 = mid,
2 3 4 4 4	4 = low.)
2 4 4 4 4	

Another difference between tone and intonation is in the way meaning is associated with them. Tone is an

integral part of a lexical item (morpheme or word) which has lexical meaning, while intonation meaning is quite different.

> English words have basic, intrinsic meanings; these lexical meanings are the ones found in the dictionary.... The intonation meaning is quite the opposite. Rather than being a stable inherent part of words, it is a temporary addition to their basic form and meaning....Rather than contributing to the intrinsic meaning of a word, it is merely a shade of meaning added to or superimposed upon that intrinsic lexical meaning, according to the attitude of the speaker. (K. Pike 1972:55)

One assumption behind the study of intonation is that pitch variation in intonation is used in a consistent way. As K. Pike stated (1972:53): "The changes of pitch which occur within a sentence are not haphazard variation. The pattern of variation, the rules of change, are highly organized." However, the perception of the pitch patterns in English intonation by English speakers is not distinctive at the lexical level as is the perception of tone by Mandarin speakers. As found in Lieberman's study (1965), even linguists familiar with the same (Trager-Smith) transcription system do not agree on the assignment of pitch levels in their transcription of recorded English sentences 60 percent of the time.

To describe the rules of intonation, the first step is to find the basic unit of intonation. However, there is much disagreement among researchers about this task.

> ...the lack of consensus on what constitutes the basic unit of intonation has led to a plethora of

> competing terms: breath group, phonemic clause, tone group, tone unit, tune, intonation group, intonation contour, intonation chunk, etc. Whether any or all of these refer to the same thing, however, is a matter for careful consideration. Depending on what defining criteria are set forth, quite different units may result. (Couper-Kuhlen 1986:76)

Nevertheless, despite the different defining criteria, the basic units of intonation in the majority of cases can be found because the native speakers tend to produce them consistently. For example, Crystal (1972) found that "when native speakers were presented with the task of repeating an utterance, there was maximum agreement (84.8 per cent) over the location of tone-unit boundaries" (Crystal 1972:110-111).

Once utterances are divided into basic intonation units, the internal structure of the basic intonation unit is most commonly viewed as consisting of one obligatory constituent, the nucleus, which is the most prominent syllable in an intonation unit, and a number of optional constituents. There are different views about how the optional constituents should be divided or described. One way is to describe the head (the pattern beginning on the first stressed syllable), the pre-head (syllables before the head) and the tail (syllables following the nuclear syllable) (Crystal 1972). Thus the structure of the basic intonation unit can be represented as follows:

(pre-head) (head) nucleus (tail)

The following example has all the possible constituents in a intonation unit (White 1980:11 adapted from Lindström 1978):

```
┌P┐ ┌--H---┐ ┌N┐ ┌----------T----------┐
She won't be 'at that school much longer.
     (' = mid-fall)
```

In describing the pitch patterns of English intonation, one way is to describe 'whole tunes', the contours produced by joining all the accents together. However, this approach is rejected by some researchers "because it misses important generalisations dependent on the occurrence of similar tones starting from the nucleus" (Cruttenden 1986:56).

Since it is generally agreed that the pitch pattern starting from the nucleus (often termed nuclear tone) carries the most important intonation meaning (Cruttenden 1986, Crystal 1972, Lindström 1978), most intonation analyses of English have been concerned with describing it. The optional pre-nuclear constituents (head and prehead) "only modify the meaning associated with the nuclear tone" (Cruttenden 1986:63) and are less studied and less understood.

Although most researchers agree that nuclear pitch patterns carry major differences of intonation meaning, they disagree in their description of nuclear pitch patterns.

> This is an area where almost every analyst varies in his judgement of what constitutes a 'major difference of meaning' and hence in the number of nuclear tones which are set up. Nor are the arguments for any one particular set of nuclear

> tones ever very convincing or indeed explicit; in fact, given that intonational meanings are often intangible and nebulous, it is difficult to see how a wholly convincing case for any one set of nuclear tones can be made. (Cruttenden 1986:58)

For example, according to Cruttenden (1986:58f), seven nuclear pitch patterns can be set up for English: high-fall, low-fall, rise-fall, low-rise, high-rise, fall-rise, and mid-level. All the seven nuclear pitch patterns are shown to carry some different meanings. For instance, "The low-fall is generally more uninterested, unexcited, and dispassionate whereas the high-fall is more interested, more excited, more involved..." (Cruttenden 1986:100). However, these nuclear pitch patterns can be grouped into falls and rises and be ascribed abstract meanings: "high-fall, low-fall, rise-fall can be grouped semantically as falls, which involve a sense of finality, completeness, definiteness and separateness, particularly when used with declaratives" (Cruttenden 1986:100). On the other hand, low-rise, high-rise, fall-rise, mid-level can be grouped as rises, and "...the use of a rising tone signals dependency or non-finality" (Cruttenden 1986:102).

A major reason that the meanings of intonation are so difficult to grasp is that they vary in different contexts. For example, Cutler (1977) provided some convincing evidence that the so-called contradiction contour (an initial rise followed by a fall followed by a final rise of some sort) proposed by Liberman and Sag (1974) can be used to express

quite different meanings such as enlightenment or impatience, and that there are alternative contours for signaling contradiction (Scott 1987:99). The following are some potential context effects identified by Cruttenden (1986:114):

> non-final v. final; declarative v. yes/no interrogative v. wh interrogative v. imperative v. exclamative; tense; agreement v. disagreement; relationship to preceding tone; and general context, including preceding utterances and physical situation. Lurking behind many of these particular factors is the more general one of speaker/listener relationship.

From the literature cited above, we can see that disagreement among analysts exists on virtually every aspect of English intonation. This is very different from the analysis of Mandarin tones, about which almost everyone agrees.

2.3 Comparison of the two systems

It has been suggested that intonation is a universal phenomenon and that its correlates may be related more to innate biological factors than to conventional grammatical factors. Bolinger (1978:471) surveyed languages of the world and found intonation universals:

> The traits of intonation shared by the majority of languages, not excepting tone languages, are both formal and semantic, and cover the two main non-tonal uses of pitch variation: to form closures (descending lines, clause-final falls and non-falls) and to form accents (obtrusions for prominence, mainly upward). Terminals are almost universally low or falling for finality and

> assertion, and high or rising for the opposite, including yes-no questions. Accents are generally set off by contrasting pitch levels, and their position in the sentence indicates both focus and climax.

In section 2.1 above, it is shown that Mandarin also exhibits these intonational universals. However, the Mandarin lexical tones are usually modified without changing their tonal characteristics. For example, a level tone may be transformed into a shallow slope, and the slope of a rising or falling tone may become steeper or shallower.

The most important difference between Mandarin and English is that pitch is distinctive primarily at the syllable (morpheme) level in Mandarin while in English the pitch functions on the word level or above. This difference (lack of lexical tone in English) is related to other differences in the use of pitch in the two languages.

For example, the pitch patterns associated with English word-stress are more flexible for native English speakers and are more susceptible to the influence of intonation (Chao 1956). On the other hand, the tonal patterns in Mandarin words are fixed and generally cannot be robbed of their basic characteristics by intonation. Thus, as Chao (1956:42) noted, in the speech of Chinese learners of English the pitch patterns associated with word-stress in isolated English words tend to be maintained in phrases and sentences.

Affective or attitudinal functions which in English are signaled by local pitch change are not signaled in the same way in Mandarin. A study by Ross, et al. (1986) reported on the influence of tone on the ability to signal affect through prosody in three tone languages (Taiwanese, Mandarin and Thai) as compared to English.

> The results...clearly indicate that the presence of tone in a language adversely impacts the free use of F0 for affective-prosodic signalling. Of the five acoustical parameters characterizing the F0 contour, three...show statistically that English singularly surpassed all the tone languages in the study by at least a factor of 2 ...Those acoustical parameters in which English singularly surpassed the others, directly or indirectly, indicate that English has greater freedom than tone languages in *locally* varying the intonational contour for affective signalling.
>
> The results involving timing and intensity measurements showed no significant influence of language type (tone vs. non-tone) on these parameters as opposed to certain F0 parameters. Thus, intensity and timing parameters may be used freely in a tone language for affective-prosodic purposes. One might even be tempted to assume that intensity and timing parameters could be used in tone languages to compensate for tone imposed constraints on affective intonation. However, our data...does not indicate that tone languages use such compensation, since English statistically performed as well as the best of the tone languages in all intensity and timing parameters. (Ross, et al. 1986:298-301)

Another difference between Mandarin and English is that the pitch range used in the two languages is different. G-T. Chen (1974) investigated the pitch range of English and Chinese speakers. His study included two types of speech

utterances: words and sentences. He reached the following conclusions (p.170):

> 1) When the test subjects spoke their native languages, the average pitch range of the four Chinese subjects was 258% wider than that of the four English-speaking subjects established by UTTERANCE WORD and the average pitch range of the four Chinese subjects was 154% wider than that of the four English-speaking subjects established by UTTERANCE SENTENCE.
> 2) When the four English-speaking subjects switched from their native language to Chinese, their pitch range increased by 123% when UTTERANCE WORDS were used and by 62% when UTTERANCE SENTENCES were used.

G-T. Chen (1974:169-170) also found that the pitch ranges of Chinese speakers increased only 4.4% from utterance words to utterance sentences. On the other hand, the pitch ranges of English speakers increased 47% from utterance words to utterance sentences when they speak English.

A contrastive analysis of Mandarin tone and English intonation by White (1980, 1981) compared the pitch patterns occurring in Mandarin tones and English intonation:

> With what is known about the relative pitch levels and contours of Mandarin tone, and through extrapolation from Chen's (1974) data, we can compare the pitch levels and contours of Mandarin and English tones....Mandarin has a high level, a rising, a low falling-rising, and a falling tone; English has at least level, falling, rising, falling-rising, and rising-falling tones. (White 1980:23)

Only the English pitch patterns which are similar to the Mandarin tones described in section 2.1, will be described below:

> The high level tone occupies a high or medium voice pitch and is not normally pitched lower than any preceding unstressed syllables. High level tones are used to give secondary prominence to words in the head of the tone-group, where they usually form a gradually descending sequence. (Lindström 1978:78)
>
> The rise is frequent as a nuclear tone. It is usually independent but may, in a compound tonic, be subordinate to a preceding nuclear tone. In prenuclear position it may serve to build up a rising head.
> The rise begins on a low or medium pitch and reaches a medium or high pitch. Its range varies a good deal. Wide rises normally start from a low pitch. A "broken" variant which occurs only as a nuclear tone, may begin on a higher pitch and is realised by a falling-rising pitch movement. (pp.79-80)
>
> The fall-rise is always independent and nuclear. It falls from a high or medium pitch to a lower pitch and then rises to a medium or high pitch. In non-final position in the utterance it tends to be high and narrow. In final position its range varies. The rising part of the tone does not normally end higher than the onset of the preceding falling element. (p.80)
>
> The fall may begin in any voice pitch region and ends up on a lower pitch. When it forms the nuclear tone of the tone-group, its range varies with the pitch of the nucleus. In the head it is usually high and narrow. Postnuclear falls are always low and narrow. (p.79)

White showed that many tonal errors produced by learners of Mandarin can be attributed to the transfer of English intonation in certain sentence types or in particular sentences. For example, the intonation on simple declarative sentences typically begins on a low, unstressed syllable, builds to the nuclear (or stressed) syllable, and then glides down. This pattern may account for the low-tone

initial syllable and gliding final syllable heard when English speakers produce Mandarin sentences. Mandarin V-not-V questions are frequently pronounced with a rising intonation by English speakers in approximation of the intonation in the English equivalent (yes/no questions). Many Mandarin descriptive adjectives or stative verbs which are not pronounced with a falling tone (e.g., *hěn* "very") are often mispronounced by English speakers because emphatic stress is associated with a high-falling tone in English, and such words are often stressed in English.

Because the function of pitch in a tone language like Mandarin and an intonation language like English is different, pitch ranges used in the two languages are different, and some specific similar pitch patterns are used in different ways for different functions in the two languages, it is not surprising that English speakers experience difficulties in their learning of Mandarin tones.

3. PERCEPTION AND PRODUCTION OF SOME ASPECTS OF L2 PHONOLOGY

In this chapter, I will first review literature on the perception and production of some aspects of L2 phonology, with emphasis on the adult learners; then I will review the literature on the perception and production of Mandarin tones by L2 learners.

3.1 Perception and production of L2 phonology by adults

Most studies on L2 phonology have primarily been concerned with production rather than perception. The major reason for this is probably that production is easier to observe. In fact, even a slight accent produced by L2 learners is easily detected by native speakers (Scovel 1988:105ff). However, since the learning of L2 phonology should include both production and perception, literature on both aspects will be reviewed.

Studies that examine the variables affecting L2 phonological learning (pronunciation) have always found age at arrival in the L2 environment the most important variable (Asher and Garcia 1969, Oyama 1976, Seliger, et al. 1975, Tahta, et al. 1981a, Piper and Cansin 1988). The younger an L2 learner starts learning in the L2 environment, the more likely he/she will achieve native-like pronunciation. The critical period hypothesis, which "is the notion that language is best learned during the early years of

childhood, and that after about the first dozen years of life, everyone faces certain constrains in the ability to pick up a new language" (Scovel 1988:2), has been controversial in many areas of L2 learning, but it is strongly supported in terms of learning native-like pronunciation (Scovel 1988).

However, despite the overwhelming evidence of relationship between age-at-arrival and L2 pronunciation, research has found that adults are better than children in learning L2 pronunciation in the short term (Olson and Samuels 1973). It is also unclear whether the younger-better effect exists at the very beginning of L2 learning. To determine whether the younger-better effect of L2 pronunciation attainment exists because younger learners are better at imitating new sounds, some studies investigated the abilities of subjects of various ages to imitate foreign sounds they were not familiar with. Some studies found an older-better effect (Loewenthal and Bull 1984; Snow and Hoefnagel-Höhle 1977, laboratory study), while others found a younger-better effect (Yamada et al. 1980; Tahta et al. 1981b). Snow and Hoefnagel-Höhle (1977, naturalistic study) reported a positive (older-better) correlation between age and L2 pronunciation among L2 learners at the beginning of their L2 learning in the L2 environment; however, this older-better effect changed over time. Thus, "despite the general initial advantage of the older subjects, both for

total scores and for many individual sounds, the younger subjects had become better at pronouncing some of the sounds after a period of 10-11 months' learning" (p.362).

Whether there is a younger-better relation in imitating foreign sounds or not, and whatever the age-related causes involved in the acquisition of L2 phonology, it is clear that there are differences between children and adults in L2 phonological learning, and it is virtually impossible for adult L2 learners to achieve native-like accent. However, the pronunciations among adult learners still vary greatly. The adult learners' learning of L2 phonology, including both perception and production, will be our main concern below.

3.1.1 Variables affecting L2 phonology learning

On the production aspect of L2 phonological learning, Purcell and Suter (1980) found that the most effective predicting variable of L2 pronunciation for adult learners from four different L1 countries is the first language (L1), followed by aptitude for oral mimicry, residency in L2 environment, and concern for pronunciation accuracy.

> The stepwise results indicate that only four of the predictors are useful in accounting for the variance of subjects' pronunciation accuracy scores. First Language was by far the most important predictor accounting for more variance than any other variable. Aptitude for Oral Mimicry was the second most important predictor. The third most important predictor was a composite variable, CV1, the only composite variable to contribute significantly to the regression. CV1 combined Years in an English Speaking Country and Months Residing with a Native Speaker of English.

> We interpret CV1 as reflecting length of residency in an English speaking environment. The last predictor to contribute significantly to the explanation of the criterion's variance was Strength of Concern for Pronunciation Accuracy. The remainder of the predictors were deemed useless, after the above four were taken into account. (Purcell and Suter 1980: 282-283)

Thompson (1984) investigated Russian immigrants who came to the United States at various ages and found that the best four-variable predictor model of pronunciation accuracy includes (1) age at arrival (0.66), (2) sex (0.77), (3) aptitude for oral mimicry (0.82), and (4) speaking proficiency in English (0.85) (p.94).

On the perception aspect of L2 phonological learning, Thompson (1984) found that the best four-variable predictor model of L2 sound discrimination includes (1) ear preference (0.20), (2) percent of English speaking friends (0.26), (3) importance of English for the job (0.31), and (4) American orientation (0.34). She noted that the battery of predictors was far more effective in accounting for differences in production than it was in explaining the differences in perception.

In the following sections, however, the variables that will be discussed are (1) phonological context, (2) task, (3) sex, and (4) L2 experience. These variables may be more important when L2 learners are adults of relatively narrow range of age and are from the same L1 background. These are

the variables that will be examined in my analyses of American learners' learning of Mandarin tones.

3.1.1.1 Phonological context

(A) Production

Phonological context has been found to have an effect on the pronunciation of L2 sounds (Flege 1988:315):

> Several L2 studies have demonstrated that authenticity of pronunciation varies as a function of phonetic context. Gatbonton (1975, discussed in Segalowitz and Gatbonton, 1977) observed that the difficult /ð/ sound of English was produced authentically by French Canadians more often when it precedes a vowel or fricative than a voiceless stop. Dickerson (1974, 1975) found that the perceived authenticity with which Japanese learners produced English sounds varied as a function of phonetic context. In both studies, the effect of phonetic context on authenticity persisted over time even though pronunciation improved overall.

L. Dickerson (1975) found that L2 learners produced L2 sounds better in word-list reading than in dialogue reading. Borden et al. (1983) investigated Korean learners' production (repeating after native speakers) of /r/ and /l/ in nonsense syllables, words, and sentences and found that "[n]onsense syllables were judged correct a larger percentage of the time than were words, and the [r] and [l] sounds in sentences were judged to be least correct" (p.509).

(B) Perception

The effect of phonological context on the perception of L2 sounds has been found in studies investigating the perception of English /r/ and /l/ by Japanese learners:

> In examining identification by Japanese learners, Shimizu and Dansuji (1983) found that English /r/ and /l/ were identified better in the context of front than back vowels (about 90% versus 80% correct), and better in isolated words than words embedded in a carrier phrase. The authors suggested this was due to a "more complete" articulatory specification of sounds in isolated words, but it is also possible that the demands of processing a longer stretch of speech makes it more difficult for listeners to identify new L2 sounds. (Flege 1988:278)

> Mochizuki (1981) also presented native English-produced minimal pairs with /r/ and /l/ to native Japanese-speaking listeners. Correct identification averaged only 82% overall, compared to the near-perfect identification by English-speaking listeners. However, percent correct identification was higher for sounds in word-final (98%) than word-initial position (88%) which, in turn, was higher than intervocalic (73%) and word-initial clusters (69%). Sheldon and Strange (1982) obtained very similar results. (Flege 1988:330)

Clearly, both production and perception of L2 phonology are likely to be affected by phonological context. The findings of L. Dickerson (1975) and Borden et al. (1983) that L2 pronunciation in isolated words was better than that in longer utterances, and Shimizu and Dansuji's (1983) finding that /r/ and /l/ were identified better in isolated words than words embedded in a carrier phrase, are of special interest since the phonological contexts examined in this study will be isolated syllables, 2-syllable words, and

short sentences. Findings in these studies suggest that sound perception and production in longer utterances are more difficult for L2 learners than in isolated words.

However, other factors may sometimes change this. Beebe (1987) found that Thai speakers produced more errors of English initial /r/ in word-list reading than in conversation. A major reason for the low accuracy rates of /r/ in word-list reading is interference from Thai. Thai has a sociolinguistic rule of using trilled /r/ in formal settings. Since word-list reading is more formal than conversation, there is more negative transfer in the word-list reading data than in the conversation. In contrast to initial /r/, final /r/ was pronounced correctly by the Thais significantly more often in word-list reading than in the conversation.

3.1.1.2 Task

(A) Production

Speaking task has also been found to influence the pronunciation of L2 sounds (Flege 1988:259):

> One other study showed that L2 pronunciation varied according to speaking task. L. Dickerson (1974, 1975) examined the ability of Japanese L2 learners to approximate the phonetic norms of English for a number of sounds. Pronunciation was more often--and more nearly--correct in word-list reading than in dialogue reading, and in dialogue reading than in conversational speech. Although the subjects' pronunciation of the sounds examined improved over the course of one year, the effect of speaking task remained. W. Dickerson (1976, 1977a, 1977b; L. Dickerson & W. Dickerson, 1977)

> hypothesized that the variability observed resulted from the same kind of tacit sociolinguistic knowledge native speakers use in L1 to vary systematically pronunciation. He asserted that the L2 learner's developing competence is best described by variable phonological rules specifying the probablistic effect of speech styles and task, as well as phonetic context, on the selection of sound variants.

Because the two production tasks performed by subjects in this study are imitation and reading aloud, the task effect of imitation and reading in L2 pronunciation is examined. However, very little has been done in this area.

E. V. Pike (1959) found that the ability to imitate foreign sounds correlates with students' final grade (based on their ability to take dictation, to pronounce sound sequences, and to understand phonetic theory) in a forty-five hour course in phonetics. The ability to imitate foreign sounds in general has also been found to be a significant predictor of specific L2 pronunciation accuracy (Purcell and Suter 1980, Thompson 1984). These findings indicate that the ability to imitate foreign sounds probably represents some phonological learning aptitude that cannot be modified easily. Nevertheless, there is evidence that with practice some improvement in imitation is immediately possible. Snow and Hoefnagel-Höhle (1977, laboratory study) found that the imitation of Dutch words by English-speaking subjects improved between the first ten trials (block 1) and the second ten trials (block 2). However, not all sounds

showed improvement. Some sounds showed significant improvement (/k/, initial /x/, /ø/, /l/ and /ɛi/), while others (final /ʃ/, /x/ in a cluster, /Ʌy/, and initial /r/) showed no improvement.

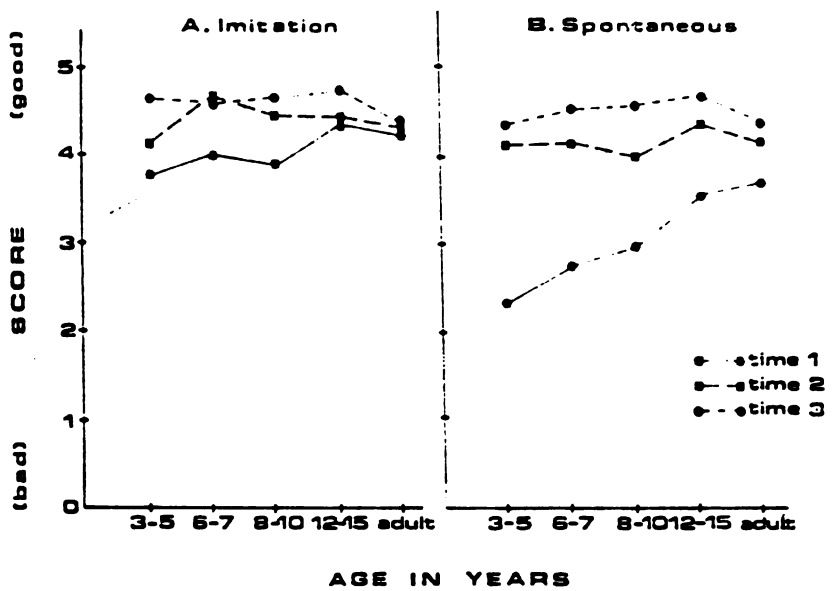


Figure 3. Median scores of two pronunciation tasks (Snow and Hoefnagel-Höhel 1977:363)

Snow and Hoefnagel-Höhle (1977, naturalistic study) used both imitation and spontaneous speech (in response to a picture) in analyzing the pronunciation of L2 learners. They did not compare the results of the two tasks. Their data are shown in two graphs in Figure 3 (Snow and Hoefnagel-Höhle 1977:363), and some observations can be made regarding the two tasks. The spontaneous scores show significant

positive correlation with age (older-better) at time 1. The imitation scores were better than the spontaneous scores for all age groups at time 1, but the differences were greater for younger groups than for adults. At time 2 the differences between the two tasks became very small as a result of greater improvement in spontaneous production than in imitation. At time 3 the scores were almost the same for the two tasks. Overall, the improvement in imitation was not as great as in spontaneous speech for all age groups. In fact, for adult learners, the imitation scores were almost unchanged from time 1 to time 3.

3.1.1.3 Sex

(A) Production

Sex is a variable which has frequently been examined in studies of L2 pronunciation. Sex was found to be a significant variable in L2 pronunciation in some studies (Asher and Garcia 1969, Thompson 1984) in which females were found to have better pronunciation than males. However, the effect seemed to diminish with length of residence (Asher and Garcia 1969). Purcell and Suter (1980) found sex not a significant variable in L2 pronunciation. Tahta, et al. (1981a) found sex contributed significantly to variance in accent, though its contribution is relatively small (2.2%).

> We might note the observation that women were less likely to show accent transfer than men. This difference did not emerge in the second multiple regression analysis: the effect (in the first

> regression analysis) might have been due to the subjects in the sample who had begun to learn English after the age of 13. However, a separate χ^2 test of this difference did not yield a significant result ($\chi^2 = 2.45$ $p > 0.05$). (Tahta et al. 1981a:271)

Studies that investigated the abilities to imitate foreign sounds has found no sex differences, either. Snow and Hoefnagel-Höhle (1977, laboratory study) found no sex differences in the abilities to imitate new sounds in subjects aged five to 31. Tahta et al. (1981b), who looked at the abilities of 5-15 year old monolingual English schoolchildren to replicate foreign pronunciation and intonation, also found no sex differences. They concluded (Tahta et al. 1981b:370):

> We found no reliable sex differences for either languages at any of the ages studied. Asher and Garcia (1969) found female superiority among fluent L2 speakers of one or more years residence, but sex differences disappeared with length of stay. Other studies (e.g. Oyama, 1976; Tahta et al., 1980, 1981) restricted observation to L2 speakers with greater length of stay and failed to report reliable sex differences. We must conclude that if there are sex differences among L2 speakers, these disappear with practice and exposure, and our findings suggest they are not attributable to reliable intrinsic differences between the sexes in abilities to replicate novel foreign language sounds.

(B) Perception

Few studies investigated the effect of sex on L2 phonological perception. Thompson (1984) found sex not significant in the perception of L2 sounds by L2 learners.

3.1.1.4 L2 experience

(A) Production

It is reasonable to assume that previous L2 learning experience will have some influence on subsequent language learning. Some people believe that learning a second language makes the learning of subsequent languages much easier (see Larsen-Freeman and Long 1991:206). Although the statement is usually made regarding the learning of a language as a whole rather than the learning of L2 phonology specifically, if prior L2 learning experience helps, it could also help in phonology learning. A few studies on L2 pronunciation learning considered prior other L2 learning experience. Suter (1976) found "number of languages the speaker is able to converse in" not a significant predictor of pronunciation accuracy. Thompson (1984) investigated variables predicting L2 pronunciation and also found knowledge of other languages insignificant.

(B) Perception

Few studies investigated the effect of prior L2 experience on the perception of L2 phonology. Thompson (1984) found knowledge of other L2s insignificant in the perception of L2 sounds.

3.1.2 Relationship between the perception and production of L2 sounds in isolation and in sequence

Sounds and suprasegmentals are produced and perceived as part of words and utterances in speech. Although one word utterances do occur in natural speech, words are usually spoken in sequence rather than in isolation. Learning a language means learning to produce and perceive natural, fluent utterances.

Since utterances are sequences of words and adult L2 learners have this knowledge from their L1, it is assumed that adult L2 learners normally learn individual words first, before they can put words together in sentences or be able to recognize and understand them in utterances. (Unanalyzed chunks or routines can be viewed as long words.) However, as we will see below, the relationship between the perception of sounds in isolated words and connected speech, and the relationship between the production of sounds in isolated words and connected speech may not be a direct and simple one because they involve some different factors. Similar relationship may exist in both perception and production between tone in isolated syllables and in sequences which will be analyzed in this study.

3.1.2.1 Perception

Acoustic studies have found that the acoustic signals of connected speech are continuous, with no clear boundaries

between words, and that sounds produced in different phonetic contexts exhibit great variation. The lack of acoustic-phonetic invariance is one of the most important problems in speech perception:

> For more than 30 years, it has been extremely difficult to identify acoustic segments and features that uniquely match the perceived phonemes independently of the surrounding context. As a result of coarticulation in speech production, there is typically a great deal of contextual variability in the acoustic signal correlated with any single phoneme. Often a single acoustic segment contains information about several neighboring linguistic segments (i.e., parallel transmission), and, conversely, the same linguistic segment is often represented acoustically in quite different ways depending on the surrounding phonetic context, the rate of speaking and the talker (i.e., context-conditioned variation). In addition, the acoustic characteristics of individual speech sounds and words exhibit even greater variabilities in fluent speech because of the influence of the surrounding context than when speech sounds are produced in isolation. (Pisoni and Luce 1986:3-4)

The acoustic-phonetic invariance is a problem for the theory of speech perception because listeners perceive the sound variants as the same. In addition to the problems arising from the lack of acoustic-phonetic invariance, there are also related problems having to do with the normalization of the speech signal. That is, despite great variations in acoustic signals produced by different speakers or different rates of speaking, listeners are able to normalize the differences and perceive sounds produced by different speakers or at different rates to be the same. The

normalization problem of speaking rate is relevant to the perception of isolated words and fluent speech:

> A second normalization problem concerns time and rate normalization in speech perception. It is well known that the duration of individual segments is influenced substantially by speaking rate. However, the acoustic duration of segments is also affected by the location of various syntactic boundaries in connected speech, by syllabic stress, and by the component features of adjacent segments in words....In addition, there are substantial differences in the duration of segments of words when produced in sentence contexts compared to the same words spoken in isolation. The duration of vowels produced in sentences is roughly half the duration of the same vowels spoken in isolation. Speaking rate also influences the duration and acoustic correlates of various phonetic features and segments. Numerous low-level phonetic and phonological effects such as vowel reduction, deletion, and various types of assimilation phenomena have been well documented in the literature. These effects seem to be influenced a great deal by speaking tempo, dialect, and surrounding phonetic context. (Pisoni and Luce 1986:11)

Perception of connected speech differs from perception of isolated words because it involves segmentation or locating word boundaries. Segmentation and recognition of words in connected speech is a complicated process that involves the interaction of sound perception and much other knowledge. Some evidence for the contribution of other knowledge was found in research on the perception of speech in noise (e.g. Miller et al. 1953), which found that words presented in noise are recognized more accurately in a sentence context than in isolation (even though words spoken

in isolation are acoustically better defined than words in fluent speech).

However, acoustic information, especially when it is clearly defined, is still an important basis of speech perception. One piece of evidence for this view is that mispronunciations can be detected by listeners. Cole's (1981) study on children's detection of mispronunciation showed that sounds in fluent speech and isolated words are not perceived equally well.

> The results showed that children detect mispronunciations in isolated words much more accurately than in fluent speech. Across all groups, mispronunciations of words spoken in isolation were detected 95% of the time, compared to 59% detection of mispronunciations in nursery rhymes. (Cole 1981:101)

Thus the sounds children have learned to perceive in isolated words have not been learned to the same extent in fluent speech.

Cole (1981) has also found that error detection performance of children was affected by the type of phonetic substitutions. The most detectable mispronunciation involved the substitution of one stop for another or the substitution of a nasal for a stop. By contrast, substitution among nasals and fricatives were detected 20% or 30% less often than were substitutions among stops. On the other hand, adults (college students) detected all word-initial mispronunciations with at least 90% accuracy. Thus, all of the phonetic substitutions are highly detectable by

adult standards. Some researchers have found similar processes in adult L2 acquisition and L1 acquisition by children (e.g. Krashen 1982). If the difference in detecting mispronunciation between adult and children indicates that children are still in the process of learning the language, we may expect adult L2 learners to behave more like children in their learning of L2 phonology.

Goldstein (1983) studied word recognition in foreign speech on the basis of acoustic signals alone. English-speaking subjects not familiar with French were trained to remember the pronunciation of four French words. They were then asked to detect the four French words in a prose passage read in French by a native speaker. Nouns were deleted from the original French passage about every four to five sentences. Each test word was in turn substituted into the passage. Mean detection rates for the four words ranged from 64% to 94%. There were significantly more false alarms in detecting the 1-syllable word *taux* than *seau*; a result may be due to a greater probability of words or parts of words that sounded like *taux*. Thus in contrast to native listeners, foreign language listeners appear to be greatly affected by the degree to which target words resemble other words and syllables phonetically. This is more likely to happen in listening to connected speech since there are more "other words" and the available processing time is likely to be shorter. A fluent French speaker listening to the same

French passage made errorless and much more rapid detection with only a few false alarms.

3.1.2.2 Production

It was mentioned that acoustic signals produced in connected speech and in isolated words have been found to be different in many respects. For instance, the average duration of segments is shorter in sentences than in isolated words, and the effects of vowel reduction and assimilation are greater (Pisoni and Luce 1986:11, see also Lyovin 1978 for similar average duration of segments in Chines).

The difference between production in connected speech and in isolated words results mainly from mechanical constraints on the movement of articulatory organs during speech production. In speech production, the shape of the vocal tract for any given sound often accommodates to the shape needed for surrounding sounds, a phenomenon called coarticulation. The result of coarticulation is referred to as the undershooting of targets. In producing isolated words, since a pause precedes and follows the words, no coarticulation occurs at word boundaries. On the other hand, connected speech is produced at a faster rate, which is also achieved by undershooting articulatory targets, and coarticulation at word boundaries is normal.

One of the differences found in the production of isolated words and connected speech is VOT (Voice Onset Time) in stop consonants. Languages have been shown to use VOT in perception and production of voiced and voiceless stops (Lisker and Abramson 1964; Williams 1979, 1980). VOT refers to the relative timing of voicing onset with reference to the release of articulatory closure. By convention, negative VOT values stand for voicing lead (onset of glottal vibration prior to articulatory release), and positive values, voicing lag (onset of glottal vibration following release) (Williams 1980:187). Locke (1983:101) cites Baran, Laufer, and Daniloff (1977) who:

> ...measured the VOTs of voiced and voiceless stops in three adults while speaking conversationally, reading a passage, and reading single-word citation forms. They found VOT values to be most discrete for citation forms, where there was no overlap between voicing categories (essentially confirming Lisker & Abramson, 1967). In speaking and reading, there was less difference between voiced and voiceless stops; in conversational speaking there was a "noted trend for increased rate of utterance and less precise articulation [p.348]."

Many of the differences found in isolated words and fluent speech, such as segment duration and VOT, are probably universal. We would expect to find similar variations in isolated words and connected speech in the production of L2 learners.

L2 learners have been found to produce L2 sounds better in word-list reading than in dialogue reading and

conversational speech, and the discrepancy remained over the course of one year, despite overall improvement in the sounds examined (L. Dickerson 1975). This indicates a systematic relation between L2 sounds produced in isolated words and in connected speech. It seems reasonable to assume that adult L2 learners when they learn to produce L2 sounds in isolated words, whether their pronunciation is accurate or not, will be able to produce them in connected speech in a similar way. If no other factors (e.g. sociolinguistic rules of /r/ in Thai) are involved, we may expect that the pronunciation errors made in isolated words are also likely to be made in connected speech, although the pronunciation in connected speech may be expected to be less accurate than in isolated words since more information needs to be encoded in producing connected speech. When some other phonological processes that apply only in connected speech (phonological rules across word boundaries, or higher level suprasegmentals) are different in L1 and L2, L2 learners are expected to experience even more difficulties in producing connected speech.

3.1.3 Relationship between perception and production of L2 phonology

The relationship between speech perception and speech production is complicated and has not been well understood. Evidence from studies of child language acquisition (e.g.,

the *fis* phenomenon, Berko and Brown (1960)) shows that perception and production are distinct aspects of language competence. The difference between speech perception and speech production can also be found easily in second language acquisition. Studies on L2 learners' perception and production of L2 sounds have found that L2 learners are able to discriminate L2 sounds without the ability to pronounce them (Thompson 1984), as well as to pronounce L2 sounds acceptably without the ability to distinguish them correctly (Goto 1971).

However, generally, it has been found that L2 phonological perception and production abilities are related in L2 learners. Snow and Hoefnagel-Höhle's (1979) study on the acquisition of Dutch by English-speaking learners in naturalistic environment found that perception and production of L2 sounds were highly correlated after one year of learning, and these skills were separate from the grammatical abilities.

> It is interesting to note how these two components separate themselves out through time. At the first test session, pronunciation loaded on the same factor as several of the grammatical abilities tests, whereas auditory discrimination loaded on a different factor. At time 2, the various grammatical abilities tests showed a relationship both to vocabulary skills and to receptive skills. Pronunciation had been separated out as a factor, but auditory discrimination still clustered with vocabulary and with some of the grammatical abilities tests. At this stage of proficiency, the correlation between pronunciation and auditory discrimination was still very low, suggesting that differences in pronunciation ability might have been the result of motor rather than perceptual

> deficits. By time 3, pronunciation and auditory discrimination correlated highly and formed one factor, suggesting that perceptual ability had become the limiting factor in pronunciation skill. (Snow and Hoefnagel-Höhle 1979:158-159)

Schneiderman, et al. (1988) found that discrimination and production abilities of French phones were significantly correlated in learners of French (beginning and low-intermediate levels) before they underwent a discrimination training program. However, the discrimination and production abilities no longer correlated significantly after the training. Goto (1971) and Sheldon and Strange (1982) both examined the perception and production of English /r/ and /l/ by Japanese learners and found that good speakers perceive the contrast better than poor speakers. Borden et al. (1983) had the same finding with Korean learners.

Williams (1979) found Spanish-speaking children learning English as an L2 showed significant changes in both perception and production of VOT towards the English monolingual pattern. Flege (1988) reviewed studies on VOT and concluded:

> Perhaps the most interesting result of the VOT studies reviewed here is that the same L2 learners who approximated L2 phonetic norms for VOT in production also showed evidence of changing perception: their phoneme boundaries were intermediate to those of monolingual native speakers of L1 and L2 (Caramazza et al., 1973; Williams, 1977b, 1979, 1980; cf. Albert & Obler, 1978; Obler, 1983). This suggests that production and perception evolve together. Many researchers believe that the changes in perception which occur

> in speech acquisition or learning "lead" corresponding changes in production. (Flege 1988:362-363)

Whether changes in perception "lead" corresponding changes in production in L2 learners is a controversial issue. Hypotheses concerning phonological perception and production in children acquiring their native language often presuppose that perception precedes production (see Sheldon and Strange 1982:243), but the evidence from L2 acquisition is quite different from that in child language acquisition.

Studies by Goto (1971) and Sheldon and Strange (1982) found that adult Japanese learners of English were better in producing identifiable /r/ and /l/ sounds than in perceiving the contrast. In Sheldon and Strange's (1982) study, almost all of the /r/ and /l/ sounds produced by five good Japanese learners were correctly identified by English speaking listeners. However, the five Japanese learners made an average of 11% errors on identifying test words produced by Americans.

Evidence that production precedes perception in L2 acquisition has also been reported in learning foreign stop and fricative contrasts (Brière 1966; cited in Sheldon and Strange 1982:254). When this relationship occurs, it has often been found that the L2 learners had been taught to pronounce the sounds in question by explicit reference to articulatory parameters.

It is possible that an L2 learner produces L2 sounds identifiable to native speakers but with non-native features, and his perception of L2 sounds is also based on some of the features that are not used by native speakers; in another words, he may not have the same "prototype" or "perceptual target" as native speakers. For example, Goto (1971:321) noted "It was found by checking the scores that some Japanese subjects got better scores on the voices of the Americans with whom they had had close contact." Mochizuki (1981:290) found that the Japanese speaker in his study perceived the /r/-/l/ contrast she produced herself better than the contrast produced by the American speaker:

> The Japanese speaker of Identification Test II could identify correctly 74% of the time in Identification Test I of an American speaker, and her tokens in Identification Test II were correctly identified 92% of the time by the American listeners. This result conforms to Goto's claim that the production of intelligible /r/ and /l/ is easier than their perception. Yet it is interesting that this Japanese speaker could identify her own tokens in Identification Test II 98% of the time, which is higher than the average identification among the American listeners.

Sheldon and Strange (1982) also found that Japanese speakers generally can identify the /r/-l/ contrast they produced themselves better than they can identify the contrast produced by American speakers.

Mochizuki (1981) also showed that the production of /r/ and /l/ by a Japanese speaker in different word positions is not equally well identified by American listeners; likewise,

Japanese learners' identification of /r/ and /l/ produced by an American speaker in different word positions is not equally good. The order of difficulty according to word position is different for production and perception. In addition, there was an asymmetry in errors made on identifying /r/ and /l/ in initial and consonant cluster positions (Sheldon and Strange 1982). Whereas /l/ was misidentified more than /r/ in initial position, /r/ was misidentified more than /l/ in consonant clusters. Thus the particular test stimuli chosen in a test could have an effect on the scores of perception and production. For example, Goto (1971) used only words with initial /r/ and /l/ or initial cluster with /r/ and /l/; these positions happened to be more difficult for perception but easier for production. This positional difference for perceptual and production difficulty may account for the greater discrepancy between perception and production found in his study.

If perception and production acquisition are related, then an important inference for L2 teaching is that training on one aspect may lead to improvement on the other. Some studies have tried to explore this possibility.

Pimsleur (1963) conducted experiments with high school students on the effect of discrimination training on production. He found that experimental groups which received perception training on the discrimination of French

nasalized vowels /ã/ /õ/ /ẽ/ scored significantly higher than the control group in their production of these phonemes. However, the experimental group which received discrimination training on the distinction between French [o] and American English [o^{u}] did not differ from the control group in their production of French [o].

Mueller and Niedzielski (1968) conducted another experiment on the effect of discrimination training on production with university students taking French. In the first nine weeks of the course the experimental students spent a total of 2 hours on discrimination training in addition to 22 hours of laboratory exercises used by both the experimental group and the control group. The results showed that among the seven variables studied, only two showed significant differences between the experimental group and the control group. The experimental group scored significantly higher in vowels which have a correlate sound in the English system (variable 3), while the control group scored significantly higher in phonemes that do not have a correlate in the English sound system (variable 1). However, the girls in the control group showed significantly higher aptitude, and general impressionistic evaluation (errors involving "mismanagement of allophones" or "deviations from phonemes", fluency, liaisons and the general intonation features, stress and rhythm features)

found the experimental group superior to the control group. Mueller and Niedzielski (1968:416) concluded:

> In view of the fact that the test was heavily weighted towards variable 1, and because the aptitude of the experimental group was so much inferior to the aptitude of the control group (girls), the impressionistic superiority of the experimental group seems to be a valid conclusion and does not seem to be contradicted by the quantitative evaluation. Discrimination training seems to be an effective training device in the learning of pronunciation.

Schneiderman et al. (1988) found that learners of French improved significantly in both discrimination and production after undergoing a discrimination training program, while the control group did not show similar improvement. However, the discrimination and production abilities no longer correlated significantly after the training.

Borden et al. (1983) investigated Korean speakers' perception and production of English /r/ and /l/ before and after two or three weekly 45-minute training sessions. They found that there was some change in the accuracy of /r/ and /l/ production during each training session (from an average of 80% correct on the probe test given at the beginning of each training session to an average of 89.4% correct on the post-training probe test), but "there was not a significant difference between the more extensive speech production test given before training and the final test, indicating that there was little carryover of training".

However, despite the lack of carryover effect of perceptual training on production, there was an indication that perception (with criteria used by native L2 speakers) precedes production (Borden et al. 1983:516):

> When subjects judged the correctness of their own [r] and [l] productions as compared with the taped model, the extent to which their judgments were in agreement with the judgments of the experimenter correlated significantly and positively with their production scores.
>
> Where there is a marked difference between the two final scores among speakers undergoing more extensive change in their speech patterns, it can be seen that self-perception scores were always higher than production scores. This indicates that the ability to perceive errors in one's own speech may precede correct production in the timetable for learning new phonemes, or that self-perception is simply an easier task than production of new speech sounds.

On the other hand, Catford and Pisoni (1970) found that subjects receiving articulatory training not only scored higher in production but also in discrimination than subjects receiving only aural training. They stated: "This obviously implies some kind of carry-over from productive competence to auditory discriminatory competence, and may, indeed be taken to be some support for a "motor theory of speech perception" (p.481). Because of lack of understanding of how perception and production interact, and since most of the training programs involved both perception and production, the results of these studies did not show conclusively what effect training in one aspect has on the other.

3.1.4 Acquisition of L2 phonology over time

L2 learning takes time. In this section, we will examine the time effect of L2 phonological learning. Perception and production will be discussed separately.

3.1.4.1 Perception

Adult L2 learners can achieve high perceptual ability similar to that of native speakers after years of experience in the L2 environment. Neufeld (1980) reported that advanced English speakers of French who accessed themselves as high in oral-aural skills in French were able to detect traces of foreign accent in French speech samples just as French native speakers can. However, less proficient English speaking learners of French were not able to do the same; they performed only at chance level.

Thompson (1984) found fluent L2 learners of English (who had lived in the USA from two to 42 years with an average of 22 years) achieved high perception ability, which was not much different from that of native speakers, as measured by discrimination and identification tests of English words.

Flege and Hillenbrand (1984) found advanced adult L2 learners who had studied French for at least 4 years in school can discriminate French /tu/ and /ty/ syllables produced by native French speakers about 90% of the time.

The effect of L2 experience on the perception of new L2 sound has been found in the learning of /r/ and /l/ by Japanese speakers. Sheldon and Strange (1982) found their Japanese subjects did better in identifying /r/ and /l/ produced by American speakers than the subjects in Goto's (1971) study. Part of the reason for this difference is that their subjects were in the United States and had more experience speaking with native English speakers.

This effect of L2 experience with native English speakers on the perception of /r/ and /l/ by Japanese speakers has also been found in the categorical perception of synthetic /r/ and /l/ stimuli.

> MacKain et al. (1981) found that Japanese learners with extensive experience speaking and hearing English, but not relatively inexperienced Japanese learners, identified and discriminated synthetic /r/ and /l/ stimuli much like native English speakers. The experienced Japanese learners showed a sharp crossover from /r/ to /l/ judgments as F3 onset frequency changed, and a peak in percent correct discrimination near the /r-l/ phoneme boundary. (Flege 1988:331)

Studies on L2 learners' perception of voicing in terms of VOT in prevocalic stop consonants also provide evidence for the effect of L2 experience on perception. Different languages use different VOT in perception and production of voiced and voiceless stops (Lisker and Abramson 1964; Williams 1979, 1980). Williams (1980) found that Spanish-speaking children with varying degrees of exposure to English in a second language environment (0-6 months, 1.5-2

years, and 3-3.5 years) labelled VOT series toward the English-like pattern gradually with increasing exposure.

Gass (1984) also found that the perception of VOT by adult L2 learners of English changes as a function of time. A group of L2 learners enrolled in an English-language training program were tested on their perception of VOT in a labelling task in an English set three times at one-month intervals. These subjects had little prior exposure to spoken English by native speakers, and their native language differed from English in VOT boundaries between stops. In contrast to native English speakers, the L2 learners' perception did not show a narrow categorical boundary; instead, there was a larger area of uncertainty, with gradual rise and slope reversals. Although the abruptness of the slope and the number of slope reversals did not change significantly over two months for the whole group, individual data from most subjects exhibited fewer slope reversals or more abrupt curves.

Short term L2 perceptual training has sometimes been found to have an effect on L2 perception. Gillette (1980) found that the identification of /r/ and /l/ by Japanese and Korean learners can be improved by training:

> Gillette (1980) studied the effect of a five-week classroom training program on the perception of English /r/ and /l/ by three native speakers each of Japanese and Korean. The training consisted of articulatory descriptions of /r/ and /l/, individual and group practice in producing /r/ and /l/, and auditory evaluation by the students of the extent to which their productions

> of /r,l/ differed from that of fellow students and the native English-speaking instructor. Their students' correct identification of /r/ and /l/ in minimal-pair words increased from 70% to 92%. (Flege 1988:330-331)

There are also indications that perception of VOT can be modified after training:

> In one study (Lisker, 1970) discrimination training enabled native speakers of Russian, a language with lead and short-lag stop categories, to label synthetic short- and long-lag stimuli as /ba/ and /pa/. (Flege 1988:354)
>
> [A] later study (Pisoni et al., 1982) showed that native English-speaking subjects could be trained to identify three categories of stops through the alternating presentation of "good" examplars of the three categories (i.e., stimuli with -70, 0, and 70 ms VOT values; see also McClasky et al., 1983). (Flege 1988:355)

However, the effect of short-term perceptual training seems to be limited in other studies. For example, Pimsleur (1963:203) stated:

> It was found in the pilot studies that students could not be trained to discriminate reliably beyond a certain point. Some students still failed, even after considerable training, in the task of telling whether *vont*, *vent*, or *vain* appeared in the sentence they heard: *Ils vont venir tout l'heure*.

Strange and Jenkins (1978) found that the perceptual training of VOT had only limited effect (cited in Flege 1988:354-355):

> Strange (reviewed in Strange & Jenkins, 1978) attempted to train English speakers to identify members of a lead category in a continuum of CV syllables synthesized with VOT values ranging from -150 to +150 ms. Native English speakers typically

> hear only two categories in this stimulus range. Six hours of training with either immediate or delayed feedback resulted in only limited improvement in subjects' ability to discriminate lead and short-lag stops, and training did not generalize to untrained places of articulation.
>
> Strange (Strange & Jenkins, 1978, p.152) concluded that "changing the perception of the VOT dimension...is not easily accomplished by techniques that involve several hours of practice."

Tees and Werker (1984, cited in Mackain 1988:57) found that after one year of experience with Hindi as a second language, the number of native English speaking subjects who reached the discrimination criterion of Hindi voice and place (retroflex vs. dental) contrasts (two of 28 prior to second-language learning) did not increase significantly.

It is clear from the literature that L2 phonological perception improves with exposure. However, the change can be very gradual, and the effect is likely to be limited with short-term exposure.

3.1.4.2 Production

Studies on the acquisition of L2 phonology in an L2 environment generally found that accent is inevitable within the first two years of living in the L2 environment, even for young children who are likely to acquire native-like pronunciation (Tahta, et al. 1981a, Snow and Hoefnagel-Höhle 1977). Studies that found a significant year of residence effect usually include subjects with less than two years of residence (Asher and Garcia 1969), while studies that found

no effect of length of residence investigated subjects with more than two years of residency (Tahta, et al. 1981a, Oyama 1976).

Snow and Hoefnagel-Höhle's (1977) study on L2 learners' pronunciation over a 10-11 months period in a naturalistic L2 environment showed that adult learners made smaller improvement in pronunciation than children, and that the improvement made during the second 4-5 months was smaller than that made during the first 4-5 months. Although from the literature we cannot be sure if and when adult L2 learners stop making improvement (and there may be some variation among learners), it seems that after two years of active learning in the L2 environment, adult L2 learners either stop making any improvement in their L2 pronunciation, or the improvement is too small to be perceived by native speakers as significant.

Compared to a second language environment, a foreign language environment is clearly not ideal for L2 pronunciation learning. Learners have few or no native-like models on which to base their hypothesis about target pronunciation. Even if native-like models are available, the number is likely to be small and may not provide adequate variation for learners to develop native-like phonological perception and production targets. Since production in adult L2 learners seldom closely resembles that of their models even in a second language environment

where native speaker input is readily available, their deviation from authentic pronunciation can be expected to be even greater in a foreign language environment.

Flege and Hillenbrand (1984, also Flege 1987) found that experienced American learners of French (professors who had used French on a fairly regular basis for an average of 10 years and had spent an average of 1.3 years in France) produced French /t/ with shorter (more French-like) VOT values than relatively inexperienced learners (undergraduate students who had spent the previous academic year in Paris), but the difference did not reach significance, and none of them produced /t/ with an average VOT value produced by monolingual native speakers of French. The more experienced learners also produced a more effective contrast between French /u/ and /y/ than less experienced learners. However, the two groups did not produce French /y/ differently: /y/ produced by experienced and inexperienced groups was correctly identified by French speakers an average of 71.5% and 70.2%, respectively, compared to 93.7% correct identification of /y/ produced by native French speakers. Only two out of seven speakers in each American group produced /y/ correctly over 93%. The confusion between /u/ and /y/ by less experienced learners was a result of their production of French /u/ (in *tous*) more like French /y/ rather than French (or even English) /u/, suggesting that they may have confused the identity of the vowel in *tous*.

The more experienced learners produced French /u/ very much like their English /u/ (as measured by F2 frequency). "Thus even after many years of experience speaking French they seem to have done little to modify their production of /u/ in the direction of French phonetic norms" (Flege and Hillenbrand 1984:196).

Since both groups of American learners were living in their native country and using English as their primary language when the study was conducted, it appears that L2 pronunciation does not improve much with experience in a foreign language environment after learners have studied the language for many years and gained considerable fluency, as might be expected from the studies on L2 pronunciation in L2 environment.

It should be noted that, in the studies cited above (Flege and Hillenbrand 1984, Flege 1987), native speakers of French who had lived in an English-speaking environment for an average of 12.2 years produced French /t/ with longer (more English-like) VOT values than those produced by monolingual French speakers. Likewise, another group of native English speakers, who had lived in France for an average of 11.7 years and used French as their principal language, produced English /t/ with shorter (more French-like) VOT values than those produced by monolingual English speakers. Thus the language environment a speaker is in has an effect on the pronunciation (at least VOT in stops) of

not only L2 but also L1. This is consistent with what was described by Obler (1982) as "the parsimonious bilingual". Obler (1982) reported that balanced bilinguals who spoke both English and Hebrew without an accent produced stop consonants in the two languages with different VOT values. However, they did not produce the consonants exactly like the monolinguals. Rather, they produced English voiced consonants closer to the Hebrew monolingual norms, and the Hebrew voiceless consonants closer to the English monolingual norms. Table 2 from Obler (1982:342) is reproduced below to show the VOT means produced by monolingual and bilingual groups.

Table 2. Group Voice Onset Time Means for Production

	Unilinguals		Bilinguals	
	Hebrew group	English group	Hebrew condition	English condition
/p/	+25.6	+77.6	+58.0	+68.8
/b/	−110.8	−8.5	−96.8	−64.8
/t/	+33.9	+77.0	+73.8	+89.6
/d/	−95.8	−6.8	−85.5	−55.8
/k/	+63.7	+89.4	+93.2	+99.4
/g/	−101.0	+15.1	−92.2	−58.3

Thus, bilinguals tend to maximize the contrast and produce equivalent sounds in two languages closer to each other.

Gass (1984) found that adult L2 learners of English produced stop consonants /b/ and /p/ with VOT values mostly within the range of possible English norms. Moreover, "when the VOT value does not fall within the possible English range, it is often not within the native language range, either. Rather, it is in a direction away from the native language" (p.69). The difference of VOT values in /b/ and /p/ produced by native English speakers and L2 learners showed mainly in the amount of variation. It was found that the production of VOT values by L2 learners, "while on the surface for the most part appearing nativelike, showed a trend from lesser variability to greater variability and back to lesser variability" (p.71).

A difficulty frequently encountered in the study of L2 pronunciation is that variations is found across speakers as well as within a speaker. For example, Borden et al. (1983:509) found Korean speakers' pronunciation of [r] and [l] as judged by the experimenter was not linear over three or four weekly testing sessions. Two subjects improved each time, two subjects improved at first and then regressed, and four subjects got worse before they improved. The causes for the non-linear development may be quite complicated, involving not only the similarities of phonetic and phonological properties of L1 and L2 sounds but also the organization of the phonological system or the distance between neighboring phonemes (Major 1987).

It is clear from the studies on /r,l/ learning by Japanese speakers that articulatory training making specific reference to articulatory parameters can help adult L2 learners to produce foreign sounds acceptably even though they may still have difficulty perceiving the contrast. Catford and Pisoni's (1970) study suggests that articulatory training is more effective than aural training in helping learners produce "exotic sounds" immediately. In their study subjects who received about two hours of articulatory training produced the exotic sounds better than subjects who received only aural training (without seeing the mouth of the teacher).

The acquisition of English /r,l/ by L2 learners also suggests that experience in a second language environment, even after many years of L2 learning in a foreign language environment, has an effect on the accuracy of production. The Japanese subjects in Sheldon and Strange's (1982) study (who had resided in the USA for 14-39 months) produced fewer errors than subjects in Goto's (1971) study (of whom only two had spent years in America). In a study by Borden et al. (1983) on Korean learners, "the half with better scores averaged about 18 months in the United States and the half with lower scores averaged almost 6 months" (p.509). This is consistent with Purcell and Suter's (1980) finding that length of residency in an L2 speaking environment is

significant in predicting L2 pronunciation, while the amount of formal L2 classroom training is not significant.

The study of Pimsleur (1963) found a difference in the effect of discrimination training on the production of phonemic distinction and the phonetic approximation in the learning of L2 sounds. Production of French nasalized vowels /ã/, /õ/, /ẽ/ improved significantly after training while production of the French vowel /o/ did not. Beebe (1987:168) believed that most pronunciation errors do not involve confusion of two phonemes. She stated that "phonemic contrasts are a very minor problem for intermediate and advanced learners. Phonetic inaccuracy, compared with a TL norm, is extensive. The phonetic deviance is heard by the teacher/listener as a phonemic error." That the advanced learners' pronunciation problem is one of phonetic inaccuracy rather than of phonemic contrast is also found in the acquisition of tone. In a study by Ioup and Tansomboon (1987), the production of Thai tone by advanced Thai learners (who had extensive formal instruction in Thai and had lived many years in Thailand as missionaries) was judged to be phonemically correct 95% of the time, compared to only 58% for the beginning learners (who were taking second quarter of elementary Thai at a university). However, only 8% of the tone produced by the advanced learners, and 3% by the beginning learners, was

judged to be phonetically indistinguishable from native Thai speakers.

3.2 Perception and production of Mandarin tones by adults L2 learners

There are only a few studies investigating the perception or production of Mandarin tones by English-speaking adults. Tone perception and production will be reviewed separately in this section.

3.2.1 Perception

Winitz (1981) conducted an experiment on the perception of Mandarin tones and sounds by monolingual English-speaking children (third-graders) and adults (college students). In the three contexts (isolation, disyllable, and sentence) tested, the adults achieved higher scores than did the children. Context was found to be significant, indicating that the tasks increased in difficulty from isolation to disyllable to sentence. Only in the perception of sounds, the adults showed no decrease between the disyllable and sentence contexts.

Broselow et al. (1987) tested English speakers' perception of Mandarin tones. A tape consisted of training and identification tasks was made by a Mandarin speaker and was presented to English speakers who had not studied any tone languages. The brief training introduced the four

lexical tones of Mandarin with the syllable *ti*, and the syllable was produced three times using each tone. The identification tasks consisted of single syllables and series of two and three syllables. A single example was presented before each set of identification trials, and subjects were given feedback after the presentation of singlets and doublets.

Tone 4 was found to be the most easily perceived tone in single tones (one syllables). Its index of detectability was significantly different from that of all the other tones. The percent of correct responses declined for all tones in longer strings. However, Tone 1 and Tone 2 showed no dramatic positional difference in the percent correct responses in 2- and 3-syllable strings, while Tone 3 and Tone 4 showed a significant difference in detectability between different positions. For both Tone 3 and Tone 4, detectability in the final position was significantly greater than detectability in non-final positions. The difference of detectability in Tone 3 can be attributed to allophonic variations in different positions. The positional difference of detectability in Tone 4, however, the authors argued, was a result of interference from English intonation. Since the contour of Tone 4 corresponds to the fall at the end of declaratives in English, "[t]he relative ease of detectability of the fourth tone in string-final position is a result of positive transfer: a familiar

item occurring in a familiar position is easy to hear. The difficulty of hearing the fourth tone in non-final position, on the other hand, is a result of negative transfer; the high falling pitch contour normally does not occur in non-final position in English strings" (p.357).

Kiriloff (1969) tested Australian students' perception of Mandarin tones in isolated syllables after two months of Mandarin learning and again after four and a half months of learning. The students were asked to listen to a recording of 20 syllables pronounced by a native speaker and to transcribe the sounds with the Pinyin symbols. The average results showed no significant difference in tone perception between the two tests. However, in a third test shortly after the second test, the students were asked to identify the tones only. The results showed that identification of tone was significantly better in the third test than in the first two tests. The χ^2 test on the errors showed that "[e]rrors involving the 2nd tone are considerably higher and those with the 1st and 4th tones much lower than expected." Thus, Kiriloff maintained that "[t]he considerable number of incorrect identifications of the second tone and a marked tendency to confuse with it the third tone indicate that the second tone presents the greatest difficulty to students of Mandarin" (p.65).

W.C.J. Lin (1985) tested tone perception in beginners (150 contact hours), intermediate (600 contact hours), and

advanced (1050 contact hours) students of Chinese. The four tones over the same segmental sequence *mā* 'mother', *má* 'hemp', *mǎ* 'horse' and *mà* 'to scold' were put in the sentence frame *Wo shuo de shi* __ ("What I said is __"), and *Tade* __ *hen youming* ("His/her __ is quite famous"). The correct identification for tones in the sentence final position was 70.83% for beginners, 83.33% (but I get 91.67% based on the figures given in his Table 2) for intermediate students, and 94.04% for advanced students. The correct identification of tone in medial position was 75% for beginners, 86.46% for intermediate students, and 92.85% for advanced students. For all groups of students and in both sentence frames, there were more errors in identifying Tone 3 than the other tones.

Lin proposed a new treatment of Tone 3: introducing the norm of Tone 3 as low-dipping (the allophonic contour occurs in non-final position) instead of the citation form low-dipping-rising. The reason was that "the tonal contour of low-dipping-rising is realized only when it occurs in isolation, in contrast with another syllable with the same sounds but different tone, or in the case of emphasis. The frequency of occurrence of these cases is very low. In the overwhelming majority of cases, including its occurrence at the end of an utterance or before a pause, the 3rd tone simply dips without rising" (p.33-34). (M-Y. Chen 1983 also found that low-dipping occurs 83% of the time.) After the

new treatment of Tone 3 was explained and some drills were given, the test in the sentence frame *Tade __ hen youming* was given. The results were 83.33% (90.97% according to my calculation) correct identification for beginners, 95.83% for intermediate students, and 98.81% for advanced students. There was a clear improvement in the identification of tones in general and improvement in the identification of Tone 3 in particular. For students who had been taught the new treatment of Tone 3 since the very beginning, the results were even better: 97.22% for beginners, 98.96% for intermediate, and 100% for advanced students.

3.2.2 Production

Neufeld (1978) reported that many English speaking adults achieved native-like pronunciation imitating Mandarin utterances after 18 video-taped one-hour training sessions which focused exclusively on phonetic material. However, on a five-point scale, only one out of twenty subjects received a five in Mandarin, and eight received a four (appears native with occasional English-like sounds). Thus his claim that "nine were judged to be native speakers of Chinese" (Neufeld 1979:234) is questionable.

Shen (1989) investigated eight American students' production of Mandarin tones in reading a text they were familiar with. Shen argued that the acquisition of tonal shapes (contours) of Mandarin does not constitute a

difficulty for L2 learners, but the tonal registers make the mastering of tones a difficult task. She agreed (Shen 1989:27) with Zhao's (1987) statement about errors in tonal register:

> the register in tonal production functions like the key in singing: if, when singing in the key of C, the singer changes suddenly to the key of F, the audience hears a deviate tune; by the same token, when the L2 learner utters a high register tone with a low register, in spite of the accuracy of the tonal shapes, L1 listeners perceive an inaccurate tonal realization.

In her study, the production data were tape recorded and judged by four native speakers regarding the correctness of the tone. When a tone error occurred (as judged by at least three judges), the type of error (register or contour) was also auditorily analyzed by the judges. The eight subjects can be divided into two groups: the group of four better L2 speakers made significantly fewer errors than the group of four worse L2 speakers. The differences in errors among the five tones (including neutral tone) were not significant for the group of better L2 speakers, but the differences were significant for the group of worse L2 speakers. "However, as for the overall errors of the two groups, the differences of the errors among the five tones are significant" (p.32). It was found that the rates of errors in the production of Tone 1 and Tone 4 were significantly higher than those of Tones 2, 3, and 0 (neutral tone). The rate of errors of Tone 4 was significantly higher than that of Tone 1, but the rates of errors of Tones 2, 3, and 0 were not significantly

different. The auditory analysis found that most errors were register errors rather than contour errors. Acoustic analysis was also carried out on tones produced by one of the subjects (the one that made the most errors), and the results were found to be in agreement with those from the auditory analysis.

Miracle (1989) also studied American students' production of Mandarin tones. His speech samples were 1- or 2-syllable words taken from sentences read by subjects. Only the words in the topic (sentence initial) position were analyzed. The tones were analyzed acoustically. The F0 range of each subject was calculated using the highest and lowest F0 found in the voiced portion of the words analyzed. Tonal errors were classified by him as either tone register errors (based on each individual's tonal space) or tone contour errors based on comparison with the F0 contours obtained from native speaker models. He found that the errors were fairly evenly divided among the five tones (including the neutral tone) and were evenly divided between tone register errors and tone contour errors. The overall error rate was 42.9%. In the tone combination (2-syllable) words, the error rate in the initial syllables (49.6%) was significantly higher than that in the final syllables (32%).

Shen (1989) and Miracle (1989) appeared to have quite different findings about American learners' learning of Mandarin tones. Miracle noted that Shen's subjects had only

had one semester of Chinese while his subjects had been studying Chinese for almost twice as long. "This fact may suggest that while the first and fourth tones are initially troublesome, with further study, the difficulties even out across all the tones." (p.56) However, other differences may also have contributed to their different results. The evaluation procedures adopted in the two studies were quite different. Miracle based his analysis entirely on acoustic information, while Shen's study was based primarily on the auditory judgement of four Chinese judges. Although acoustic analysis was also carried out by Shen on the production of one subject and the results were found to be in agreement with those from the auditory analysis, it has generally been found that physical differences in acoustic signals do not always correspond to differences in human perception. For instance, a slightly falling F0 contour has frequently been produced as Tone 1 by native speakers in some environments (Miracle, 1989; Shen 1985; Dreher and Lee, 1968) and would be perceived as Tone 1 (level) by native Mandarin speakers (Tseng, 1981). Furthermore, the F0 range of each subject in Miracle's study was based on the Mandarin words analyzed (words taken out from sentence initial position), which may not be the same range if the subject were a native Mandarin speaker.

Another important difference between the two studies was the material read by subjects. Miracle restricted his

data to words in the topic position "to reduce the possible interference of English sentence intonation" (p.51). On the other hand, the sentences read by Shen's subjects were longer and connected. For example:

Yǒu shíhou tā gēn péngyoumen qù kàn diànyǐng,
Sometimes she and friends go see movie,

yǒu shíhou gěi bàba, māma hé péngyou xiě xìn.
sometimes to Dad, mom and friend write letter

These sentences were very likely to have been influenced by English sentence intonation.

In long stretches of connected speech, some tonal contours (and other phonological features) may not be preserved well in native speakers' production because of the shorter duration of syllables in connected speech (Tseng,1981; Lyovin, 1978). Thus in the perception of tone in long stretches of connected speech when contour features are not clear, tonal register may be more important than tonal contours (Massaro et al. 1985). Although we do not know how fast the American subjects' in Shen's study read the text, it is quite possible that at least some syllables were short since the text they read was familiar to them. If that was the case, then it is not surprising that more register errors were found by the Chinese judges.

Basically, studies on Mandarin tone learning by English speakers were concerned with the errors the learners made, the source of errors, and the ways to correct the errors. Only Kiriloff's (1969) study compared the results of

identification tests conducted at two different times and found no improvement.

4. ISSUES AND HYPOTHESES

It has been mentioned in the Introduction that the purpose of this study is to see how tones are perceived and produced by English speaking adults at the beginning of their learning of Mandarin. More specifically, this study will try to answer the following questions:

(1) Is tone perception and production affected by variables of tonal context, task, sex, and L2 learning experience?

(2) Is there a relationship between the perception of tone in isolated syllables and in sequences, and a relationship between the production of tone in isolated syllables and in sequences?

(3) Is there a relationship between tone perception and tone production?

(4) Do tone perception and tone production improve over one term of learning?

(5) Among the four Mandarin tones, is there a difficulty order for English speaking learners? If there is, what can account for the order of difficulty?

(6) Are there general patterns in errors of tone made by English speaking adults learning Mandarin? If such patterns can be found, what accounts for them?

These question are important questions for researchers in second language acquisition and for teachers of Mandarin as a second language.

The first question concerns two external variables (phonological context and task) which have been found to affect L2 performance. But no studies have been done on the effect of the particular phonological context (isolated single syllables, 2-syllable words and short sentences) and task (imitation and reading aloud), or on two learner variables (sex, L2 experience) in the perception and production of Mandarin tone by L2 learners. The second question concerns an assumption often made by L2 researchers and L2 teachers that the perception or production of a sound or tone in its citation form is closely related to its perception or production in utterances. But few studies have been done on this. The third question is about the relation between perception and production, which has often been asked in L2 acquisition research as well as in other related areas (e.g., L1 acquisition and psycholinguistics) because of its theoretical importance and its implication for L2 learning and teaching. The fourth question concerns the development of L2 phonological perception and production in L2 learners, which may be especially important in developing an explanation for the wide-spread phenomenon among adult L2 learners of foreign accent. The fifth question is about the difficulty order of the four Mandarin

tones for L2 learners, and the sixth question is about the type of errors L2 learners make. Both questions have been investigated by L2 researchers in many aspects of L2 acquisition (phonology, morphology, syntax, etc.) in many different languages, and theories of L2 acquisition have been proposed and modified based on the findings of these studies. Findings on the difficulty order of Mandarin tones and the type of errors L2 learners make in learning tone also need to be accounted for by theories of L2 acquisition, as they may provide valuable evidence for or against particular theories.

In this chapter, based on the literature reviewed in the previous two chapters, I will propose some hypotheses concerning the above questions. Except for question (1), which includes four variables each of which corresponds to one hypothesis (Hypotheses 1 to 4), questions (2) to (4) each corresponds to one hypothesis. Thus, Hypothesis 5 correspond to question (2), Hypothesis 6 to question (3), Hypothesis 7 to question (4). These hypotheses will then be examined with the data obtained. No specific hypotheses will be proposed for questions (5) and (6), concerning the difficulty order of Mandarin tones and error patterns, which will be obtained by analysis of the data.

4.1 Effects of some variables on L2 tone perception and production

In research on variables that predict accuracy of L2 pronunciation, while age of arrival in L2 environment has been found to be the best predictor for accuracy of pronunciation (Asher and Garcia 1969, Oyama 1976, Seliger et al. 1975, Thompson 1984), native language (L1) was found to be the best predictor for foreign accent among adult L2 learners (Purcell and Suter, 1980). There are few studies that investigate variables predicting L2 phonological perception. One study found ear preference to be the best predictor for perception ability but it was not a strong predictor (Thompson 1984).

In this study, all subjects are native speakers of American English, and all subjects have passed puberty (the presumed critical or sensitive period for achieving native-like accent) and the age range is relatively narrow. Individual differences are kept constant by testing the same subjects twice on tone perception and production, although some subjects were absent or dropped out for the second test. The variables to be analyzed in this study are tonal context (tone in isolated syllables, 2-syllable words and short sentences), task (imitation and reading), sex, and L2 learning experience.

4.1.1 Tonal context

It is noted by many experienced teachers of Mandarin as an L2 that many students can pronounce Mandarin tones in citation forms quite well, yet cannot do as well when speaking a sentence. The same may also be true in tone perception. Winitz (1981) and Broselow et al. (1987) found that English speakers with no knowledge of Mandarin perceived Mandarin tones in citation form better than in two or more syllables.

In this study, tone in three tonal contexts (isolated syllables, 2-syllable words and short sentences of three or four syllables) will be studied. The difference in utterance length may affect L2 perception and production (see 3.1.1.1), and other factors, such as transfer of L1 suprasegmental patterns (Odlin 1989), are also expected to come into play in different contexts. For instance, English sentence intonation is more likely to be transferred to a Chinese sentence than to a syllable or a word. The English stress pattern of 2-syllable words is more likely to be transferred to 2-syllable Chinese words rather than to a 1-syllable words. Thus the perception and production of Mandarin tones in isolated single syllables are expected to be easier for English learners of Mandarin than perception and production in 2-syllable words or short sentences.

Hypothesis (1): Subjects perceive and produce tones with significantly higher scores (see Chapter 5 below)

in isolated single syllables than in 2-syllable words and sentences.

4.1.2 Task

L2 learners' performance can vary for different tasks (L. Dickerson 1975, Beebe 1987, Sato 1985). The production data in this study will be collected on two different tasks: imitation and reading from a list. Both imitation and reading are common tasks used in soliciting subjects' speech samples because the control of other variables (e.g. grammar, vocabulary, phonological context) is easier in imitation and reading than in free speech. However, if the two tasks result in different performance, then results obtained from studies using different tasks may not be directly comparable.

It is commonly observed that beginning L2 learners can pronounce L2 sounds much better when repeating immediately after an auditory model than when pronouncing the same sounds without such a model. Data from a study using both methods showed pronunciation in imitation better than that in spontaneous production (Snow and Hoefnagel-Höhle 1977, see 3.1.1.2). In L2 tone production, just as in other aspects of L2 pronunciation, imitation is expected to produce better results than reading, at least in the beginning.

Hypothesis (2): Task, imitation or reading, affects subjects' performance in tone production: imitation results in significantly higher scores in tone production than does reading.

4.1.3 Sex

Sex has often been included as a variable in studies of L2 acquisition probably because a sex difference is usually present and many people believe that females are better than males in learning languages (e.g., Bolinger 1975:333). It will be interesting to see if tone learning is affected by sex.

As reviewed in 3.1.1.3, sex was found to be a significant variable in L2 pronunciation in some studies (Ash and Garcia, 1969, Thompson 1984), in which females were found to have better pronunciation than males. However, the effect seemed to diminish with length of residence (Asher and Garcia 1969). In some other studies (e.g. Purcell and Suter 1980) sex was found not significant in L2 pronunciation. There is no evidence that L2 phonological perception is affected by sex.

Hypothesis (3): Subjects' performance in Mandarin tone perception or production is not affected by their sex.

4.1.4 L2 experience

Knowledge of a previously learned L2 has been reported to facilitate the learning of another closely-related L2 (Rivers 1979, Singleton and Little 1984). Some people believe that learning a second language makes the learning of subsequent languages much easier (Larsen-Freeman and Long 1991:206). However, there is no evidence in the literature on L2 acquisition that this is true in the learning of phonology (see 3.1.1.4).

Originally it was planned to test this hypothesis by grouping subjects into two groups: those with previous L2 learning experience and those without such experience, assuming that students taking Chinese 101 have no prior Chinese learning experience and no experience in learning another tone language. However, it turned out that many of the students had taken Chinese classes before, either in high school or in college, though for no more than one year. It was decided to include them in the study as well, but they are classified into a third group: those with Mandarin learning experience.

It is expected that students with prior Mandarin learning experience will do better since they had already studied the tones before. Students with other L2 learning experience and students with no L2 learning experience are not expected to learn tones differently.

Hypothesis (4): (A) Subjects with prior Mandarin learning experience score significantly higher in Mandarin tone perception and production than those without such experience. (B) For subjects without Mandarin learning experience, other L2 learning experience does not affect performance in tone perception or production.

4.2 Relationship between the perception and production of tone in isolated syllables and in sequences

In most beginning Chinese classes, only tone in isolated syllables and the tone sandhi rule for Tone 3 (Tone 3 becomes Tone 2 before another Tone 3, see 2.1.2) are specifically taught. In addition, some proponents of teaching tones in sequences emphasize the third tone and the neutral tone, which are realized phonetically differently depending on the preceding or following tones (see 2.1.2).

The focus of teaching tone in isolated syllables is not surprising since multisyllabic words or phrases and sentences are basically a succession of syllables with their lexical tones. If there is a strong positive correlation between the acquisition of tone in isolated syllables and in sequences, then the teaching of tone in isolated syllables may be quite adequate. If, however, there is no significant correlation or the correlation is weak, then the teaching of

tones in connected speech may need to be done more than is usual.

Acoustic studies on Mandarin tones found that the basic tonal contours of the four tones generally remain the same in isolated 2-syllable words and sentences as in their citation forms (Ho 1976a, M-Y. Chen 1983). However, in spontaneous running speech or in fast speech, the tonal characteristics may not be well preserved in all syllables (Tseng 1981, Wu 1982). As a result, tone in syllables taken out of such context becomes difficult to identify even for native speakers (Tseng 1981, Lyovin 1978). Therefore sequences in this study are normal speed utterances of limited length, 2-syllable words and short sentences of three or four syllables; this eliminates the tonal variation that results in longer sentences (see 2.1.3).

In this limited definition of tone in sequences, the tonal characteristics of tone are expected to be clearly produced, and the identification of tone based on acoustic signals alone should be possible, just as is the identification of tone in isolated syllables. Two to four syllables should be within the short term memory span of an average adult, especially when all his attention is directed to the tone. Native language may have some different effect on the perception or production of tone in isolated syllables and in sequences (see 4.1.1), but the effect is expected be similar for subjects speaking the same L1. It

is assumed that the ability to identify a tone in isolated syllables is a prerequisite to the ability to identify tone in sequences. Thus it is expected that those who do better in identifying tone in isolated syllables will also do better in identifying tone in 2-syllable words and short sentences.

On the production side, the ability to pronounce tones in isolated syllables correctly also seems to be a prerequisite to pronouncing tones in a sequence of two to four syllables correctly. Thus it is expected that L2 learners who pronounce tones in isolated syllables well will tend to do better in pronouncing tones in 2-syllable words and short sentences.

> **Hypothesis (5):** There is positive correlation between perception of tone in isolated single syllables and syllable sequences, and positive correlation between production of tone in isolated syllables and syllable sequences.

4.3 Relationship between L2 tone perception and tone production

Phonological perception and production are distinct aspects of language competence. L2 learners have been found to be able to discriminate L2 sounds without the ability to pronounce them (e.g. Thompson 1984) as well as able to

pronounce sounds acceptably without the ability to distinguish them fully (e.g. Goto 1971).

However, generally, it has been found that L2 phonological perception and production ability are related in L2 learners. L2 learners who have better pronunciation also perform better in sound perception and vice versa (see 3.1.3). Tone belongs to suprasegmental phonology yet it is similar to segmental phonology in that it is distinctive at the lexical level. Based on previous studies of phonological perception and production in L2 learners, it is expected that L2 tone perception ability and tone production ability are related. Learners doing better in one test (e.g. tone in isolated syllables) of tone perception are expected to do better in the counterpart tone production test.

It will also be interesting to compare general performance in tone production and tone perception since in L2 phonological learning the question of whether perception is ahead of production is a controversial issue. In this study, only the distinctive (phonemic) aspect of tone will be analyzed.

Hypothesis (6): There is positive correlation between subjects' Mandarin tone perception scores and tone production scores.

4.4 Improvement of tone perception and production

It is usually expected that an L2 learner's L2 competence will improve with experience over a long period of time. However, concerning the acquisition of L2 phonology by adults, it is not clear how long L2 phonology will continue to improve. The commonly observed phenomenon that adult learners typically speak an L2 with a foreign accent seems to indicate a cessation of improvement in their L2 phonological learning.

Studies investigating the L2 pronunciation accuracy of subjects who have lived in the L2 environment for many years have usually found little effect of length of residency (Oyama 1976, Tahta et al. 1981a). However, for L2 learners who have lived in the L2 environment for less time there seems to be improvement with increased experience (Borden et al. 1983). There is also evidence that adults learn faster and have better control of pronunciation than children at the beginning of L2 learning (Olson and Samuels 1973, Snow and Hoefnagel-Höhle 1977). However, adults' improvement in pronunciation slows down after the initial stage and children's pronunciation becomes better than adults' in less than a year (Snow and Hoefnagel-Höhle 1977). Even for younger learners who are likely to improve more than adults and eventually achieve a native-like accent, the learning process can take years. It has been found that native-like pronunciation generally cannot be achieved by L2 learners

regardless of age within two years in the L2 environment (Tahta, et al. 1981a, Snow and Hoefnagel-Höhle 1977). Studies on L2 learners' pronunciation have found a gradual approximation toward the native speakers' norm (Flege 1987). All these studies indicate that the learning of pronunciation is a gradual process. It can be expected that the progress of adults after the initial stage will be even slower if they indeed continue to improve as children do.

On the perception aspect of L2 phonology, there is evidence that even bilinguals who can speak two languages with no accent have different perceptive patterns from monolingual native speakers (e.g. categorical boundary of VOT) (Obler 1982). Nevertheless, as far as the identification of real L2 sounds is concerned, advanced L2 adult learners have been reported to have good perceptive abilities that are not much different from those of native speakers (Flege and Hillenbrand, 1984). However, there is also evidence that no significant change in perception occurs after one year of foreign language learning (Tees and Werker 1984). W.C.J. Lin's (1985) test results of Mandarin tone perception by beginning (150 hours), intermediate (450 hours) and advanced (600 hours) learners suggested a continual improvement in learners as a group.

Studies on the variables contributing to L2 pronunciation in a second language environment have typically found little effect of formal L2 teaching (e.g.,

Purcell and Suter 1980), and studies on the perception of L2 learners who have little experience in an L2 environment also suggest that foreign language classroom experience may not have much effect on the perception of an L2 contrast not found in the native language (e.g., Goto 1971). However, since many adults will begin their learning of L2 in a foreign language environment, such as taking a foreign language class in the university, it is important for us to understand how perception and production of L2 phonology develop in adults in a foreign language classroom.

There is evidence that suprasegmentals may be more difficult than segmentals for adults to learn, since many adults who speak L2 with only a slight accent are said to have suprasegmental deviations rather than segmental errors. The ability to imitate L2 intonation has been found to decline rapidly between ages 8 and 11 (Tahta, Wood and Loewenthal 1981b). The learning of tone may be even more difficult since tone is utilized in a totally different way from intonation. Research comparing the learning of tone and other aspects of phonology and grammar (Ioup and Tansomboon 1987) has confirmed the difficulty of tone for adult learners.

Since learning of L2 phonological perception and pronunciation is gradual, and may be even slower for suprasegmentals, it is difficult to predict if the improvement in tone perception and tone production will be

significant within a short time such as one term. However, based on studies on L2 phonological learning which show that improvement occurs more rapidly for adults at the beginning of L2 learning, we will hypothesize that the learning of Mandarin tone will improve over one term of learning.

Hypothesis (7): Subjects' scores in perception and production of Mandarin tone are significantly higher at the end than at the beginning of the term.

4.5 Order of difficulty among the Mandarin tones

Early studies on the acquisition of some English grammatical morphemes by L2 learners have shown that there is a similar developmental sequence or order of accuracy among L2 learners (e.g., Dulay and Burt 1973), and this order has been found by some to correspond to the acquisition order of children acquiring their first language (Krashen 1977). Thus the order found by comparing success/error rates in L2 performance is not random and the order can reveal important differences among the elements under study. Theories of L2 acquisition as well as relevant linguistic theories need to explain this fact.

A few studies on the learning of Mandarin tones by L2 learners have discussed the relative difficulty of the four tones (see section 3.2). However, the available data are limited and not all are in agreement. On the perception aspect, one study found Tone 2 to be the most difficult

(Kiriloff 1969) while another found Tone 3 the most difficult (W.C.J. Lin 1985). On the production aspect, one study found the four tones were equally difficult (Miracle 1989), while another found Tone 1 and Tone 4 to be more difficult than Tone 2 and Tone 3 (Shen 1989).

This study will provide us with more data and hopefully can shed some light on the issue. The results will be compared with the developmental order of tone in children acquiring Mandarin as their L1, and explanation will be sought to account for the results.

4.6 Error patterns

Previous studies have found some patterns of tone errors in English learners of Mandarin (see 3.2). For example, in the perception aspect, confusion between Tone 2 and Tone 3 has frequently been reported (W.C.J. Lin 1985, Kiriloff 1969). On the production aspect, some found tone register errors predominant (Shen 1989) while some found that both register errors and contour errors are frequent (Miracle 1989).

The classification of tone errors in this study will be based on the Mandarin tonal system (i.e., a tone is misperceived or mispronounced as another Mandarin tone or is unidentifiable), and the categorization could give us some clues as to whether a register or contour error is involved. An examination of the tone errors in this study will give us

data in different tonal contexts at two different times, and the error patterns involved may be seen more clearly.

Theories of L2 acquisition will be employed to account for the errors. Both transfer and developmental factors have been found to be important in the acquisition of L2 phonology (Hecht and Mulford 1982), and both are expected to play a role in errors of tone perception and production discovered in this study.

5. METHODS AND PROCEDURE

The data used in this study were collected from students taking Elementary Chinese 101 at Michigan State University in the Fall term, 1989. Both the tone perception test and the tone production test were conducted twice: the first time two or three weeks after the course had begun when the sound system of Mandarin had been mostly covered and some basic words and sentences had been taught, and the second time near the end of the ten-week term.

5.1 Tone Perception Test

There were three parts to the tone perception test:
(A) Identification of tone in isolated syllables,
(B) Identification of tones in isolated 2-syllable words,
(C) Identification of sentences with the correct tones in a multiple choice format. Each part will be discussed below.

(A) Identification of tone in isolated syllables

Perception of tone in isolated syllables is usually explicitly taught in a Chinese-as-a-Second-Language class. In this part of the perception test, students were asked to identify the tones of isolated syllables spoken twice by a native speaker. For example, the students would hear the syllable "*dī*". They would then have to identify the tone of this syllable and write down the corresponding tone mark (-,

/, ˇ, \), or tone number (1, 2, 3, 4). In this case, the correct answer is 1 or -. More examples can be found in a complete perception test in Appendix A.

(B) Identification of tones in isolated 2-syllable words.

The difference between parts (A) and (B) is that a tone in 2-syllable words is either preceded or followed by another tone. This tonal context may affect the perception of tone by an English speaking learner.

In this part of the test, students were asked to identify both tones in a 2-syllable word spoken by a native speaker. For example, the students would hear the word "*fāxiàn*" spoken three times. They would have to identify the tone of the first syllable "*fā*" and the second syllable "*xiàn*", and write down the corresponding tone numbers or tone marks of both syllables. In this case, the correct answer would be 14 or -\. See Appendix A for more examples.

(C) Identification of short sentences with correct tones in a multiple choice format.

In contrast with the identification of tone in isolated syllables or two syllable words, in this case subjects were not asked to identify all the tones in a sentence. Although native Mandarin speakers can easily identify the tones of all the words in a sentence, they don't necessarily listen for these tones before making an identification. It is more

likely that they process the sentence by taking in all the information available, and may identify the tones based on other levels of information, especially when a sentence is spoken at a fast tempo (see section 2.1.3). If we give learners of Mandarin a sentence that they do not understand and ask them to identify the tones, the exercise seems to be unnatural and therefore unjustifiable. We can predict that the identification of tones will be less successful simply because there are more tones to pay attention to in a sentence. In Broselow, et al. 1987, tone identification of three syllables was less successful than that of two syllables. However, if we give learners a sentence that they have learned and ask them to identify the tones, we cannot be sure that they are perceiving the tone correctly. They could be processing the sentence by taking in other (segmental phonological, syntactic, etc.) information, and be making the identification based on their memorization of the tones for each word. A native Mandarin speaker can understand sentences which differ only in tones. For example, *Tā mài shū*, "He sells books" and *Tā mǎi shū*, "He buys books" differ only in the tone of "*mai*". The perception of tone is crucial in understanding the sentences correctly. A native Mandarin speaker can also detect a tone mispronounced in a sentence and may ask for clarification if uncertainty or misunderstanding arises. A good L2 learner of Mandarin should, to some extent, be able to do the same.

Consequently, in this third part of the perception test, students were asked to choose from four versions of a Mandarin sentence spoken by a native Mandarin speaker. The four choices differed only by one tone, and only the sentence with the correct tones was the correct answer. The oral English translation of the sentence preceded the four choices. For example, the students would hear the English sentence, "Are you busy?", followed by "A. *Nǐ màng ma?* B. *Nǐ măng ma?* C. *Nǐ máng ma?* D. *Nǐ māng ma?*". All the sentences in the task were short (three words) and, in contrast to words in parts (A) and (B), had been taught in class. The English translation of each correct sentence as well as its spelling in Pinyin were written on the answer sheet, so that the students would know the meaning and the correct tones of the sentence they were going to hear even if they did not remember what they had learned in the class. For the example given above, the answer sheet would read:

Are you busy? (Nǐ máng ma?) a b c d

All they had to do was to pick the correctly spoken sentence when they heard it. In this case, the correct answer should be c, and the letter c should be circled. See more examples in Appendix A.

5.2 Tone Production Test

In the early weeks of the term, it would be impossible to elicit spontaneous speech from the students. Even simple

conversation probably would be too difficult for some. Thus, for this study, the use of imitation and reading tasks seemed to be the best choice for obtaining speech samples. This would have three advantages. First, exactly the same speech samples, elicited from the recorded and written material provided, would be obtained for all subjects. Second, the length and the grammatical or phonological difficulties of the tasks could be controlled. Third, the speech samples obtained in the first and second test would be similar and therefore comparable.

There were also three parts to the production test: (A) isolated syllables, (B) 2-syllable words, and (C) short sentences of three or four syllables. The students were asked to perform two tasks for each part: (1) to imitate a recorded model produced by a native speaker, and (2) to read from a word/sentence list without an immediate auditory model. For each part, they would first repeat the word or sentence spoken by a native speaker as they heard it on the tape, while silently scanning the word/sentence list provided to them. After they had finished repeating all the words or sentences in a particular part, they would then read the same items aloud from the list. All the words and sentences were written in Pinyin with tone marks on them. In contrast to the words in parts (A) and (B), all the short sentences had been taught in class and the sentence translations were also given, so that the students would

know what sentence they were reading even if they had forgotten what they had learned in class. It was hoped that by first repeating the words while having a chance to look at them spelled out in Pinyin, the students would be better prepared to read them aloud on their own within a limited time. A production word/sentence list can be found in Appendix A.

5.3 Test Material

The test material used in both perception and production tests were similar. Both consisted of three types of test stimuli: (A) isolated syllables, (B) 2-syllable words, and (C) short sentences of three or four syllables.

For isolated syllables and 2-syllable words, the test items were prepared based on four considerations. First, in order to reduce any possible distraction of unfamiliar consonants or vowels in the perception or production of tones, sounds that are similar in both English and Mandarin were used as much as possible, especially for the tests conducted early in the term and in the production test. Second, certain Pinyin spellings that were likely to cause difficulties or errors were not used in the production test, which required students to read words spelled in Pinyin. These Pinyin spellings include consonant initials *x* (which represents a palatal fricative [ɕ]) and *q* (which represents

an aspirated palatal affricate [tɕʻ]), and *i* after the consonant initial *z*, *c*, and *s* (where *i* represents an apical vowel, or a syllabic [z]). They tend to cause problems because they represent very different sounds in the English spelling, which students at the level of those who took the first test had not yet mastered. A few words whose spelling may also be likely to cause errors (*i* in *shi*, where *i* represents a retroflex vowel) were nontheless used because they had been taught in class. Third, the four Mandarin tones have equal occurrences in parts (A) and (B), so that the number of errors would not be affected by the unequal distribution of certain tones. Fourth, in the perception test, the 2-syllable words chosen were not familiar to the students. This was done to eliminate the possibility that the tone of the words would be known without perception. For the production test, there were simply too few 2-syllable words that had been learned by the students. While most of the 2-syllable words were new to the students, a few 2-syllable words that had been taught in the class were used as test items in the production test to reduce the difficulty of reading all unfamiliar words. The same familiar items were used in the first and the second production tests.

For the short sentences of both tests, it was decided that unfamiliar sentences would not be used since either perception or production of totally incomprehensible

sentences would be unnatural and might not represent the students' performance in a real situation. However, given the limited vocabulary and syntax that the students had learned before the first test, it was difficult to control the segmental environment or the tone in each position. Thus the criteria for choosing the sentences were: (1) familiarity to the students, and (2) sentence length of just three or four syllables. Although different sentence types (statements, wh-questions and yes-no questions) were included, and the tone to be tested in the perception test occurred in different sentence positions, the occurrences of the four tones in different sentences types or in different positions were not equal, because of the limitations already mentioned. The same sentences were used in the first and second tests, but both the order of the items and the order of the choices was different for the two tests.

In the perception test, there were 20 isolated syllables in part (A), five of each of the four tones, and 16 2-syllable words in part (B), including all the tone combinations possible in 2-syllable words. (However, because tone 3 is realized as tone 2 when followed by another tone 3, there are actually only 15 patterns). There were eight short sentences in part (C). In the production test, there were ten isolated syllables in part (A); excluding the first two as practice, there were two of each of the four tones. There were also 16 2-syllable words in part (B), including

all the tone combinations, and eight short sentences in part (C).

5.4 Subjects

All the students taking Chinese 101 in the Fall, 1989 term took the tests during their regular class hours as part of their course requirement. Most of them were undergraduate students and a few were graduate students. No information about their age was obtained. However, they were probably between 18 and 30 years of age. Based on the information regarding the students' language background (as given by the students in a questionnaire requested by the instructor), those students whose native language is not English and those students who can speak a tone language fluently were later excluded. This resulted in a total of 31 native English speaking subjects, 15 male and 16 female. Many of them had previous second language learning experience, including seven who had taken Chinese before, although their Chinese learning experience had not exceeded one year. Some students dropped the course after the first tests. There were also students who came late to class and missed part of a test, or who were absent and did not take a test. As a result, two subjects took only part of the first perception test, four did not take the second perception test, six did not take the second production test and one

took only part of it. Their scores were still included in the analyses when available.

5.5 Judges

The subjects' production was evaluated by three Chinese judges. These judges are native Mandarin speakers from Taiwan with backgrounds in linguistics and experience in teaching Chinese. One judge has a Ph.D. degree in linguistics and over seven years of Chinese teaching experience. Another judge is a Ph.D. student in TESOL, who has taken courses in linguistics (including phonetics) and who has had experience teaching Chinese to American-born Chinese children. The author is the third judge, who has five years of experience in teaching Chinese to Americans as well as to American-born Chinese children. None of the judges was teaching the Chinese 101 course when the tests were conducted and thus the judges could not be influenced by the students' performance in the class.

5.6 Procedures

The first tone perception test was conducted as part of a quiz about two weeks after the course began. The first tone production test was administered a week after that. Originally, the first production test was conducted at the same time as the first perception test. However, some of the students did not understand the instructions very well

and went on to read the words from the part (B) 2-syllable list aloud immediately after they had finished repeating the part (A) isolated syllables, thus skipping the unprompted production task of part (A). Other students who started reading the isolated syllables aloud also switched to 2-syllable words after a few words because they heard others reading different words. As a result, the recordings could not be used and the production test had to be conducted again the following week.

The second production test was conducted in the last week of the class, and the second perception test was conducted during the final examination after the ten week course ended. There were six weeks between the first and the second production test, and about seven weeks between the two perception tests.

The tones and most segments of Mandarin had been introduced and practiced before the first perception test was conducted. A tone identification test of isolated syllables and 2-syllable words was given in the class by the instructor the week before the first perception test, so that the students were familiar with the format of the first two parts of the perception test.

5.6.1 Testing

The first perception test and both production tests were conducted in the language laboratory, where Sony ER-840

Booth recorders and headphone sets were used. When a test was administered in the language laboratory, all the instructions and test materials were pre-recorded on the tape. The instructions for the production tests were also written on the students' word/sentence list. Students were first instructed to put on their headphones and put their tape in place, and to adjust the position of the microphone. A brief oral description of the test was given through the headphones before the tape was played. After the instructions were played, the tape was stopped to allow for questions before proceeding to the test.

For the perception test, the 20 isolated syllables were spoken twice each and the 16 2-syllable words were spoken three times each. The students wrote down the tone mark or tone number on their answer sheets as they heard the word. In the short sentence part, they chose the sentence with the correct tones by circling on their answer sheet the letter preceding the correct choice. Each of the four choices was spoken once, and there were eight sentences in total. The perception test took a total of about ten minutes.

The second perception test was conducted in a classroom instead of the language lab because the final examination was scheduled to be held in the classroom. The test items were read by the instructor rather than pre-recorded. However, the content and the format were otherwise similar to the first perception test.

For the production test, after the pre-recorded instructions were played, the students listened to the isolated syllables and immediately repeated each syllable. After they had finished repeating all of them, the students then read the same syllables aloud from the word list one by one after hearing the corresponding number on the tape. When they had finished this reading, they would then listen to the 2-syllable words and repeat them, and subsequently read these words aloud from the list. They then went on to the short sentences. The students' repetitions and readings were recorded automatically from the control panel at the front of the language lab.

Admittedly, the testing and recording environment of the language lab was not ideal for the production test. The students could definitely hear others during the production test and sometimes this background noise was even recorded on the tape. However, since the imitation task provided them with a good model to imitate, the students presumably would not pay much attention to the pronunciation by other classmates, whose production could well be less than adequate. Even for the reading task, there would be little motivation for a student to listen to others before reading a word unless he knew the other student was pronouncing the word correctly. As a matter of fact, most students read the words or sentences at about the same time. Some students did try to make a modification right after the first

attempt. However, they were not always successful, and very often the second attempt sounded just like the first one. It is not clear whether they were affected by other students' productions or simply were not satisfied with their own.

5.6.2 Evaluation

In order to check the validity of the perception test, three Chinese listeners, one female from Beijing and two males from Taiwan, took both the first and the second perception tests after they had been administered to the students. The recording of the second perception test was obtained by using a minirecorder in the classroom when the test was conducted. To check the validity of the production test, two native speakers, one male speaker from Beijing and one female speaker from Taiwan, also made the same recordings as the subjects did. The perception test questions were objective questions, but the production test questions needed to be evaluated. The recordings of the students' production were good enough for a listener to hear the tones clearly, even when more than one voice was recorded on the tape. The recordings were copied and edited according to task and subtest. For example, all the isolated syllables produced in the imitation task were put together. Each subject was given a number and their recordings were preceded by their number on the edited

tapes. The two native speakers' recordings were also given a number and mixed with those of the subjects. Recordings of two of the subjects were also chosen randomly to be copied twice and to be given a second number. The edited tapes were then judged by three native speakers.

The judges were instructed to judge only the tones, and were told that the correctness of the segments should not be a concern in their judgement. None of them seemed to have any problem with the isolation of the tone from the segments. When a word or a sentence was repeated or modified by the subject, the judges were asked to judge only the version with the better tone.

For each tone recorded, the judges were asked to decide: (1) whether it could be identified as one of the four Mandarin tones (phonemic judgement), and (2) whether it was well pronounced (phonetic judgement). If a tone was judged to be one of the Mandarin tones, the judges would write down the tone number or tone mark on the evaluation sheet. If it could not be identified as a Mandarin tone, they would write a question mark. If a tone was identifiable, but deviated from the norm, they would circle the tone number or tone mark. For isolated syllables, the judges were asked to write down each tone regardless of the correctness of the tone. For 2-syllable words and short sentences, the judges were asked to write down only the incorrect or deviant tones (for example, if a tone 1 was

pronounced as a tone 4, or when a tone was not well pronounced, or not identifiable), since they probably would recognize the target word or sentence when they heard the recording, even though it might not have been pronounced completely correct. By not writing down all the correct tones, they could therefore concentrate on the incorrect ones.

The recordings total about seven hours in length, so the evaluation was done at each judge's own convenience. It was often necessary for them to listen to an item more than once in order to make a judgement, so that it took much longer than seven hours to complete the work.

Ioup and Tansomboon (1987) used a similar evaluation procedure to evaluate tone production of learners of Thai. Three native Thai judges rated tones in a taped conversation and a humming test based on a three-point scale:

Level 1: phonemically and phonetically incorrect

Level 2: phonemically correct but with phonetic deviation

Level 3: indistinguishable from native speakers

The three judges concurred completely on their evaluation of the phonemic accuracy of tones, and they agreed on 93 percent of their judgements in evaluating the phonetic quality of the tones (p.340). This argues that tone evaluation by native judges is a reliable procedure.

5.7 Computing the scores

The responses of each subject to each test item, and the evaluations given by the three judges on each test item for the production test, were entered in a computer file. The scores were calculated by computer by comparing the correct answers with the responses given by the subjects.

5.7.1 Perception test

There was only one correct answer to every item in the tone perception test. The percentage of correct responses in the two perception tests was computed separately for each of the three parts--(A) isolated syllables, (B) 2-syllable words, and (C) short sentences--and the percentage of overall correct responses in the perception tests was also computed. For 2-syllable words, the percentages of correct responses for the first syllable and for the second syllable were also computed separately.

5.7.2 Production test

For the production tests, each tone produced by the subjects received one judgement from each of the three judges. Two different methods were used to compute the scores. First, the percentage of correct tone production in each of the six parts (2 tasks [imitation and reading] × 3 contexts [(A) isolated syllables, (B) 2-syllable words, and (C) short sentences]), was computed separately based on the

average score of the three judges. The percentage of correct production of tones in the first and second syllables of 2-syllable words and the percentage of correct production of tones in the first (1-syllable) word, second (1-syllable) word, and the third (1- or 2-syllable) word of a sentence were also computed separately. Each subject also had an overall correct tone production score for imitation and reading. In computing the scores, only the phonemic judgement was considered, and any phonetic deviation was ignored. Thus a tone judged to be correct but not very good would be the same as a tone judged to be correct and good. However, it was often the case that a tone judged to be not very good by one or two judges would be judged to be incorrect by another judge. Thus the average of the three judges could reflect how good the production of a subject was.

Another way of calculating the scores was to take the majority opinion (two out of three judges) in deciding the correctness of each tone production. Thus a tone would be considered Tone 2 if at least two judges judged it to be Tone 2 even though one judge judged it to be Tone 3. If all three judges disagreed with one another, then the tone would be considered unidentifiable. This way of deciding the answer is necessary to compare the error/success rates of the four tones and to find the tone error patterns for each tone. The percentage of correct production in each of the

six parts and the overall correct percentage for the two tasks (imitation and reading) in the two tests were all also computed this way.

5.8 Statistical Analyses

The statistical analyses were done with the SAS statistical computer package on the IBM 3090 mainframe computer at the Computer Center of Central Michigan University, Mount Pleasant, Michigan. For all the statistical analyses, the significance level was set at 0.95 (i.e., $p<0.05$). The production scores obtained by the two different methods (the average score of the three judges and majority opinion) were both analyzed when applicable.

Some experimental data were uncollectable: two subjects took only part of the first perception test, four did not take the second perception test, six did not take the second production test and one took only part of it. These data were considered as missing values. When analyses of variance were computed separately for each subset of the tests, the missing values in one subset did not affect the analyses of the other subsets of the tests. However, in the analyses of Correlations and Repeated Measures, because the calculation was based on paired values in two subsets, all the pairs with at least one missing value were treated as missing values.

5.8.1 Analyses of the effects of some variables on tone perception and tone production

The analysis of variance (ANOVA) was computed by using the SAS generalized linear model (GLM) procedure to analyze the effects of tonal context, task, sex and L2 experience on perception and production scores. Three variables were analyzed for the perception test: tonal context (isolated syllables, 2-syllable words, short sentences), sex (male, female), L2 experience (no L2 learning experience, previous Mandarin learning experience, other L2 learning experience). For the production test, the above three variables (tonal context, sex and L2 experience) plus the variable of task (imitation, reading) were analyzed. The significance level was set at 0.95.

Tukey's studentized range (HSD) test at the significance level of 0.95 was used with the ANOVA to compare the means of the multiple levels of each variable. Thus, if a variable was found significant, the way in which the means of the dependent variable was affected by the multiple levels of the variable could be determined.

The ANOVA was first performed separately on each subset. When the repeated measures were performed to compare the corresponding subtests in the first and the second tests, the ANOVA was again performed on each of the two corresponding subsets, but the missing values in one subset would cause the corresponding values in the other

subset also to be treated as missing values. No missing values were used in the repeated measures.

5.8.2 Analysis of correlations between the perception and production of tone in isolated syllables and in sequences

The SAS CORR procedure was used to compute the correlation coefficients (Pearson's *r*) between all the subsets of both the perception and production tests (e.g., reading of isolated syllables in the first test would be one subset). The significance level was set at 0.95. The relationship between the perception or production of tone in isolated syllables and tone in sequences (2-syllable words or short sentences) would be determined by examining the correlation coefficient obtained between them.

5.8.3 Analysis of correlations between tone perception and tone production

The relationship between the perception and production of tone in the same tonal context (e.g. 2-syllable words) would be determined by the correlation coefficient obtained between them using the SAS CORR procedure. The significance level was set at 0.95.

5.8.4 Analysis of improvement in tone perception and tone production

To determine whether improvement occurred between the first test (Time 1) and the second test (Time 2), the SAS GLM procedure with a Repeated option (Time as the repeated measure) was performed on the perception and production data sets as well as on each subset of the perception and production tests. Missing values at one time would cause the corresponding values at the other time to be treated as missing also. No missing values were used in the repeated measure analysis. The significance level for this analysis was set at 0.95.

5.8.5 Analysis of order of difficulty among the Mandarin tones

To determine whether the four Mandarin tones were equally difficult for subjects to perceive or produce, the SAS FREQ procedure with a chi-square test was used to compute the two-way cross-tabulation tables of the number of correct/incorrect responses to each of the four tones in each test subset. The Pearson chi-square values obtained by the FREQ procedure were used to determine if the correct/ incorrect responses were affected by the variation of Mandarin tones. The significance level was set at 0.95 for the chi-square test. To determine whether the pattern of correct/incorrect responses to each of the four tones in

each part remained the same in both the first and second tests, the SAS FREQ with the Cochran-Mantel-Haenszel (CMH) option was used to compute the two-way cross-tabulation tables in corresponding subsets of the two time-varied tests. The general association values obtained by the FREQ procedure were used to determine if the pattern of correct/incorrect responses to the four tones in each subset remained the same in the first and the second tests. The significance level was set at 0.95 for the general association values.

6. RESULTS AND DISCUSSION

6.1 Test validity and reliability

As controls for the perception test, three native speakers of Chinese were asked to take both the first and the second perception tests. One control's answers were completely correct. Two made one mistake each on their tests. However, when they were asked to do part of the test again (which included the item on which they had made the mistake), they changed their answers. Even with the mistakes, the percentage of correct answers of native speaker controls was over 99%.

For the production test, the judges agreed almost completely on the two native speakers' production. Their production was judged to be all correct and good except for one isolated syllable, which was judged to be not very good by one judge. The two native speakers were recognized easily by the other two judges besides the author.

The production of two subjects was randomly chosen to be judged twice. A comparison was made between these two judgments by each of the three judges. The average consistency rate for these judgments of the three judges was 95±1.66%, when only the phonemic judgment was considered (96.15±1.66% for Judge 1, 93.85±3.72% for Judge 2, and 97.11±2.03% for Judge 3). If phonetic judgment was also taken into consideration, then the three judges had an

average consistent rate of 93.46 ±2.75% (93.66±1.7% for Judge 1, 92.50±3.84% for Judge 2, and 94.23±2.84% for Judge 3).

For all subjects, the three judges agreed on 89% of their judgments on the phonemic identification: Judge 1 and Judge 2 agreed on 90.15% of their judgment, Judge 1 and Judge 3 agreed on 92.16%, and Judge 2 and Judge 3 agreed on 91.39%. Because the phonetic judgments of the production was not used in computing the scores, we did not calculate the percentage of agreement among judges on the phonetic judgments. Obviously it would be lower than the percentage of agreement on phonemic judgments only.

Subjects' scores, based on the average of the three judges, can be found in Appendix B.

6.2 Variables affecting tone perception and production

The effects of tonal context, task, sex, and L2 experience on tone perception and production were determined by the results of ANOVA, with Tukey's test revealing the difference between the various levels of each variable.

However, the results of Tukey's test are not always consistent with those of ANOVA. A variable considered significant in ANOVA (usually with p close to 0.05) sometimes turns out to be insignificant in Tukey's test.

6.2.1 Tonal context

Perception

The ANOVA procedure revealed that subjects performed significantly differently (p=0.0001) in the three tonal contexts ((A) isolated syllables, (B) 2-syllable words, (C) short sentences) in the first perception test. Tukey's test revealed that the differences were significant between all three contexts (A and B, B and C, as well as A and C). However, the differences between the three contexts in the second perception test narrowed and the results of ANOVA did not reach significance. The mean scores of tone perception in the three contexts in the two perception tests are shown below:

	A	B	C
1st perception	85.52	64.69	50.40
2nd perception	92.78	87.96	85.19

This change was also reflected in a repeated measures analysis, which showed significant Time*Context interaction effect (p=0.0001).

Production

There was a significant context effect in both the first and the second production tests. In the first test, the difference between part A (isolated syllables) and part B (2-syllable words) as well as part A (isolated syllables) and part C (short sentences) was significant. The difference between part B and part C was not significant.

In the second test, only the difference between part A and part B remained significant. The difference between part A and part C was no longer significant.

There was also a significant Context*Task interaction effect found in both the first and the second test. When the two tasks were analyzed separately, the Context factor was significant only in the reading task in both the first test ($p=0.0001$) and the second test ($p=0.0013$), but not in the imitating task. The mean scores of reading and imitation are shown below:

	A	B	C
1st reading	93.15	72.31	76.04
2nd reading	88.67	74.18	78.61
1st imitation	98.52	96.88	95.70
2nd imitation	97.17	97.17	95.06

Hypothesis 1, "Subjects perceive and produce tones with significantly higher scores in isolated single syllables than in two-syllable words and sentences," is confirmed for perception in the first test and for reading production, but not for imitation. The results indicate that L2 tone perception and reading production, just as L2 sound perception and production, are affected by different phonological contexts (Flege 1988:278, 315, see 3.1.1.1). As expected, tone in isolated syllables was perceived and produced with significantly higher scores than tone in 2-syllable words and sentences, just as L2 sounds in isolated words are perceived and produced better than in longer

utterances (Shimizu and Dansuji 1983, L. Dickerson 1975, Borden et al. 1983). However, this context effect soon disappeared for perception. At the end of the term, although the perception scores in isolated syllables are still higher than those in 2-syllable words and short sentences, the difference did not reach significance. Thus the effect of different tonal contexts (as defined in this study) on tone perception decreased in less than two months as learning progressed.

It should also be noted that while the perception of tone in 2-syllable words was much better than that in short sentences at the beginning, the reading of tone in short sentences was a little better than that in 2-syllable words. The fact that all short sentences had been taught in the class, but most words had not, may have some effect on the scores. Thus, for the production aspect, difference in familiarity with the test items seems to have affected subjects' performance. Consequently, the context effect that may normally be expected (tone in 2-syllable words is produced better than tone in sentences) did not show up. For the perception aspect, familiarity with the test items did not seem to have the same effect as in production, probably because it is necessary to produce the tone in producing a sentence but it is not necessary to perceive the tone when listening to a sentence. Another possible confounding factor is that perceiving tone in 2-syllable

words and in short sentences are different tasks (see 5.1) and the task of perceiving tone in short sentences might be more difficult.

6.2.2 Task

The variable task (imitation and reading) was analyzed for the production test. Results of ANOVA showed that the differences between imitation and reading were significant ($p<0.02$) in all three contexts (isolated syllables, 2-syllable words, and short sentences) in both the first and the second production tests. The mean scores on the two tasks are listed below:

1st test	A	B	C	Mean
imitation	98.52	96.88	95.70	97.03
reading	93.15	72.31	76.04	80.50
2nd test				
imitation	97.17	97.17	95.06	96.48
reading	88.67	74.18	78.61	80.60

Hypothesis 2, "Task, imitation or reading, affects subjects' performance in tone production: imitation results in significantly higher scores in tone production than reading," is clearly confirmed. The discrepancy between reading and imitation was greater in tone sequences (2-syllable words and short sentences) than in isolated-syllable tones.

This result is similar to that obtained by Snow and Hoefnagel-Höhle (1977, naturalistic study) from L2 learning in an L2 environment, which also showed higher scores of

pronunciation in imitated than in non-imitated speech (in response to a picture) (see 3.1.1.2). However, their data also showed a tendency for non-imitation scores to improve faster and gradually catch up with the imitation scores as learning progressed. The lack of improvement in reading scores in the present study (see 6.5.2 below) does not show the same tendency for tone learning.

Nevertheless, there were other signs that reading and imitation were getting closer. Examining the correlation between reading and imitating tone in the same context in the first production test, we found no significant correlation between the two tasks. However, in the second production test, a significant correlation was found between reading and imitating of isolated syllables ($r=0.66208$, $p=0.0003$), and a lower correlation was found between reading and imitating of short sentences ($r=0.46459$, $p=0.0222$).

Also found significant was the correlation between reading of tone in isolated syllables in the first test and the imitation of tone in isolated syllables in the second test ($r=0.61164$, $p=0.0012$), a correlation better than that between imitation in the two tests. A significant correlation was also found between the reading of short sentences in the first test and the imitation of short sentences in the second test ($r=0.60441$, $p=0.0018$).

Thus, the imitation of tone in isolated syllables and short sentences seemed to become more related to the reading of tone in corresponding contexts as learning progressed.

6.2.3 Sex

Perception

Results of ANOVA on the perception tests showed that the difference between male and female was significant (p=0.0319) in the first perception test. The T-test procedure on each of the three subtests, however, revealed that only part C (short sentences) showed a significant difference between male and female (p=0.0413). The sex difference in part A (isolated syllables) and part B (2-syllable words) did not reach significance. In the second perception test, sex difference did not reach significance in any of the three parts. The scores are listed below:

1st perception	A	B	C	Mean
male	89.00	68.33	57.50	71.62
female	81.79	61.04	43.75	61.35
2nd perception				
male	95.42	90.63	88.54	91.52
female	90.67	85.83	82.50	86.33

However, when a repeated measure analysis was done, and no observations with missing values were used, then sex was not a significant variable in either of the two perception tests. This suggests that the significant difference between the sexes was probably due to the small number of subjects which does not represent the general population of

students. Perhaps the few individuals who dropped the course later in the term or who came late and missed part of a test affected these results significantly.

Production

Based on the result of ANOVA, sex should be a significant (p=0.0319) variable in the first production test. However, Tukey's test indicated that the difference between males (mean=89.41) and females (mean=88.15) did not reach significance at the 5% level. The second production test did not show a significant sex difference.

When each part was analyzed separately, only part (A) (isolated syllables) of the first production test showed a significant sex difference (p=0.0416) according to ANOVA. Again, the difference was significant only in ANOVA, Tukey's test showing the difference between males (mean=96.39) and females (mean=95.31) to be insignificant at the 5% level. The scores are shown below:

1st production	A	B	C	Mean
Male	96.39	85.90	85.96	89.41
Female	95.31	83.37	85.79	88.15
2nd Production				
Male	94.27	86.55	87.18	89.20
Female	91.67	85.34	86.42	88.06

When each part was analyzed according to Task (imitation or reading), we found that the imitation task of part A (as well as part B and part C) did not show a significant sex difference. Only the reading task, in part

A of the first production test, showed a significant (p=0.0490) sex difference. Again, the difference between males (mean=94.44) and females (mean=91.93) did not reach significance in Tukey's test. The scores on the reading task are shown below:

1st reading	A	B	C
Male	94.44	74.72	77.07
Female	91.93	70.05	75.08
2nd reading			
Male	89.24	74.43	77.46
Female	88.14	73.96	79.59

When male and female production data were analyzed separately, for the first test, the same significant differences were found in Task (imitation or reading) and Context (isolated syllables, 2-syllable words, or short sentences), as well as the Task*Context interaction (see Task and Context above). However, there were also significant differences found only in the female data in L2 Experience (p=0.0034) and L2*Task interaction (p=0.0176), but not in the male data. For the second test, only Task remained a significant factor for both male and female production. A Time*Context interaction was also found for both male and female data.

Hypothesis 3, "Subjects' performance in Mandarin tone perception or production is not affected by their sex," cannot be clearly confirmed or rejected. The difference between sexes was greater in the first test than in the second test. While the difference in the second test was

insignificant, the significance of sex difference in the first test cannot be definitely determined because the results from two statistical procedures were inconsistent. The seemingly better performance in males than females in tone learning is opposite to the findings of some other studies on L2 pronunciation, which showed that females had better pronunciation than males (Asher and Garcia 1969, Thompson 1984). However, in the second test, females seemed to perform a little better than males in reading short sentences, which had been taught in class and were most familiar to subjects (see 5.3), although the difference was not significant. This might suggest a greater improvement for females than males in reading test items they were most familiar with. More research is needed to determine whether sex has an effect on tone perception and production by L2 learners.

6.2.4 L2 experience

Perception

ANOVA on the perception test indicated that L2 experience did not make a significant difference in either the first or the second perception tests. Subjects with Mandarin learning experience or with other L2 learning experience performed better than subjects with no second language learning experience; however, the difference was not significant.

Production

L2 is not a significant factor in the subjects' tone production performance in either of the two production tests. When ANOVA was carried out for different contexts and tasks, the same insignificant results were found, with only one exception. A significant result (p=0.0013) was found in the imitating of short sentences in the second production test. Tukey's test showed that the significant difference occurred between scores of subjects with Mandarin learning experience (mean=89.00) and scores of subjects with no L2 learning experience (mean=96.67) as well as scores of subjects with other L2 experience (mean=96.17). Subjects with previous Mandarin learning experience scored significantly lower than the other two groups.

L2 experience	No L2	Other L2	Mandarin
	96.67 (N=4)	96.17 (N=16)	89.00 (N=4)

L2 experience and L2*Task interaction were also found to be significant in the female data in the first test, but not in the male data. However, the number of subjects is probably too small to make a valid generalization.

Hypothesis (4A), "Subjects with prior Mandarin learning experience score significantly higher in Mandarin tone perception and production than those without such experience," was not confirmed. There were no significant differences between subjects with and without prior Mandarin learning experience in any of the perception and reading

production tests. The only significant difference found between the two groups was in imitating sentences, and the scores were contrary to expectation: subjects with prior Mandarin learning experience scored significantly lower than subjects without such experience. Remember that imitation scores were not much related to perception and reading scores at the beginning, but the scores on imitating single syllables and short sentences became closer to counter perception and reading scores in the second test (see 6.2.2).

It appears that two possible factors may account for these lower scores of subjects with prior Mandarin learning experience. First, when subjects were very familiar with the test items, they may already have had a "prototype" in their mind and they paid less attention to the auditory model they were imitating. Second, they may already have formed a habit of producing tones in their own way and it became harder to modify this. In fact, it has been observed that L2 learners with prior learning experience with non-native speakers in a foreign language environment often have worse pronunciation than those who do not have such experience (Seliger et al. 1975:22). The lower scores of subjects with prior Mandarin learning experience may be considered consistent with the trend of closer relationship between imitation and non-imitation in utterances they had been taught. It has been found that greater variation in

production is characteristic of an active learning process, and the variation tends to be reduced resulting in an increased amount of within-subject stability (Gass 1984).

The insignificance of Mandarin learning experience on perception seemed to contradict the result concerning Time effect found in the perception tests. Among the seven subjects who had had Mandarin learning experience, their experience ranged from less than one quarter to one year of high school, and may have included classes like "Chinese culture and language", so it was not clear what they had learned before coming to the Chinese 101 class in Fall 1989. The great variation found in tone perception and the small number of subjects in this group may also account for the insignificant result.

Hypothesis (4B): "For subjects without Mandarin learning experience, other L2 learning experience does not affect performance in tone perception or production," was supported by the results obtained. This finding is consistent with the findings of other studies on L2 pronunciation, which found no significant effect of other L2 experience (Suter 1976, Thompson 1984). However, since L2 learners have been reported to rely on prior-learned L2 in learning additional related languages (Rivers 1979, Singleton and Little 1984), it is possible that prior L2 phonological learning will have some effect on learning additional similar L2 phonological structure. Thus the

learning of non-tone languages may not have much effect on the learning of a tone language, the learning of another tone language may have greater effect on the learning of additional tone languages. However, as in many aspects of L2 learning, similarity may result in transfer, both negative and positive, and because of the interaction of tone and intonation, it may not be easy to sort out the effect of prior L2 learning experience.

6.3 Relationship between the perception and production of tone in isolated syllables and in sequences.

6.3.1 Perception

In the first perception test, the perception of tone in isolated syllables correlated with the perception of tone in 2-syllable words ($r=0.77826$, $p=0.0001$), especially with perception of tone in the second syllable ($r=0.84243$, $p=0.0001$). The correlation with the first syllable was also highly significant but not as high ($r=0.52571$, $p=0.0001$). The correlation between the perception of tone in isolated syllables and tone in short sentences was also significant, but even lower ($r=0.47642$, $p=0.0090$). There was also a significant correlation between perception of tone in 2-syllable words and short sentences ($r=0.56774$, $p=0.0011$); this correlation was higher than that between tone in isolated syllables and tone in short sentences.

In the second perception test, the perception of tone in isolated syllables correlated with the perception of tone in 2-syllable words (r=0.60031, p=0.0009). However, unlike in the first test, the correlations with the first syllable (r=0.57955, p=0.0015) and with the second syllable (r=0.55999, p=0.0024) were not much different. The correlation between the perception of tone in isolated syllables and the perception of tone in short sentences was no longer significant. But the correlation between the perception of tone in 2-syllable words and short sentences continued to be significant (r=0.52464, p=0.0050).

In addition, the perception of tone in isolated syllables in the first test correlated with the perception of tone in sequences in the second test. The correlation was highly significant with the perception of tone in 2-syllable words (r=0.79424, p=0.0001) and significant with the perception of tone in short sentences (r=0.47437, p=0.0166) in the second test.

The perception of tone in 2-syllable words in the first test also correlates with the perception of tone in short sentences in the second test (r=0.58891, p=0.0016). The perception of tone in short sentences in the first test correlated with the perception of tone in 2-syllable words in the second test (r=0.42983, p=0.0252).

The perception of tone in isolated syllables in the first and the second tests correlated significantly

(r=0.60779, p=0.0013); so did the perception of tone in 2-syllable words (r=0.68293, p=0.0001). But the perception of tone in short sentences did not correlate significantly between the two tests.

The results indicate that the perception of tone in isolated syllables at the beginning of the acquisition process correlates with the perception of tone in sequences at the beginning as well as at a later time. The correlation did not change over the seven weeks of learning, despite improvement in scores in tone perception of 2-syllable words and short sentences. On the other hand, improvement in the perception of tone in isolated syllables results in weakening (2-syllable words) or insignificance (short sentences) of the correlation of this with the perception of tone in sequences in the second test.

The first part of Hypothesis 5, "There is positive correlation between perception of tone in isolated single syllables and syllable sequences", was generally confirmed. The perception of tone in isolated syllables was definitely related to the perception of tone in 2-syllable words. It was also correlated with the perception of tone in sentences, but only at the very beginning of the acquisition process, and the correlation was not high. On the other hand, the correlation between tone perception in 2-syllable words and short sentences was stronger than that between tone perception in isolated syllables and in sentences,

suggesting that the sequencing of tone, whether it was in 2-syllable words or in sentences, was the common factor. The correlation between the perception of tone in 2-syllable words at the beginning of learning and the perception of tone in short sentences did not seem to change after seven weeks of learning.

The assumption that perception of tone in isolated syllables is a prerequisite to the perception of tone in sequences seems to be supported by the significant correlation found between them. However, the results also show that the correlation became lower when the tone sequences became increasingly longer and structurally more complicated. It should be remembered that only short sequences of tone, which supposedly resembles tone in citation form, was tested in this study. The insignificance of correlation between tone perception in isolated syllables and short sentences in the second test may be partially due to the high scores on tone perception in isolated syllables, but it is also possible that the teaching of these short sentences in the class has somehow affected the correlation. Thus after L2 learners had been taught and had been actively learning certain constructions of L2 for some time, their performance on these constructions may have been strongly affected by variables other than their perception of the citation form.

6.3.2 Production

Since the results of the two production tasks, reading and imitating, are quite different, they will be discussed separately.

6.3.2.1 Reading task

The reading of tone in isolated syllables did not correlate significantly with the reading of tone in 2-syllable words or short sentences in the first test or in the second test. On the other hand, the reading of tone in 2-syllable words and in short sentences were significantly correlated in the first test ($r=0.76239$, $p=0.0001$) and still significant though lower in the second test ($r=0.47544$, $p=0.0189$).

The reading of tone in 2-syllable words in the first production test also correlated with the reading of tone in short sentences in the second test ($r=0.68713$, $p=0.0002$) and with the reading of tone in isolated syllables in the second test ($r=0.40045$, $p=0.0473$). Note that scores for the reading of tone in isolated syllables decreased in the second test.

The reading of tone in short sentences in the first test correlated with the reading of tone in 2-syllable words in the second test ($r=0.78581$, $p=0.0001$).

Between the two tests, the reading of tone in parallel contexts were always correlated. The reading of isolated syllables was correlated ($r=0.62552$, $p=0.0008$) as was that

of the 2-syllable words (r=0.83865, p=0.0001) and short sentences (r=0.77655, p=0.0001).

The second part of Hypothesis 5, "There is positive correlation between production of tone in isolated syllables and syllable sequences", was not confirmed for reading. Production of tone in isolated syllables did not correlate significantly with production of tone in sequences. Thus, the ability to produce Mandarin tone in isolated syllables was not related to the ability to produce Mandarin tone in sequences. As mentioned in 4.1.1, L1 transfer may have a different effect on the production of tone in isolated syllables, 2-syllable words, and sentences. The results suggest that L1 transfer, or some other factors that may affect tone production, did not affect all L2 learners in the three tonal contexts in the same way. However, the production of tone in 2-syllable words and short sentences were very much related at the beginning. The relation became less close at the end of the term. Subjects' increased familiarity with the test items in short sentences through extended practice during the term (see 5.3) may have affected the relation between producing tone in 2-syllable words and short sentences.

6.3.2.2 Imitation task

The imitation task showed no significant correlation between any contexts in either the first or the second test.

Between the two tests, the imitation of tone in parallel contexts was not significantly correlated except for isolated syllables (r=0.57763, p=0.0025).

Hypothesis 5 was not confirmed for imitation, either. Imitation seemed to have reached a near peak of performance in all three parts of the test from the beginning. (Scores do not improve, presumably because they were so high on the first test that there was little if any room for improvement. Such absence of improvement is a so-called *ceiling effect*). This may explain the lack of significant correlation of imitation with reading and perception, and the lack of correlation between the two tests or different parts of the same test.

6.4 Relationship between tone perception and tone production

Relationship between tone perception and tone production will be discussed according to the three tonal contexts.

6.4.1 Tone in isolated syllables

For tone in isolated syllables, a significant correlation was found between perception and the reading task in the first test (r=0.63099, p=0.0002). The correlation between perception and the imitation task was not significant. Perception of tone in isolated syllables in the second test correlated with both reading (r=0.61137,

p=0.0012) and imitation (r=0.72245, p=0.0001) of tone in isolated syllables.

There were also significant correlations found in the perception and production of isolated syllables across the two tests. A significant correlation was found between the perception of isolated syllables in the first test and the production of isolated syllables in the second test for reading (r=0.72925, p=0.0001), and for imitation (r=0.78515, p=0.0001). The perception of isolated syllables in the second test correlated significantly with the production of isolated syllables in the first test for reading (r=0.56153, p=0.0023) and imitation (r=0.48150, p=0.0110).

The perception and production of tone in isolated syllables were positively correlated (except for imitation in the beginning). This result is similar to previous findings on the correlation between perception and production of L2 sounds (Goto 1971, Snow and Hoefnagel-Höhle 1979, Borden et al. 1983, Schneiderman et al. 1988).

It is interesting to note that the correlation between the first perception test and the second production test seems higher than that between the first perception test and the first production test. There may be a delayed effect between the development of perception and production of tone, and a similar effect has been proposed between input frequency and accuracy order relationship on other aspects

(morpheme, syntactic structure) of L2 learning (e.g., Skehan 1982, cited in Larsen-Freeman and Long 1991:134).

6.4.2 Tone in 2-syllable words

The correlation between perception and reading of 2-syllable words was significant in the first test ($r=0.58289$, $p=0.0007$) and in the second test ($r=0.69208$, $p=0.0002$). Imitation did not correlate significantly with perception in either test.

There were also significant correlations found between perception and production across the two tests. The first perception test correlated with the second reading test ($r=0.64227$, $p=0.0010$), and the second perception test also correlated with the first reading test ($r=0.67733$, $p=0.0001$). Again, the correlation between the first perception test and the second production test seems higher than that between the first perception test and the first production test, suggesting a delayed effect, though the difference was not as great as in isolated syllables.

6.4.3 Tone in short sentences

The correlation between perception and reading of tone in short sentences was significant ($r=0.44455$,$p=0.0122$) in the first test but not significant in the second test. The imitation task did not show significant correlation with the perception in either test.

Tone perception in short sentences in the second perception test, however, showed significant correlation with tone perception in short sentences in the first reading test (r=0.47454, p=0.0124).

The percentage of overall correct tone perception and tone production in the reading task and the imitation task was also computed. The correlation between the perception and reading of tone was significant for the first test (r=0.62424, p=0.0003) and the second test (r=0.69484, p=0.0002). No significant correlation was found between perception and imitation in either of the two tests. The first perception test also correlated with reading in the second test (r=0.72090, p=0.0002) and imitation in the second test (r=0.52237, p=0.0126). The second perception test correlated with reading in the first test (r=0.63795, p=0.0003).

There was a significant correlation between the first and the second perception tests (r=0.77090, p=0.0001) and between reading in the first and the second tests (r=0.95260, p=0.0001). But imitation in the two tests was not significantly correlated. Reading and imitation tasks were not significantly correlated in the first test and significantly but not highly correlated in the second test (r=0.43977, p=0.0315).

Hypothesis 6, "There is positive correlation between subjects' Mandarin tone perception scores and production

scores," was generally confirmed between tone perception and tone reading but not for imitation. The significant correlation between perception and reading scores indicates that the two aspects of tone learning are very much related, a finding consistent with the findings on the perception and production of L2 phonology in general (see 3.1.1). The significant correlations often found between perception and reading across the two tests, sometimes higher than those found in tests conducted at about the same time, also suggest that there were interactions between them which occurred over time.

The only exception was found between perception and reading of short sentences in the second test. The correlation concerning short sentences was weak at the beginning and became insignificant after one term of learning. Remember that all short sentences had been taught in the class. The lack of correlation in short sentences at the end of the term reminds us of the results reported by Schneiderman et al. (1988), who found significant correlation between discrimination and production of French phones in learners of French before they underwent discrimination training, but no significant correlation after the training. Since we have very little understanding of how L2 phonological perception and production interact in L2 learners, and how they are affected by classroom instruction, we cannot explain how classroom instruction may

have affected the relation between perception and production of L2 phonology. However, the results obtained here suggest that instruction has some effect on the relationship between L2 phonological perception and production.

6.5 Improvement over one term of learning

6.5.1 Perception

A repeated measure analysis for perception showed a significant time effect between the first perception test and the second perception test ($p=0.0001$). There was also a significant Time*Context interaction ($p=0.0001$) as mentioned in 6.2.1. When a repeated measure was done for each subtest separately, the same significant time effect was found in all three parts ($p<0.05$). The scores of the two tests are shown below:

	A	B	C	Mean
1st test	86.40	65.38	49.07	66.48
2nd test	92.60	87.74	85.19	88.41

The perception aspect of Hypothesis 7, "Subjects' scores in the perception of Mandarin tone are significantly higher at the end than at the beginning of the term", is confirmed. This indicates that English-speaking learners' perception of Mandarin tones improved in about seven weeks (five 50-minute classes per week) of Mandarin learning. The significant time effect found in all three parts of the perception test means that improvement of tone perception

occurred generally in isolated syllables, 2-syllable words, and short sentences.

L2 learners' L2 phonological perception has been found to improve with their L2 experience (MacKain et al. 1981) and some of them can achieve native-like perceptual ability (Neufeld 1980, see 3.1.4.1). However, the effect of short-term perceptual training seems limited (e.g., Pimsleur 1963). For studies on the perception of Mandarin tones by English-speaking learners, Kiriloff (1969) found no significant improvement for tone in isolated syllables between two tests, the first conducted after two and a half months of Mandarin learning and the second after four months of learning. In the present study, the second perception test was conducted ten weeks (about two and a half months) after (most) subjects started learning Mandarin. It is possible that significant improvement can only occur before two and a half months of learning, and that after two and a half months (at least till four months) no significant improvement may be made.

W.J.C. Lin (1985) found that for a syllable embedded in a sentence, beginners (150 contact hours) made more errors in tone identification than intermediate (600 contact hours) students, who in turn made more errors than advanced (1050 contact hours) students. Most of the subjects in the present study had had only about 50 hours of instruction when the second perception test was conducted. They seemed

to have made great progress in their perception of tone in short sentences in a relatively short time. If the test scores reflect the perceptual ability for tone of L2 learners, English-speaking learners seemed to make greater improvement in short sentences than in 2-syllable words, and they made greater improvement in 2-syllable words than in isolated single syllables. Of course, there was much more room for improvement in short sentences than in 2-syllable words, and in 2-syllable words than in isolated single syllables.

However, we should remember that the two perception tests were conducted under different situations: the first in a language lab using an audio-tape pre-recorded by a speaker whose voice the subjects were not familiar with, and the second in a classroom read by an instructor whose voice the subjects were familiar with. It is possible that some other variables might also be responsible for the significantly higher scores in the second test. If the perception aspect of hypothesis 7 is to be tested again in future studies, the testing situations should be made as similar as possible.

6.5.2 Production

A repeated measure analysis showed that the difference between the first and the second production tests was not significant. When the scores of the two tasks, reading and

imitation, were analyzed separately, there was still no significant difference in either the imitation task or the reading task. The repeated measure analysis was also carried out on each of the subtests [(A) isolated syllables, (B) 2-syllable words, (C) short sentences] separately, and none of the parts showed significant time effect. The scores are listed below:

Reading Task	A	B	C	Mean
1st test	93.15	72.31	76.04	80.50
2nd test	88.67	74.18	78.61	80.60
Imitation task				
1st test	98.52	96.88	95.70	97.03
2nd test	97.17	97.17	95.06	96.48

The production aspect of Hypothesis 7, "Subjects' scores in production of Mandarin tone are significantly higher at the end than at the beginning of the term," is not confirmed for either reading or imitation. This means no improvement in tone production was made between the beginning and the end of the term. This result in imitation can be attributed to the ceiling effect reached from the beginning of the term.

It is widely observed that adult L2 learners often cease to improve well before achieving native-like pronunciation, and a lack of improvement in L2 pronunciation has been reported in many aspects of L2 phonology (e.g., VOT in stops and vowel quality, see Flege and Hillenbrand 1984). Studies have also found that the amount of formal classroom training is not a significant variable predicting L2

pronunciation (e.g., Purcell and Suter 1980). However, the lack of significant improvement in the production of tone at the beginning (the third to the tenth week) of Mandarin learning may still be a little surprising. After all, adult L2 learners have been reported to be able to produce phonemic distinctions better with increased L2 experience (Beebe 1987), including tone (Ioup and Tansomboon 1987), and there is some evidence that adult L2 learners' pronunciation improves the most at the beginning of their L2 learning (e.g., Snow and Hoefnagel-Höhle 1979).

If the results in this study are valid, then one possible explanation is that most of the learning had taken place before the first test (the third week of the class). It is also possible that there was improvement which could not be measured easily by impressionistic judgment of the correctness of a tone. For instance, there may be a trend from more variation to less variation (Gass 1984:70), or from more phonetic deviation to less phonetic deviation.

The results suggest that improvement in tone learning occurs more rapidly in the perception aspect. It seems that at the beginning of Mandarin learning, tone production is ahead of perception (even if we disregard imitation scores here, which did not appear to be related much to learning). Students could pretty much read an isolated syllable with correct identifiable tone but made more errors in identifying it. This finding is similar to that concerning

the production and perception of the /r/-/l/ contrast by Japanese speakers (Goto, 1971; Sheldon and Strange, 1982). Just as pronunciation training may have helped adult L2 learners producing sound contrasts which they could not perceive reliably, it is possible that the teaching of tone in isolated syllables, including specific description of the tonal contour of the four tones, may have helped many of the learners to produce tones better before they were able to perceive the distinction reliably.

While producing acceptable Mandarin tones in isolated single syllables did not seem to be too difficult to achieve for most adult English speaking learners, tone production in 2-syllable words and short sentences was much more difficult; however, reading was still better than perception at the beginning of the term. At the end of the term, perception of tone in 2-syllable words and short sentences improved significantly while the reading scores had not changed much. As a result, the perception of tone was better than reading at the end of the term.

6.6 Order of difficulty of the four tones

Recall the nature of the four tones in Mandarin:

Tone 1 high-level (˥)

Tone 2 high-rising (˧˥)

Tone 3 low-dipping (˨˩˦) in final position

low-falling (˨˩) in non-final position

and not followed by another Tone 3

high-rising (↑) when followed by Tone 3

Tone 4 high-falling (↓)

Comparison of the error/success rates of the perception or production of the four Mandarin tones in a given test often reveals an order of difficulty of the four tones. The success rates of the perception of the four tones in the perception tests were apparent, since there was only one correct answer to each test item. Production, however, was subjectively evaluated by the three judges and disagreement was not uncommon. Therefore, the success rate of tone production was based on majority opinion (two out of three judges). A tone produced by a subject was identified by judges as one of the four tones, or a neutral tone, or unidentifiable. When at least two judges agreed on the tone category (phonetic deviation was ignored), that would be the final decision. Thus the correctness of a tone produced by a subject or how it was mispronounced (e.g. Tone 1 mispronounced as Tone 2) would be decided this way.

The percentage of correct identification and production for the four tones in different contexts will be presented below. The number of successes and errors of the four tones for each task in each context was analyzed with a chi-square test to see if there was any significant difference among the four tones. The results were significant ($p<0.05$) for most parts of the tests, indicating that the four tones had

significantly different success rates for perception or production. When the success rates of the perception or production of the four tones were not significantly different, this will be made clear.

6.6.1 Perception

6.6.1.1 Tone in isolated syllables

The percentage of correct answers in the perception of tone in isolated syllables is shown below:

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st test	90%	72%	96%	83%	86%
2nd test	95%	90%	92%	95%	93%

In the first perception test, in isolated syllables Tone 2 was the least successfully identified tone, while Tone 3 was the most successfully identified. In the second perception test, all four tones in isolated syllables were correctly identified over 90% of the time. A chi-square test showed that the difference in success rates of the perception of the four tones was not significant.

Kiriloff (1969) also studied students' tone perception in isolated syllables (see 2.3). He found that "errors involving the 2nd tone are considerably higher and those with the 1st and 4th tones much lower than expected" and thus "[t]he considerable number of incorrect identifications of the second tone and a marked tendency to confuse with it the third tone indicate that the second tone presents the greatest difficulty to students of Mandarin" (p.65).

In this study, in agreement with Kiriloff's finding, Tone 2 was also the least successfully identified tone in isolated syllables in the first test, and continued to be in the second test, though the difference between errors on Tone 2 and on other tones was insignificant. Error rates of the perception of Tone 1 and Tone 4 were not always much lower. Tone 3 was correctly identified at a higher rate than Tone 1 and Tone 4 in the first test. However, judging from the lower identification rate for Tone 3 in the second test and the error patterns of Tone 2 and Tone 3 in the two tests (see 6.7.1), it seems that the high identification rate of Tone 3 in the first test was due to the tendency of subjects to identify both Tone 2 and Tone 3 as Tone 3. As they learned to adjust the boundary between Tone 2 and Tone 3, Tone 2 was more correctly identified but Tone 3 was often misidentified as Tone 2. Nonetheless, error rates for the perception of the four tones in the second test were not significantly different.

6.6.1.2 Tone in 2-syllable words

The percentages of correct answers in the perception of tone in 2-syllable words, including the first syllable and the second syllable as well as the overall scores, are listed below:

1st test

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st syll	68%	54%	46%	73%	61%
2nd syll	67%	51%	86%	72%	69%
Overall	67%	53%	69%	73%	65%

2nd test

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st syll	93%	88%	86%	82%	88%
2nd syll	93%	84%	94%	82%	88%
Overall	93%	86%	91%	82%	88%

Chi-squire shows that in the first perception test, the four tones were identified at significantly different rates in both the first and the second syllable of 2-syllable words. Tone 3 was more difficult to identify than the other tones in the first syllable but much easier to identify than the other tones in the second syllable. Tones 1, 2, and 4 seemed to be correctly identified at a similar rate in the first syllable and the second syllable. There are at least two possible reasons for the different success rates in identifying Tone 3 in different positions. The first is the allophonic variation of Tone 3 in (emphasized) final and non-final positions (see 2.1.2). Since the variation that occurred in non-final position was different from that in isolation, while the variation which occurred in final position was the same as that in isolation, it is natural that the tone which occurred in final position was identified better than in non-final position. Another reason for the better identification rate of Tone 3 in the second syllable (final position) was probably due to the same reason as in single syllables: both Tone 2 and Tone 3

tended to be identified as Tone 3. Overall, as when tone was identified in isolated syllables, Tone 2 had the lowest identification rate, and Tone 4 had the highest identification rate.

In the second perception test, the chi-square test shows that the successful identification rate for the four tones was not significantly different in the first syllable of 2-syllable words but was significantly different in the second syllable, in which Tone 1 and Tone 3 were identified more correctly than Tone 2 and Tone 4. However, Tone 3 still had the biggest discrepancy of successful identification rate between the first and the second syllable, though the discrepancy was much smaller than that in the first test. Tones 1, 2, and 4 in the first syllable and the second syllable still seemed to be correctly identified at similar rates. The overall identification rates of all four tones in the second test were much better than those in the first test. But now Tone 1 seemed to be the most successfully identified tone and Tone 4 was the least successfully identified, the result of much more improvement in the other tones than in Tone 4.

6.6.1.3 Tone in short sentences

The percentages of correct perception of tone in sentences are shown below:

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st test	24%	50%	68%	52%	50%
2nd test	91%	72%	90%	85%	85%

The Chi-square test shows that the four tones had significantly different success rates on both the first and the second perception test. In the first perception test, the success rates for tone perception in short sentences were not high for all four tones, and the success rate of Tone 1 perception was not higher than chance. In the second test, however, success rates of identification for all four tones greatly improved. Tone 2 was the least successfully identified tone, and Tone 1 became the most successfully identified.

W.C.J. Lin (1985) tested tone perception of his students with two sentence frames: *Wo shuo de shi* __. "What I said is __" and *Tade* __ *hen youming* "His/her __ is quite famous". Only the syllable *ma* with four different tones was tested (see 3.2.1). His test was somewhat similar to the test of tone perception in short sentences in this study, although the test in this study involved recognizing the sentence with the correct tones rather than identifying the tone of a syllable in a sentence. However, because of the limitations mentioned in 5.3, in this study the four tones did not always occur in the same position or with the same syllable.

According to Lin's results Tone 3 was the most problematic tone for his students to identify in both

sentence frames, and he suggested a new treatment of it. However, even though the treatment seemed to be effective and the students' performance improved, Tone 3 was still the most difficult to identify. In the present study Tone 3 was not the least successfully identified tone in short sentences, though it was the least successfully identified tone as the first syllable of 2-syllable words. The difference of the test methods employed in the two studies mentioned above may be part of the reason for this difference.

In the study of Broselow, et al. (1987), Tone 4 was the tone most correctly identified in isolated syllables by American English speakers. However, Tone 4 was identified at a significantly lower rate in non-final position than in final position. This difference cannot be explained by allophonic variation as can be that for Tone 3, and as was argued as evidence for perceptual transfer of American English intonation (see 3.2.1). Broselow et al.'s subjects had no Mandarin Chinese learning experience and lacked knowledge of the phonological characteristics of the four tones (e.g., falling, level, etc.) other than the examples they were given before the test. Their results may be generally valid for American listeners with no Mandarin learning experience, but presumably after learning the Mandarin tonal system and being exposed to more examples, the learners' perception would change. Tone 4 was not the

most correctly identified tone in isolated syllables in our first perception test, less than three weeks after the class began, and the positional difference was not found.

There may be more complicated factors involved in the perception of tone. The position of the tone and the tonal context can certainly affect perception. Other factors include syllable structure and degree of familiarity with the vocabulary. For example, examining the original data, I found a particularly high error rate for tones in words with the same syllable structure as words the students had learned right from the beginning. The syllable *hao2* was perceived to be Tone 3 (*hao3* "good") and the syllables *ni2* and *ni4* were perceived to be Tone 3 (*ni3* "you") at a much higher rate than other Tone 2 and Tone 4 words. The fact that tests were conducted in the class may have lead the students to expect that the words they have learned would be more likely to occur in the test.

The actual variation of tone production the students were exposed to may also have some effect. They may have developed some criteria that were more effective in identifying tones produced by speakers whose voices they were familiar with. Since the students were not familiar with the tape-recorded voice in the first perception test, but were familiar with the voice of the instructor who read the test items in the second perception test, their perception could have been influenced by this fact.

6.6.2 Reading

In the production of tone, the two tasks of reading and imitation produced quite different results and results on the reading task were more closely related to those of perception (see 6.2.2 and 6.4).

6.6.2.1 Tone in isolated syllables

The percentage of correct production of tone in isolated syllables in the reading task is shown below:

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st test	97%	92%	94%	97%	95%
2nd test	92%	68%	96%	96%	88%

In the first production test, the rates of correct tone reading were not significantly different for the four tones, and all four tones were correctly produced over 90% of the time. However, the second test showed a dramatic change for Tone 2; only 68% of Tone 2 produced by subjects was judged to be correct. The other three tones did not show such change from the first test.

6.6.2.2 Tone in 2-syllable words

The percentage of correct tone production in 2-syllable words, both the first syllable and the second syllable as well as the overall scores, is shown below:

1st test

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st syll	73%	65%	52%	81%	69%
2nd syll	73%	83%	86%	64%	76%
Overall	73%	73%	71%	73%	73%

2nd test

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st syll	91%	77%	48%	69%	72%
2nd syll	75%	75%	94%	60%	76%
Overall	83%	76%	71%	65%	74%

In the first production test, the overall success rates for reading the four tones in 2-syllable words did not differ significantly. However, the production of the four tones had significantly different success rates in the first syllable as well as in the second syllable. And except for Tone 1, every tone had quite different success rates of production in the two syllables. In the first syllable Tone 4 was the most successfully produced by subjects, and Tone 3 was the least successfully produced. The second syllable showed exactly the opposite picture. Tone 3 was the most successfully produced while Tone 4 was the least successfully produced.

In the second production test, the reading of the four tones had significantly different overall success rates. Tone 1 was the most successfully produced tone, and Tone 4 was the least successfully produced. The success rates of the production of the four tones in the first and the second syllable were different too: Tone 1 was the most successful in the first syllable and Tone 3 the least successful. Tone 3 was the most successful in the second syllable and Tone 4 the least successful.

One similarity between the two production tests was that Tone 3 was the least successfully produced tone in the

first syllable but the most successfully produced tone in the second syllable. Another is that Tone 4 was more successfully produced in the first syllable than in the second syllable, where it was the least successfully produced among the four tones. This appears to be a transfer from the English word stress pattern. The similarity of unstressed syllables in English words and Tone 3 in Mandarin has been found in the perceptions of Mandarin speakers (Cheng, 1968). Cheng found that English words with an initial unstressed syllable would cause tone sandhi in the preceding Tone 3 (e.g., *Hao professor bu duo* "There are not many good professors"), just as a Tone 3 in Mandarin would, while a word starting with a stressed syllable would not. A 2-syllable English word usually has the stress on the first syllable, and the second syllable is unstressed. The stressed syllable usually is high in pitch, and the unstressed syllable low in pitch. It appears that this high-low pitch pattern was carried over to the 2-syllable Mandarin words: Tone 3 (a low tone) was produced correctly only about half of the time in the first syllable, but much more frequently in the second syllable. The tone that was most successfully produced in the first syllable always starts high (Tone 4, high falling; Tone 1, high level). The high falling Tone 4 was the most difficult to produce in the second syllable. Most Tone 4 errors in the second syllable were judged unidentifiable in the first test and as Tone 3

in the second test, indicating that Tone 4 was probably low in pitch, a transfer of the English stress pattern.

Results in Miracle's (1989) study on 1- and 2-syllable words seems similar to that in the reading of 1- and 2-syllable words in this study, although in his study the words were those of the topic (sentence initial) position rather than words produced in isolation. In addition, Miracle relies in his analysis on acoustic information while this study relies on the auditory judgments of 3 Chinese judges.

Miracle found that the errors made by his subjects were evenly divided among the five tones (including the neutral tone). The same was found in this study in isolated syllables and 2-syllable words in the first test. However, this does not apply to the second test or to the first word in a sentence (see 6.6.2.3 below). He also found that in 2-syllable words students made significantly more errors in initial syllables (49.6% error rate) than in final syllables (32% error rate). This was also true in this study though the difference was much smaller: 31% vs. 24% error rate in the first test and 28% vs. 24% in the second test.

6.6.2.3 Tone in short sentences

The percentages of correct tone production in reading short sentences are shown below:

1st test

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st syll	83%	45%	61%	85%	76%
2nd syll		50%	89%	75%	72%
3rd word	77%	89%	73%	61%	77%
Overall	81%	65%	77%	76%	75%

2nd test

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st syll	93%	67%	67%	88%	85%
2nd syll		56%	73%	84%	75%
3rd word	65%	88%	83%	92%	78%
Overall	83%	71%	76%	84%	79%

The frequency of occurrence for each tone in short sentences was not controlled for position as in 2- syllable words. Tone 1 did not occur in the second syllable at all. However, exactly the same sentences were produced by the subjects in the two tests; therefore, the results in the two tests should be comparable.

The success rates of tone production varied for the four tones in each position. In the first syllable, Tone 1 and Tone 4 seemed to be the most successfully produced in the first test and Tone 2 the least successfully. In the second test, Tone 1 was the most successfully produced tone while Tone 2 and Tone 3 the less successfully. In the second syllable (Tone 1 did not occur in this position), Tone 2 was still the least successfully produced in both tests, but Tone 3 was most successfully produced in the first test, and Tone 4 the most successfully produced in the second test. In the third or fourth syllable (one 2-syllable word occurred in one sentence), Tone 4 was produced the least successfully and Tone 2 the most successfully in

the first test, but Tone 1 the least successfully and Tone 4 the most successfully in the second test. Overall, success rates for the production of the four tones were: Tone 2 the least successfully produced tone in both tests; Tone 1 the most successfully produced in the first test and Tone 4 and Tone 1 the most successfully produced in the second test.

The task in Shen's (1989) study (see 3.2.2) is similar to the reading of sentences in this study. However, there are also important differences. Shen's study was based on the reading of a lesson; thus the sentences formed a coherent text. All the sentences were statements and their lengths varied from 3 words (6 syllables) to over 10 words. On the other hand, the sentences read by subjects in this study were all short (3 or 4 syllables). Both questions (yes-no and question word questions) and statements were included. Some were sequenced like questions and answers, but each sentence was preceded by a number, and there was no necessary connection between sentences.

Shen (1989) found that error rates for Tone 1 and Tone 4 were significantly higher than those for Tones 2, 3, and 0 (neutral tone), and that the errors originated principally from tonal register rather than from tonal contour. It is clear from an examination of the success rate in reading short sentences in this study that errors for Tone 1 and Tone 4 were not more common than those for Tone 2 and Tone 3. On the contrary, Tone 2 (and Tone 3 in the second test)

seemed more difficult than Tone 1 and Tone 4, though Tone 1 and Tone 4 seemed to have more errors in the final (third or fourth) syllable in the second test and Tone 4 seemed more difficult in the third syllable in the first test.

Shen argues that Tone 1 and Tone 4 were produced with more errors than Tone 2 and Tone 3 because Tone 1 and Tone 4 are less marked and are more likely to be subject to L1 transfer than the other tones. This explanation obviously cannot be applied to the results found in this study.

In studies on children's acquisition of Mandarin tones (Li and Thompson 1977, 1978, Clumeck 1980), Tone 2 and Tone 3 have been found to be acquired later than Tone 1 and Tone 4 in both perception and production, even though there are individual differences in children acquiring their L1 (rising tone has been found to be produced first, Clumeck 1980). Besides the general tendency for rising tones to be more difficult than level and falling tones for children acquiring tone languages, the phonetic similarities between Tone 2 and Tone 3, and the tone sandhi rule for Tone 3 (realized as Tone 2 before another Tone 3), may account for their difficulty for children learning Mandarin as L1. Tone 2 and Tone 3 also appeared to be more difficult for L2 adult learners in this study, and the same reasons may account for these difficulties. However, it should not be surprising that adult L2 learners exhibit even more variation. Besides, error rate may not always be a good indicator of difficulty.

For instance, Tone 3 may seem to be easier to perceive than Tone 1 and Tone 4 at the beginning since it was perceived at the lowest error rate. In fact this may be a result of confusion between Tone 2 and Tone 3 with a boundary closer to Tone 2: all the errors were made in the direction of misidentifying Tone 2 as Tone 3.

Since the difficulty order of tones may be different in different contexts (e.g., Tone 3 in the first and second syllables of 2-syllable words), the difficulty order of tones can not be easily determined if we consider contexts other than isolated syllables.

6.6.3 Imitation

It is clear that imitation was a much easier task than reading: most tones were produced with over 95% success. The differences between the four tones and the difference between different contexts were therefore small and mostly insignificant.

Isolated syllables

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st test	100%	97%	100%	100%	99%
2nd test	98%	100%	100%	100%	100%

There were no significant differences in the production of the four tones in imitating isolated syllables. The success rates in imitating the four tones were all very high, 100% or very close, in both the first and the second test.

2-syllable words
1st test

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st syll	97%	91%	99%	95%	95%
2nd syll	99%	100%	98%	100%	99%
Overall	98%	95%	98%	98%	97%

2nd test

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st syll	98%	92%	100%	95%	96%
2nd syll	97%	100%	96%	100%	98%
Overall	97%	96%	98%	97%	97%

The average success rates of imitating the four tones in 2-syllable words were not significantly different. However, the difference between the four tones was significant in the first syllable. Tone 2 was less successfully imitated. The overall success rates of imitating the four tones in 2-syllable words did not change significantly between the two tests.

Short sentences
1st test

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st syll	90%	97%	100%	100%	94%
2nd syll		90%	97%	97%	95%
3rd syll	98%	98%	98%	100%	99%
Overall	92%	95%	98%	98%	96%

2nd test

	Tone 1	Tone 2	Tone 3	Tone 4	Average
1st syll	98%	96%	96%	92%	96%
2nd syll		81%	94%	95%	91%
3rd syll	100%	94%	100%	100%	98%
Overall	99%	89%	97%	95%	95%

The success rates in imitating the four tones in the first syllable of short sentences were significantly different in the first test: Tone 1 was less successfully imitated than the other tones. In the second test, however,

the success rates in imitating the four tones were not significantly different, a result of improvement for Tone 1 and slightly drop for the other three tones. The second syllable did not show significant differences among the three tones (Tone 1 did not occur in this position) in the first test, but the difference was significant in the second test, in which Tone 2 was produced at a significantly lower rate than Tone 3 and tone 4. The third and fourth syllables did not show a significant difference for the four tones in either the first or the second test, though Tone 2 was somewhat more difficult than the other tones in the second test. The overall success rates in imitating the four tones were different in the first as well as in the second test. Tone 1 was less successfully imitated in the first test, and Tone 2 was less successfully imitated in the second test.

However, the success rate in imitating any tone in any part of the imitating task is higher than its counterpart in the reading task, and, except for Tone 2 in the second syllable in the second test, all tones were imitated with 90% or higher success rate. Differences in success rates in producing the four tones were much smaller in any part of imitation than in reading, and when a difference is significant, it does not always correspond to a difference in the reading task, indicating quite different processes involved in imitation and reading.

The following table shows the percentage of overall correct production of the four tones in reading and imitation in all syllables.

Reading

	Tone 1	Tone 2	Tone 3	Tone 4	Average
Test 1	79%	73%	77%	77%	76%
Test 2	84%	73%	76%	76%	78%

Imitation

	Tone 1	Tone 2	Tone 3	Tone 4	Average
Test 1	96%	95%	98%	98%	97%
Test 2	98%	94%	98%	97%	97%

The success rates in producing the four tones did not differ significantly in the reading task in the first test, but they differ significantly in the second test: Tone 1 was more successfully produced than the other tones. The success rates in producing the four tones also differ significantly in the imitation task. Tone 2 was produced less successfully than the other tones.

6.7 Error Patterns

When there is a tendency for a group of learners to make the same errors in a certain aspect of L2, it indicates that a general language learning process may be involved. A clear understanding of the data will help us to learn more about the possible L2 learning processes involved.

Is there a tendency for American English speakers to make certain common errors in learning Mandarin tones? We will examine the data and try to find any general patterns. In the following sections, the number as well as the

percentage of correct and incorrect responses from the students will be listed according to tones. Different tasks in different contexts will be discussed separately.

6.7.1 Perception

6.7.1.1 Tone in isolated syllables

Test 1 (isolated syllables)

	Perceived as Tone 1	Tone 2	Tone 3	Tone 4
T1	131 (90%)	8 (6%)	2 (1%)	4 (3%)
T2	2 (1%)	105 (72%)	37 (26%)	1 (1%)
T3	0 (0%)	2 (1%)	139 (96%)	4 (3%)
T4	3 (2%)	13 (9%)	8 (6%)	121 (83%)

Test 2 (isolated syllables)

	Perceived as Tone 1	Tone 2	Tone 3	Tone 4
T1	128(95%)	5 (4%)	1 (1%)	1 (1%)
T2	2 (1%)	121(90%)	12 (9%)	0 (0%)
T3	0 (0%)	11 (8%)	124(92%)	0 (0%)
T4	5 (4%)	2 (1%)	0 (0%)	128(95%)

In the first perception test, when errors were made in identifying a tone in isolated syllables, the most obvious error pattern was that Tone 2 (rising) tended to be misidentified as Tone 3 (falling-rising). There also seems to be a tendency for Tone 1 (high level) to be misidentified as Tone 2 (rising), and Tone 4 (high-falling) to be misidentified as Tone 2 (rising).

In the second test, there still was a clear tendency for Tone 2 to be misidentified as Tone 3, and in addition, a clear tendency for Tone 3 to be misidentified as Tone 2. Tone 1 was still most often misidentified as Tone 2, but Tone 4 was misidentified as Tone 1 more often than Tone 2.

A comparison of the two tests showed that Tone 1 was most often misidentified as Tone 2 on both the first and the second test, though the number of errors was relatively small (6% and 4%). Tone 2, the least successfully identified tone on both tests, was overwhelmingly misidentified as Tone 3 in the first test as well as in the second test. Tone 3 was correctly identified 96% of the time in the first test and the few misidentifications involved both Tone 4 (4 times) and Tone 2 (twice). However, the error rate of Tone 3 increased in the second test and all the errors were misidentifications as Tone 2. The most likely explanation for this change is that some students developed an incorrect phonemic boundary between Tone 2 and Tone 3 closer to the Tone 2 end before the first perception test, so all the errors were made in one direction: Tone 2 being misidentified as Tone 3. After several weeks of learning their perception boundary shifted toward the correct boundary and errors were made in both directions. Tone 4 was more often misidentified as Tone 2 in the first test, but as Tone 1 in the second test, though the number was small (4%).

Kiriloff's study (1969), on the tone perception of isolated syllables, specifically mentions Tone 2's "marked tendency" to be confused with Tone 3. From his data (as shown in Kiriloff, 1969, table II, p.66) we can see that besides Tone 2 being often misidentified as Tone 3 and Tone

3 as Tone 2, Tone 1 was most often misidentified as Tone 2, and Tone 4 as Tone 3. Comparing the error patterns in his data to those found in our data, we find that Tone 1 and Tone 2 error patterns are the same in both studies, and the Tone 3 error pattern in his study is consistent with that in our second test. However, Tone 4 was more often misidentified as Tone 2 than Tone 3 in this study (first test), and it was rarely misidentified as Tone 3 in all other contexts.

Some studies on the perception of tones used synthesized tonal continua between Tone 1 (high level) and Tone 2 (rising) to test both native speakers and non-native speakers (Leather, 1987; Chan, et al. 1975). It has been found that native Mandarin speakers display a sharp category boundary at about the middle of the continuum, while for non-native speakers the boundary was not as sharp and it was shifted more toward the level end of the continuum. This may account for some of the errors made in identifying Tone 1 as Tone 2.

According to Gandour (1978), there are five dimensions in tonal perception: average pitch, direction, length, extreme endpoint, and slope. These five dimensions were the results of the analysis of tone perception data from three language groups: Thai, Yoruba, and American English. The dimension average pitch is the most important dimension for all three language groups. However, the relative importance

attached to this dimension is significantly different for tone language groups, Thai and Yoruba; and the non-tone language group, English. The English group placed relatively more emphasis on average pitch and extreme endpoint in comparison with either the Thai or Yoruba groups. On the other hand, the dimension of direction was more important to tone language groups than to the English group. The errors made in identifying Tone 4 (high falling) as Tone 2 (rising) may be due to this relatively lesser importance attached to direction than average pitch by American English speakers.

6.7.1.2 Tone in 2-syllable words

Test 1

1st syllable

	Tone 1	Tone 2	Tone 3	Tone 4
T1	81(68%)	11 (9%)	1 (1%)	27(23%)
T2	19(13%)	81(54%)	23(15%)	26(17%)
T3	7 (8%)	17(19%)	41(46%)	24(27%)
T4	15(13%)	14(12%)	3 (3%)	88(73%)

2nd syllable

	Tone 1	Tone 2	Tone 3	Tone 4
T1	80(67%)	34(28%)	1 (1%)	5 (4%)
T2	8 (7%)	61(51%)	36(30%)	15(13%)
T3	5 (4%)	5 (4%)	103(86%)	6 (5%)
T4	14(12%)	20(17%)	0 (0%)	86(72%)

In the first syllable of 2-syllable words in the first perception test, Tone 1, Tone 2 and Tone 3 were all most often misidentified as Tone 4, and Tone 4 was the most successfully identified tone. Thus there seemed to be a tendency for all tones in the first syllable to be

identified as Tone 4. However, the misidentification of Tone 2 as Tone 4 (17%) may not represent a meaningful pattern since there were almost as many misidentifications as Tone 3 (15%) and Tone 1 (13%). The misidentification of Tone 1 as Tone 4 was probably due to the similar starting pitch level of the two tones and the shorter duration of the first syllable of 2-syllable words, since the level and falling pitch contours of the two tones are likely to be less prominent in this position. This possibility was also seen in the misidentification of Tone 4 as Tone 1. It should be noted that there were a substantial number of errors made in identifying Tone 1 and Tone 4 besides misidentifying one of them as the other, and the second most frequent error seems to be the same as the most frequent error found in isolated syllables: Tone 1 and Tone 4 were often misidentified as Tone 2. The misidentification of Tone 3 as Tone 4 was probably due to the similarity of falling pitch contour (low-falling in Tone 3 and high-falling in Tone 4) in this position.

In the second syllable, Tone 1 was most often misidentified as Tone 2, Tone 2 was most often misidentified as Tone 3, Tone 3 has the lowest error rate and was misidentified as all other tones at similar rate, and Tone 4 was most often misidentified as Tone 2. The error pattern for the second syllable was the same as that for isolated syllables in the first perception test.

Test 2

1st syllable

	Tone 1	Tone 2	Tone 3	Tone 4
T1	100(93%)	7 (6%)	0 (0%)	1 (1%)
T2	5 (4%)	119(88%)	6 (4%)	5 (4%)
T3	2 (2%)	2 (2%)	70(86%)	7 (9%)
T4	5 (5%)	13(12%)	1 (1%)	89(82%)

2nd syllable

	Tone 1	Tone 2	Tone 3	Tone 4
T1	100(93%)	5 (5%)	0 (0%)	2 (2%)
T2	5 (5%)	91(84%)	6 (6%)	6 (6%)
T3	0 (0%)	5 (5%)	102(94%)	1 (1%)
T4	2 (2%)	14(13%)	2 (2%)	89(82%)

In the first syllable of 2-syllable words in the second test, Tone 1 was most often misidentified as Tone 2, Tone 2 was misidentified as all other tones at similar rate, Tone 3 was often misidentified as Tone 4, and Tone 4 was often misidentified as Tone 2. This error pattern for Tone 2 and Tone 3 was the same as in the first test. However, the tendency to confuse Tone 1 and Tone 4 clearly decreased. The confusion between Tone 1 and Tone 4 mostly disappeared since Tone 1 was misidentified as Tone 4 only 1% of the time, and Tone 4 was misidentified as Tone 1 5% of the time. As a result, the error pattern for Tone 1 and Tone 4 was similar to that in isolated syllables and in the second syllable of 2-syllable words in the first test.

As the second syllable of 2-syllable words, Tone 1 was most often misidentified as Tone 2, Tone 2 was misidentified as all other tones, Tone 3 was most often misidentified as Tone 2, and Tone 4 was most often misidentified as Tone 2. That Tone 3 was most often misidentified as Tone 2 was as in

isolated syllables in the second test, suggesting a shift of perception boundary as mentioned in 6.7.1.1, and also indicating a consistent identification criterion for tone in isolated syllables and tone in 2-syllable words.

6.7.1.3 Tone in short sentences

Test 1

	Tone 1	Tone 2	Tone 3	Tone 4
T1	15(24%)	17(27%)	14(23%)	16(26%)
T2	14(23%)	31(50%)	8(13%)	9(15%)
T3	2 (2%)	26(28%)	63(68%)	1 (1%)
T4	6(19%)	1 (3%)	8(26%)	16(52%)

Test 2

	Tone 1	Tone 2	Tone 3	Tone 4
T1	49(91%)	0 (0%)	0 (0%)	5 (9%)
T2	1 (2%)	39(72%)	10(19%)	4 (7%)
T3	2 (2%)	5 (6%)	73(90%)	1 (1%)
T4	0 (0%)	1 (4%)	3(11%)	23(85%)

In short sentences, the perception task was a little different from that in isolated syllables and 2-syllable words. When a misidentification occurred, a sentence with an incorrect tone was chosen, rather than a correctly produced tone being misidentified as another tone.

In the first test, more instances of Tone 2 were misidentified as Tone 1, but in fact all four tones were chosen at a similar rate; more instances of Tone 1 were misidentified as Tone 2, more instances of Tone 2 were misidentified as Tone 3, and more instances of Tone 3 were misidentified as Tone 4. Except for Tone 1 (which was correctly identified at only about chance level), the

pattern was also the same as found in isolated syllables and 2-syllable words in the first test.

In the second test, more instances of Tone 4 were misidentified as Tone 1, more instances of Tone 3 were misidentified as Tone 2, more instances of Tone 2 were misidentified as Tone 3, and more instances of Tone 3 were misidentified as Tone 4. This is the same pattern as for isolated syllables in the second test. The major difference between the first test and the second test was in Tone 1 and Tone 2. There was more confusion between Tone 1 and Tone 2 in the first test but more between Tone 2 and Tone 3 in the second test.

In W.C.J. Lin's (1985) data for beginners (150 contact hours, closest to our subjects), Tone 3 was the most often misidentified tone and it was mostly confused with Tone 2 in sentence-final position (W.C.J. Lin 1985, table 1, p.36). In sentence-medial position ("*Tade __ hen youming*"), however, though Tone 3 was still the most difficult tone, it was most often misidentified as Tone 4 (W.C.J. Lin 1985, table 4, p.38).

The tendency for students to confuse Tone 3 with Tone 2 in emphasized sentence final position, and to confuse Tone 3 with Tone 4 in sentence medial position, can be attributed to the pitch contours Tone 3 exhibits in these two positions. Since the falling-rising contour in final position is similar to Tone 2 (rising), and the low-falling

contour in non-final position is similar to Tone 4 (high-falling), it was not surprising that Tone 3 was more easily confused with Tone 2 in final position and with Tone 4 in non-final position.

To sum up, in the perception aspect, Tone 1 was most likely to be misidentified as Tone 2, though misidentification of Tone 1 as Tone 4 was also common in the first syllable of 2-syllable words at the beginning. It was rarely misidentified as Tone 3. Tone 2 was most likely to be misidentified as Tone 3, though misidentification as Tone 4 and Tone 1 was also common in 2-syllable words. Tone 3 was most likely to be misidentified as Tone 4 in the first syllable of 2-syllable words. Otherwise misidentification as Tone 2 was more common. The positional difference was most likely due to the allophonic variation of Tone 3 in these positions. Tone 4 was most likely to be misidentified as Tone 2, though misidentification as Tone 1 was also common in 2-syllable words at the beginning.

6.7.2 Reading

6.7.2.1 Tone in isolated syllables

Test 1

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	60(97%)	1 (2%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)
T2	0 (0%)	57(92%)	4 (6%)	1 (2%)	0 (0%)	0 (0%)
T3	0 (0%)	1 (2%)	58(94%)	2 (3%)	1 (2%)	0 (0%)
T4	2 (3%)	0 (0%)	0 (0%)	60(97%)	0 (0%)	0 (0%)

Test 2

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	46(92%)	2 (4%)	1 (2%)	1 (2%)	0 (0%)	0 (0%)
T2	2 (4%)	34(68%)	14(28%)	0 (0%)	0 (0%)	0 (0%)
T3	0 (0%)	2 (4%)	48(96%)	0 (0%)	0 (0%)	0 (0%)
T4	0 (0%)	0 (0%)	0 (0%)	48(96%)	2 (4%)	0 (0%)

In the reading of isolated syllables, only a few tokens were judged to be phonemically incorrect by at least two judges. When the errors made by subjects were few, they can not indicate a general error pattern. Therefore, only the two errors exceeding 5% will be discussed. In the first test, Tone 2 tended to be mispronounced as Tone 3 (6%). The same tendency was found in the second test, and the number of errors increased dramatically (28%). In the two perception tests, the error patterns of Tone 2 and Tone 3 seem to indicate a shift of boundary from closer to Tone 2 toward Tone 3. The increase of Tone 2 mispronounced as Tone 3 may indicate the same shift of boundary.

6.7.2.2 Tone in 2-syllable words

Test 1

1st syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	91(73%)	4 (3%)	7 (6%)	19(15%)	3 (2%)	0 (0%)
T2	14 (9%)	101(65%)	25(16%)	15(10%)	0 (0%)	0 (0%)
T3	1 (1%)	39(42%)	48(52%)	4 (4%)	1 (1%)	0 (0%)
T4	7 (6%)	7 (6%)	5 (4%)	101(81%)	4 (3%)	0 (0%)

2nd syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	90(73%)	19(15%)	3 (2%)	4 (3%)	8 (6%)	0 (0%)
T2	15(12%)	103(83%)	2 (2%)	0 (0%)	4 (3%)	0 (0%)
T3	5 (4%)	3 (2%)	107(86%)	5 (4%)	4 (3%)	0 (0%)
T4	3 (2%)	11 (9%)	9 (7%)	79(64%)	15(12%)	7 (6%)

In reading 2-syllable words in the first test, the tone error patterns in the first syllable and the second syllable were different. In the first syllable Tone 1 was most often mispronounced as Tone 4, and Tone 2 was most often mispronounced as Tone 3, though mispronunciation as Tone 1 and Tone 4 was also common. Tone 3 was most often mispronounced as Tone 2, and Tone 4 was often mispronounced as Tone 1 and Tone 2. In the second syllable Tone 1 was most often mispronounced as Tone 2, Tone 2 was most often mispronounced as Tone 1, and Tone 4 errors were mostly unidentifiable, though mispronunciation as Tone 2 and Tone 3 was also common.

Test 2

1st syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	87(91%)	3 (3%)	1 (1%)	3 (3%)	2 (2%)	0 (0%)
T2	14(15%)	74(77%)	4 (4%)	0 (0%)	4 (4%)	0 (0%)
T3	3 (3%)	45(47%)	46(48%)	1 (1%)	1 (1%)	0 (0%)
T4	10(10%)	12(13%)	5 (5%)	66(69%)	3 (3%)	0 (0%)

2nd syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	72(75%)	8 (8%)	4 (4%)	3 (3%)	9 (9%)	0 (0%)
T2	11(11%)	72(75%)	6 (6%)	2 (2%)	4 (4%)	1 (1%)
T3	1 (1%)	1 (1%)	90(94%)	3 (3%)	1 (1%)	0 (0%)
T4	2 (2%)	0 (0%)	23(24%)	58(60%)	6 (6%)	7 (7%)

In the second test, in the first syllable of 2-syllable words, Tone 2 was most often mispronounced as Tone 1, Tone 3 was most often mispronounced as Tone 2 (actually, mispronunciation as Tone 2 and correctly pronounced Tone 3 were about the same in number), and Tone 4 was most often mispronounced as Tone 1 and Tone 2. In the second syllable, Tone 1 errors were mostly unidentifiable or mispronounced as Tone 2, Tone 2 was most often mispronounced as Tone 1, and Tone 4 was most often mispronounced as Tone 3.

Comparing the error patterns of the two tests, we can see in the first syllable that there was a decrease of Tone 1 mispronunciation as Tone 4. In the first syllable, Tone 2 mispronunciation as Tone 3 and Tone 4 decreased but mispronunciation as Tone 1 increased. In the second syllable, there was a decrease of Tone 1 mispronunciation as Tone 2 but an increase of unidentifiable errors. Tone 2 seemed to be mispronounced as Tone 3 a little more in the second test than in the first test. A similar but greater change also occurred in isolated syllables. Tone 4 mispronunciation as Tone 2 decreased while Tone 4 mispronunciation as Tone 3 increased.

Some studies (Miracle 1989, Shen 1989) classified tonal errors into register and contour errors. The production data in this study were judged on the tone category alone and not classified as register or contour errors. However, some mispronunciation errors were apparently tonal contour

errors. Tone 1 (level) mispronounced as Tone 4 (falling) was such an error, and often occurred when the second syllable started lower in pitch (Tone 2 or Tone 3). Tone 4 mispronounced as Tone 1 was another such error. Tone 4 (falling) mispronounced as Tone 2 (rising) and Tone 2 mispronounced as Tone 4 were also tonal contour error. Some mispronunciation errors were tonal register errors, such as Tone 4 mispronounced as Tone 3 (half Tone 3, low-falling) and vice versa. Tone 2 mispronounced as Tone 1 usually involved a start too high in pitch. Though the rising contour may be maintained to a lesser degree in some cases, it was nonetheless perceived as Tone 1 (level). Tone 2 mispronounced as Tone 3 usually resulted from starting too low or dipping too low, and Tone 3 mispronounced as Tone 2 was a result of not falling low enough at the beginning. Some errors may involve both tonal register and tonal contour. The unidentifiable tone errors usually lacked clear tonal contours and showed ambiguity of pitch level between two tones.

The much higher error rate in producing 2-syllable words than isolated syllables may be due to difficulty in the control of pitch patterns in 2-syllable words, since the length of syllables in 2-syllable words is shorter. However, some other factors seem to be involved. Transfer of the English word-stress pattern stressed-unstressed (high-low in pitch) may account for the different error rates of

the four tones. The error patterns in 2-syllable words, except for the Tone 2 mispronunciation as Tone 3, which also occurred in single syllables, seem to be consistent with this account in that most errors in the first syllable tended to be high in pitch and the errors in the second syllable tended to be low in pitch.

6.7.2.3 Tone in short sentences

1st syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	103(83%)	0 (0%)	1 (1%)	17(14%)	3 (2%)	0 (0%)
T2	4(13%)	14(45%)	0 (0%)	10(32%)	3(10%)	0 (0%)
T3	5(16%)	4(13%)	19(61%)	2 (6%)	1 (3%)	0 (0%)
T4	2 (3%)	3 (5%)	2 (3%)	53(85%)	2 (3%)	0 (0%)

2nd syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T2	7(11%)	31(50%)	17(27%)	6(10%)	1 (2%)	0 (0%)
T3	2 (3%)	3 (5%)	55(89%)	1 (2%)	1 (2%)	0 (0%)
T4	4 (3%)	0 (0%)	8 (6%)	93(75%)	19(15%)	0 (0%)

3rd and 4th syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	48(77%)	9(15%)	1 (2%)	4 (6%)	0 (0%)	0 (0%)
T2	6(10%)	55(89%)	0 (0%)	0 (0%)	1 (2%)	0 (0%)
T3	1 (2%)	11(18%)	45(73%)	3 (5%)	2 (3%)	0 (0%)
T4	1 (3%)	6(19%)	2 (6%)	19(61%)	3(10%)	0 (0%)

Test 2

1st syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	89(93%)	0 (0%)	0 (0%)	5 (5%)	2 (2%)	0 (0%)
T2	7(29%)	16(67%)	0 (0%)	0 (0%)	1 (4%)	0 (0%)
T3	2 (8%)	5(21%)	16(67%)	1 (4%)	0 (0%)	0 (0%)
T4	0 (0%)	1 (2%)	0 (0%)	42(88%)	5(10%)	0 (0%)

2nd syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T2	8(17%)	27(56%)	7(15%)	4 (8%)	2 (4%)	0 (0%)
T3	2 (4%)	6(13%)	35(73%)	5(10%)	0 (0%)	0 (0%)
T4	3 (3%)	1 (1%)	0 (0%)	81(84%)	11(11%)	0 (0%)

3rd and 4th syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	31(65%)	10(21%)	1 (2%)	2 (4%)	3 (6%)	1 (2%)
T2	4 (8%)	42(88%)	0 (0%)	0 (0%)	2 (4%)	0 (0%)
T3	0 (0%)	4 (8%)	40(83%)	2 (4%)	2 (4%)	0 (0%)
T4	1 (4%)	3(13%)	1 (4%)	18(75%)	1 (4%)	0 (0%)

In reading short sentences of three or four syllables, in the first syllable, Tone 1 was most often mispronounced as Tone 4, Tone 2 was also most often mispronounced as Tone 4, and Tone 3 was often mispronounced as Tone 1 and Tone 2. In the second syllable (no Tone 1 occurred in this position), Tone 2 was most often mispronounced as Tone 3, though there were also quite a few mispronounced as Tone 1 and Tone 4. Tone 3 was most often mispronounced as Tone 2, and Tone 4 errors were most often unidentifiable. When a mispronounced Tone 4 was identifiable, it was most often identified as Tone 3. In the third or fourth syllable, Tone 1 was most often mispronounced as Tone 2, Tone 2 was most often mispronounced as Tone 1, Tone 3 was most often mispronounced as Tone 2, and Tone 4 was most often mispronounced as Tone 2.

In the second test, Tone 1 in the first syllable was most often mispronounced as Tone 4 (as in the first test), Tone 2 was most often mispronounced as Tone 1, Tone 3 was

most often mispronounced as Tone 2, and Tone 4 errors were most often judged unidentifiable. In the second syllable, Tone 2 was most often mispronounced as Tone 1 and Tone 3, Tone 3 was most often mispronounced as Tone 2, and Tone 4 mispronunciation was most often judged unidentifiable. In the third or fourth syllable, Tone 1 was most often mispronounced as Tone 2, Tone 2 was most often mispronounced as Tone 1, Tone 3 was most often mispronounced as Tone 2, and Tone 4 was most often mispronounced as Tone 2.

Comparing the error patterns in the two tests, we can see some changes. The mispronunciation of Tone 1 as Tone 4 in the first syllable was the most frequent error in both tests but the error rate became lower in the second test. The mispronunciation of Tone 2 as Tone 4 in the first syllable of a sentence did not appear in the second test. The mispronunciation of Tone 3 as Tone 1 in the first syllable also decreased in the second test. In the second syllable, the mispronunciation of Tone 2 as Tone 3 decreased in the second test but the mispronunciation as Tone 1 did not. The mispronunciation of Tone 3 as Tone 2 and Tone 4 in the second syllable increased in the second test. However, most Tone 4 errors in the second syllable remained unidentifiable in both tests. In the third and fourth syllable, the error patterns remained similar.

It is clear that the error rates of the four tones and the error patterns in each position of the sentence were

different, and that there were changes between the two tests although exactly the same sentences were produced in the two tests. Some errors may be the result of transfer from English intonation. For example, the word *shei* ("who", Tone 2) as the first word of a sentence was frequently mispronounced as Tone 1 or Tone 4 in the first test and as Tone 1 in the second test, but never as Tone 3. It is possible that this is because the equivalent English sentence would start with a high pitch, so that subjects tended to mispronounce this word with a high tone (both Tone 1 and Tone 4 start high). Another example concerns the question particle *ma* (neutral tone, not listed above), which was most often misproduced as Tone 2 (rising). This error also seems to be a transfer of English rising intonation at the end of yes/no questions. These transfer errors can be explained by comparing the intonation pattern of the equivalent English sentence.

The study of Shen (1989) reported that American learners of Mandarin made more errors in producing Tone 1 and Tone 4 (actually this claim could only be applied to the poorer L2 speakers in her study), and Tone 1 and Tone 4 errors were all register errors rather than contour errors. Thus she maintains that "the errors originated principally from tonal register" (p.33). However, other studies (e.g., Miracle 1989 and this study) do not always find Tone 1 and Tone 4 the most difficult, and Tone 1 and Tone 4 errors are

not all register errors. Since in English pitch is used to express affect or grammatical functions and is often affected by stress, it is likely that an American learner of Mandarin would tend to produce tones differently depending on the position of the tone in a sentence or the type of sentence it is in, and possibly the position of a multi-syllable word it is in (as we have found in reading isolated 2-syllable words). Thus more data obtained on all possible conditions are needed before we can confidently make a generalization about the error patterns made by American learners of Mandarin.

To sum up findings in the reading aspect in this study:

1. Tone 1 was most likely to be mispronounced as Tone 4 or Tone 2.

2. Tone 2 was most likely to be mispronounced as Tone 3 in isolation, but also as Tone 1 and Tone 4 (in the first syllable of a short sentence).

3. Tone 3 was most likely to be mispronounced as Tone 2. However, Tone 1 and Tone 4 were also common.

4. Tone 4 was most likely to be mispronounced as Tone 2, though Tone 1 and Tone 3 were also common mispronunciations. And Tone 4 was more often than the other tones judged to be unidentifiable.

The errors made by subjects in this study included both register and contour errors. Some of these errors can be found in children learning Mandarin as their first language,

such as the confusion between Tone 2 and Tone 3 and the substitution of level for contour pitch patterns in non-final position (Clumeck 1980:261), and may be developmental. Some errors such as a high tone in the first syllable and a low tone in the second syllable of 2-syllable words, and the rising contour at the end of questions, probably represent transfer from English word-stress and intonation.

6.7.3 Imitation

isolated syllables

Test 1

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	62(100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
T2	0 (0%)	60(97%)	2 (3%)	0 (0%)	0 (0%)	0 (0%)
T3	0 (0%)	0 (0%)	62(100%)	0 (0%)	0 (0%)	0 (0%)
T4	0 (0%)	0 (0%)	0 (0%)	62(100%)	0 (0%)	0 (0%)

Test 2

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	49(98%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
T2	0 (0%)	50(100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
T3	0 (0%)	0 (0%)	50(100%)	0 (0%)	0 (0%)	0 (0%)
T4	0 (0%)	0 (0%)	0 (0%)	50(100%)	0 (0%)	0 (0%)

2-syllable words

Test 1

1st syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	120(97%)	2 (2%)	0 (0%)	2 (2%)	0 (0%)	0 (0%)
T2	1 (1%)	141(91%)	12 (8%)	0 (0%)	1 (1%)	0 (0%)
T3	0 (0%)	1 (1%)	92(99%)	0 (0%)	0 (0%)	0 (0%)
T4	6 (5%)	0 (0%)	0 (0%)	118(95%)	0 (0%)	0 (0%)

2nd syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	123(99%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
T2	0 (0%)	124(100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
T3	0 (0%)	1 (1%)	121(98%)	2 (2%)	0 (0%)	0 (0%)
T4	0 (0%)	0 (0%)	0 (0%)	124(100%)	0 (0%)	0 (0%)

Test 2

1st syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	96(98%)	0 (0%)	0 (0%)	2 (2%)	0 (0%)	0 (0%)
T2	1 (1%)	91(92%)	6 (6%)	0 (0%)	1 (1%)	0 (0%)
T3	0 (0%)	0 (0%)	98(100%)	0 (0%)	0 (0%)	0 (0%)
T4	5 (5%)	0 (0%)	0 (0%)	94(95%)	0 (0%)	0 (0%)

2nd syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	96(97%)	1 (1%)	0 (0%)	0 (0%)	2 (2%)	0 (0%)
T2	0 (0%)	99(100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
T3	0 (0%)	1 (1%)	94(96%)	3 (3%)	0 (0%)	0 (0%)
T4	0 (0%)	0 (0%)	0 (0%)	98(100%)	0 (0%)	0 (0%)

Short sentences

Test 1

1st syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	111(90%)	2 (2%)	1 (1%)	10 (8%)	0 (0%)	0 (0%)
T2	1 (3%)	30(97%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
T3	0 (0%)	0 (0%)	31(100%)	0 (0%)	0 (0%)	0 (0%)
T4	0 (0%)	0 (0%)	0 (0%)	62(100%)	0 (0%)	0 (0%)

2nd syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T2	5 (8%)	56(90%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)
T3	1 (2%)	0 (0%)	60(97%)	0 (0%)	1 (2%)	0 (0%)
T4	1 (1%)	1 (1%)	2 (2%)	120(97%)	0 (0%)	0 (0%)

3rd and 4th syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	61(98%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
T2	1 (2%)	61(98%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
T3	0 (0%)	1 (2%)	61(98%)	0 (0%)	0 (0%)	0 (0%)
T4	0 (0%)	0 (0%)	0 (0%)	31(100%)	0 (0%)	0 (0%)

Test 2
1st syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	94(98%)	1 (1%)	0 (0%)	1 (1%)	0 (0%)	0 (0%)
T2	0 (0%)	23(96%)	0 (0%)	1 (4%)	0 (0%)	0 (0%)
T3	0 (0%)	1 (4%)	23(96%)	0 (0%)	0 (0%)	0 (0%)
T4	0 (0%)	2 (4%)	0 (0%)	44(92%)	2 (4%)	0 (0%)

2nd syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T2	8(17%)	39(81%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)
T3	0 (0%)	3 (6%)	45(94%)	0 (0%)	0 (0%)	0 (0%)
T4	1 (1%)	3 (3%)	1 (1%)	91(95%)	0 (0%)	0 (0%)

3rd and 4th syllable

	Pronunciation judged as					
	Tone 1	Tone 2	Tone 3	Tone 4	Uniden	Tone 0
T1	48(100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
T2	2 (4%)	45(94%)	0 (0%)	0 (0%)	1 (2%)	0 (0%)
T3	0 (0%)	0 (0%)	48(100%)	0 (0%)	0 (0%)	0 (0%)
T4	0 (0%)	0 (0%)	0 (0%)	24(100%)	0 (0%)	0 (0%)

The tone error rates in imitation tasks were generally low, usually below 5% for all four tones in any part, and the errors in any category rarely exceeded those in the counterpart reading task. An error rate exceeding 5% in the imitation task occurred mostly with Tone 2: it was mispronounced as Tone 3 in the first syllable of 2-syllable words in the first test (8%) and the second test (6%), and as Tone 1 in the second syllable of a short sentence in the first test (8%) and the second test (17%). Tone 1 was imitated as Tone 4 in the first syllable of a short sentence 8% of the time in the first test. These errors were not different from those that occurred in the reading task and will not be discussed further.

7. CONCLUSIONS

7.1 Summary of findings

The results of this study can be summarized as follows:

1. Tonal context (isolated syllables, 2-syllable words, and short sentences) has clearly influenced the perception and production of tone by adult L2 learners. However, the effect on perception was greater at the beginning and decreased as learning progressed.

Tonal context has a similar effect on tone perception and reading, but not on imitation; tone perception and production in isolated syllables was better than in sequences (2-syllable words and short sentences). Tone perception in 2-syllable words was better than in short sentences. However, tone production in short sentences was better than in 2-syllable words. Subjects' familiarity with the sentences seemed to have changed the expected context effect in reading.

2. Different tasks (imitation and reading) have influenced L2 learners' tone production. Imitation was better than reading in all contexts from the beginning to the end of the term, indicating imitation was a much easier task than reading. The discrepancy between reading and imitation was greater in tone sequences (2-syllable words and short sentences) than in isolated syllables.

There was no significant correlation between the two tasks at the beginning of Mandarin learning, suggesting that they were not related. However, correlations between them in isolated syllables and short sentences became significant at the end of the term, thus they seemed to get more related as learning progressed.

3. Whether L2 tone perception and production is affected by sex cannot be determined by this study. Contrary to findings in some other studies on L2 pronunciation, male subjects scored higher than female subjects in tone perception and production. However, statistical results were inconclusive. The difference between males and females also decreased as learning progressed. Females seem to have a greater improvement than males in short sentences, which had been taught in class and were familiar to the subjects, though the difference was not significant.

4. Prior experience of learning other L2s has no significant effect on the perception or production of tone. This is consistent with findings in some studies that other L2 experience has no effect on L2 pronunciation or sound perception.

Prior Mandarin learning experience (no more than a year) has no significant effect on the perception or production of tone, with only one exception. This exception was in imitating short sentences at the end of the term;

subjects who had learned Mandarin before did worse than others in imitating short sentences they had learned. It is possible that they paid less attention to the auditory model when they imitated items they were familiar with, or they may have formed a habit of producing tones and it became harder to modify it. Because of the small number of subjects with prior Mandarin learning experience in this study, future studies on more L2 learners with different lengths of experience will be needed to provide further information on the effect of L2 learning on the pronunciation of imitating L2 utterances.

5. The correlation between the perception of tone in isolated syllables and in sequences (2-syllable words and short sentences) was generally significant. The correlation was higher between tone in isolated syllables and 2-syllable words, and between 2-syllable words and short sentences. The correlation between tone in isolated syllables and in short sentences was low at the beginning and became insignificant at the end of the term. The ceiling effect of tone perception in isolated syllables, and the teaching of short sentences in the class, may have affected the correlation.

On the other hand, the correlation between the production of tone in isolated syllables and in sequences was not significant, possibly because of the ceiling effect of the production of tone in isolated syllables. The

correlation between reading 2-syllable words and short sentences was significant, but the correlation was higher at the beginning of the term than at the end of the term.

6. The correlation between tone perception and tone reading was significant in isolated syllables and 2-syllable words. Significant (and higher) correlation was also found between tone perception and reading across the two tests, suggesting a delayed effect between the development of perception and production of tone. The correlation between tone perception and tone reading in short sentences was significant at first but became insignificant at the end of the term. Since short sentences had been taught in the class, this suggests that instruction may have some effect on the relation between perception and reading. Tone production in the imitating task generally did not correlate significantly with tone perception.

7. There was significant improvement in tone perception between the two tests conducted at the beginning and at the end of the term. The improvement occurred in all isolated syllables, 2-syllable words, and short sentences. However, there was no significant improvement in tone production, either in reading or imitation, between the beginning and the end of the term. The ceiling effect appear to have been reached for imitation from the beginning. Even if we disregard imitation, production was better than perception

at the beginning of the term, but perception became better at the end of the term.

8. The four Mandarin tones caused different degrees of difficulty for L2 learners as measured by error rates. However, the order of difficulty seemed different in different contexts at different times. Tone 2 was the most difficult tone to learn in isolation. Tone 3 was the most difficult in the first syllable of 2-syllable words but the easiest in the second syllable of 2-syllable words.

9. The tone error patterns also differ depending on the context. Variation found in the same context shows that there is probably more then one factor explaining the errors. Both register and contour errors were found. Errors may be developmental or transfer from word-stress (see 6.6.2.2 and 6.7.2.2) and intonation (see 6.7.2.3) of English.

Basically the learning of Mandarin tones by English-speaking learners is not different from the learning of other aspects of L2 phonology by L2 learners. Of course, specific tonal errors and tonal contexts are unique to the learning of Mandarin.

7.2 Suggestions for teaching Mandarin tones

Some suggestions of teaching Mandarin tones to adult L2 learners can be made based on these conclusions:

1. Tones in 2-syllable words should be taught more. The teaching of tones in citation forms is the basis of teaching tones in longer utterances. However, the correlations found between tones in citation form and tones in short sentences were usually low or insignificant. On the other hand, the correlations found between 2-syllable words and short sentences were always significant. Since tones in 2-syllable words either precede or are followed by another tone, they are more like tones in longer sequences. Tone sandhi rules that apply in utterances basically apply in sequences of two tones. Thus the learning of tones in 2-syllable words may have some positive effect on learning tones in longer utterances. Besides, 2-syllable words are almost as common as 1-syllable words in Mandarin and both are the basic building blocks of longer utterances. Lin (1985) proposed to teach the norm of Tone 3 as low-falling (see 3.2.1) and showed the effectiveness of this new treatment of Tone 3. In fact, since the low-falling-rising contour of Tone 3 will inevitably occur in the citation form and in an emphasized position, the best way to teach the low-falling variant is to teach it in the right context, and 2-syllable words provide such a context. The suggestion of teaching tones in 2-syllable words has been proposed in other studies on the teaching of Mandarin tones (M-Y. Chen 1983).

2. More time should be spent on the perception rather than the production of Mandarin tones. The production of acceptable isolated Mandarin tones seemed to be easily learned at the beginning of the term; however, the production of tones did not seem to be easily improved within one term. On the other hand, the perception of tones seemed more difficult at the beginning of learning, but it can be improved more easily than production over one term of learning, especially for tones in sequence. It thus seems more effective to spend more time on perception than production. Since the perception and the production (reading) of tone correlated significantly, even though evidence of a transfer of improvement in perception into improvement in production is inconclusive, it is quite possible that time spent on tone perception may also have some positive effect on tone production. This suggestion should not be taken to mean that production training is not needed. In fact, because production does not improve easily, it may be even more important to help students begin to learn correct pronunciations from the very beginning.

3. Students should be made aware of transfer errors from English word-stress and sentence intonation. Tonal errors made in producing isolated syllables were relatively few compared to errors made in tone sequences. Transfer of English word-stress and sentence intonation appears to be an important cause of tonal errors which students made in

reading 2-syllable words (see 6.6.2.2 and 6.7.2.2) and short sentences (see 6.7.2.3).

There is evidence that form-focused instruction has an effect on the accuracy of L2 learners' use of L2 structures (Lightbown and Spada 1990), supporting the notion that learners can benefit from "consciousness raising" (Rutherford 1987, Rutherford and Sharwood-Smith 1988). Since English (or other non-tone language) speakers may be less conscious about the use of pitch in word-stress and sentence intonation in their native language, they may benefit from explicit knowledge of the differences in pitch range and pitch contour between English and Mandarin. While errors may be inevitable in learning an L2, awareness of the source of errors may be helpful for adult learners reducing or correcting their errors. Because English sentence intonation is superimposed on the whole sentence and the pitch patterns vary depending on the meaning of a particular sentence, a particular intonation pattern is likely to be transferred to Mandarin sentences with similar meanings. Thus it may be necessary and more effective to call students' attention to intonation transfer with particular Mandarin sentences as students learn them.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abramson, A. S. and Svastikula, K. 1983. Intersections of tone and intonation in Thai. Haskins Laboratories Status Report on Speech Research, Apr.-Sep., 143-154.

Albert, M. and Obler, L. 1978. The Bilingual Brain. New York: Academic.

Asher, J. J. and Garcia, R. 1969. The optimal age to learn a foreign language. Modern Language Journal 53, 334-342.

Baran,J. A., Laufer, M. Z. and Daniloff, R. 1977. Phonological contrastivity in conversation: A comparative study of voice onset time. Journal of Phonetics 5, 339-350.

Beebe, L. M. 1987. Myths about interlanguage phonology. Interlanguage Phonology. G. Ioup and S. Weinburger (eds.), 165-175. Cambridge: Newbury House.

Beebe, L. M. 1987. Sociolinguistic variation and style shifting in second language acquisition. G. Ioup and S. Weinburger (eds.), 378-388. Cambridge: Newbury House.

Berko, J. and Brown, R. 1960. Psycholinguistic research methods. Handbook of Research Methods in Child Development. P. Mussen (ed.). New York: Wiley.

Boliger, D. L. 1958. A Theory of Pitch Accent in English. Word 14, 109-149.

Bolinger, D. L. (ed.) 1972. Intonation. Baltimore: Penguin Books Inc.

Boliger, D. L. 1975. Aspects of Language. Second edition. New York: Harcourt Brace Jovanovich, Inc.

Boliger, D. L. 1978. Intonation across languages. Universals of Human Language, vol. 2. J. Greenberg (ed.), 471-524. Stanford, CA: Stanford University Press.

Bolinger, D. L. 1986. Intonation and Its Parts. Stanford, CA: Stanford University Press.

Borden, G., Gerber, A. and Milsark, G. 1983. Production and perception of the /r/-/l/ contrast in Korean adults learning English. Language Learning 33, 499-526.

Brière, E. 1966. An investigation of phonological interference. Language 44, 768-796.

Broselow, E., Hurtig, R. and Ringen, C. 1987. The perception of second language prosody. Interlanguage Phonology. G. Ioup and S. Weinburger (eds.), 350-361. Cambridge: Newbury House.

Caramazza, A., Yeni-Komshian, G., Zurif, E., and Carbone, E. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. Journal of the Acoustical Society of America 54, 421-428.

Carroll, J. B. 1981. Twenty-five years of research on foreign language aptitude. Individual Differences and Universals in Language Learning Aptitude. K. C. Diller (ed.), 83-118. Rowley: Newbury House Publishers.

Catford, J. C. and Pisoni, D. B. 1970. Auditory vs. Articulatory Training in Exotic Sounds. Modern Language Journal 54, 477-481.

Chan, S. W., Chuang, C.-K and Wang, W. S.-Y. 1975. Cross-language study of categorical perception for lexical tone. Manuscript of paper read at the 90th Meeting of the Acoustical Society of America, 1975. (Abstracted in Journal of Acoustical Society of America, 58, S119.)

Chang, Nien-Chuang T. 1958. Tones in Intonation in the Chengtu Dialect (Szechuan, China). Phonetica, 2, 59-84.

Chao, Yuen-Ren. 1930. A system of tone letters. Le maître phonétique 45, 24-27.

Chao, Yuen-Ren. 1933. Tone and intonation in Chinese. Bulletin of the Institute of History and Philology 4, 121-134.

Chao, Yuen-Ren. 1948. Mandarin Primer. Cambridge: Harvard University Press.

Chao, Yuen-Ren. 1956. Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. For Roman Jakobson. M. Halle, H. G. Lunt, H. Mclean and C. H. Van Schooneveld (eds.), 52-59. The Hague: Mouton.

Chao, Yuen-Ren. 1968. A Grammar of Spoken Chinese. Berkeley: University of California Press.

Chen, Gwang-tsai. 1974. The pitch range of English and Chinese speakers. Journal of Chinese Linguistics 2, 159-71.

Chen, Ming-yuan. 1981. Shuli tongji zai Hanyu yanjiu zhong de yingyong (Application of statistics in the study of Mandarin). Zhongguo yuwen 6, 466-474.

Chen, Ming-yuan. 1983. Hanyu zuowei di er yuyan jiaoxue zhong de yuyin wenti (Phonetic problems in the teaching of Chinese as a second language). Language Learning and Communication 2:3, 249-265.

Cheng, Chin-chuan. 1968. English stresses and Chinese tones in Chinese sentences. Phonetica 18, 77-88.

Chiang, T. 1979. Some interferences of English intonations with Chinese tones.International Review of Applied Linguistics 17, 245-250.

Chomsky, N. and Halle, M. 1968. The Sound Pattern of English. New York: Harper and Row.

Clumeck, H. 1980. The acquisition of tone. Child Phonology, Vol.1, Production. G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (eds.), 257-275. New York: Academic Press.

Cole, R. A. 1981. Perception of fluent speech by children and adults. Native Language and Foreign Language Acquisition. H. Winitz (ed.), 92-109. New York: The New York Academy of Sciences.

Connell, B., Hogan, J. and Rozsypal, A. 1983. Experimental evidence of interaction between tone and intonation in Mandarin Chinese. Journal of Phonetics 11, 337-351.

Couper-Kuhlen, E. 1986. An Introduction to English Prosody. Baltimore: Edward Arnold.

Cruttenden, A. 1986. Intonation. New York: Cambridge University Press.

Crystal, D. 1972. The intonation system of English. Intonation. Bolinger, D. L. (ed.), 110-136. Baltimore: Penguin Books Inc.

Cutler, A. 1977. The context-dependence of "intonation meanings". Chicago Linguistic Society 13, 104-115.

Dickerson, L. 1974. Internal and external patterning of phonological variability in the speech of Japanese learners of English: Toward a theory of second language acquisition. Ph.D. dissertation, University of Illinois.

Dickerson, L. 1975. The learner's interlanguage as a system of variable rules. TESOL Quarterly 9, 401-407.

Dickerson, W. 1976. The psycholinguistic unity of language learning and language change. Language Learning 26, 215-231.

Dickerson, W. 1977a. Interlanguage phonology: Current research and future directions. The Notions of Simplification, Interlanguages and Pidgins. S. Corder and E. Roulet (eds.). Neuchatel: Universite de Neuchatel.

Dickerson, W. 1977b. Language variation in applied linguistics. International Review of Applied Linguistics 35, 43-66.

Dickerson, L. and Dickerson, W. 1977. Interlanguage phonology: Current research and future directions. The Notions of Simplification, Interlanguages and Pidgins and Their Relation to Second Language Pedagogy. S. Corder and E. Roulet (eds.), 18-29. Geneva: Droz.

Dreher, J. J. and Lee, Pao-Chen. 1968. Instrumental investigation of single and paired Mandarin tonemes. Monumenta Serica 20, 343-373.

Dulay, H. and Burt, M. 1973. Should we teach children syntax? Language Learning 23, 245-258.

Elman, J., Diehl, R., and Buchwald, S. 1977. Perceptual switching in bilinguals. Journal of the Acoustical Society of America 62, 971-974.

Flege, J. 1987. The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. Journal of Phonetics 15, 47-65.

Flege, J. 1988. The production and perception of foreign language speech sounds. Human Communication and Its Disorders. H. Winitz (ed.), 224-401. Norwood, NJ: Ablex Publishing.

Flege, J. and Davidian, R. 1985. Transfer and developmental processes in adult foreign language speech production.

Journal of Applied Psycholinguistic Research 5, 323-347.

Flege, J. and Hillenbrand, J. 1984. Limits on phonetic accuracy in foreign language speech production. Journal of the Acoustical Society of America 76, 708-721.

Fry, D. B. 1955. Duration and intensity as physical correlates of linguistic stress. Journal of the Acoustical Society of America 27, 765-768.

Fry, D. B. 1958. Experiments in the perception of stress. Language and Speech 1, 126-152.

Gandour, J. 1978. The Perception of Tone. Tone: A Linguistic Survey. Fromkin, V. (ed.), 41-76. New York: Academic Press.

Gao, Yu-zhen. 1980. Beijing hua de qingsheng wenti (Issues of the neutral tone in Beijing Mandarin). Yuyan Jiaoxue yu Yanjiu 2, 82-98.

Gårding, E., Kratochvil, P., Svantesson, J. and Zhang, J. 1986. Tone 4 and Tone 3 discrimination in Modern Standard Chinese. Language and Speech 29, 281-293.

Gass, S. 1984. Development of speech perception and speech production abilities in adult second language learners. Applied psycholinguistics 5, 51-74.

Gass, S. and Madden, C. (eds.) 1985. Input in Second Language Acquisition. Rowley, Mass.: Newbury House.

Gatbonton, E. 1975. Systematic variation in second language speech: A sociolinguistic study. Ph.D. Dissertation, McGill University.

Gillette, S. 1980. Contextual variation in the perception of L and R by Japanese and Korean speakers. Minnesota Papers in Linguistics and the Philosophy of Language (No.6, 59-72). University of Minnesota, Minneapolis, MN.

Goldstein, H. 1983. Word recognition in a foreign language: A study of speech perception. Journal of Psycholinguistic Research 12, 417-427.

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

Harris, J. 1969. Spanish Phonology. Cambridge: MIT Press.

Hecht, B. F. and Mulford, R. 1982. The acquisition of a second language phonology: interaction of transfer and developmental factors. Applied Psycholinguistics 3, 313-328.

Hieke, A. E. 1990. Toward listener strategies for decoding fluent speech. International Review of Applied Linguistics 28, 221-233.

Ho, Aichen T. 1976a. The acoustic variation of Mandarin tones. Phonetica 33, 353-367.

Ho, Aichen T. 1976b. Mandarin tones in relation to sentence intonation and grammatical structure. Journal of Chinese Linguistics 4, 1-13.

Ho, Aichen T. 1977. Intonation variations in a Mandarin sentence for three expressions: interrogative, exclamatory, and declarative. Phonetica 34, 446-456.

Howie, J. M. 1976. Acoustical studies of Mandarin vowels and tones. New York: Cambridge University Press.

Ioup, G. and Tansomboon, A. 1987. The acquisition of tone: A maturation perspective. Interlanguage Phonology. G. Ioup and S. Weinberger (eds.), 333-349. Cambridge: Newbury House.

Kiriloff, C. 1969. On the auditory perception of tones in Mandarin. Phonetica 20, 63-67.

Klatt, D. H. and Stevens, K. N. 1973. On the automatic recognition of continuous speech: implications from a spectrogram-reading experiment. IEEE Transactions on Audio and Electroacoustics, AU-21, 210-217.

Krashen, S. D. 1977. Some issues relating to the Monitor Model. On TESOL '77. H. Brown, C. Yorio, and R.Crymes (eds.), 144-158. Washington D.C.: TESOL.

Krashen, S.D. 1981. Aptitude and attitude in relation to second language acquisition and learning. Individual Differences and Universals in Language Learning Aptitude. K. C. Diller (ed.), 155-175. Rowley: Newbury House Publishers.

Krashen, S.D. 1982. Principles and Practice in Second Language Acquisition. New York: Pergamon.

Krashen, S. D. and Terrell, T. D. 1983. The Natural Approach. New York: The Alemany Press.

Larsen-Freeman D. and Long, M. H. 1991. An Introduction to Second Language Acquisition Research. New York: Longman.

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

Leather, J. 1987. F0 pattern inference in the perceptual acquisition of second language tone. Sound Patterns in Second Language Acquisition, A. James and J. Leather (eds.), 41-80. Netherlands: Foris Publications.

Lehiste, I. 1970. Suprasegmentals. Cambridge: the M.I.T. Press.

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

Li, C. N. and Thompson, S. A. 1978. The acquisition of tone. Tone: A linguistic Survey. V. A. Fromkin (ed.), 271-284. New York: Academic Press.

Liberman, M. and Sag, I. 1974. Prosodic form and discourse function. Papers from the Tenth Regional Meeting, Chicago Linguistic Society, 416-427.

Lieberman, P. 1960. Some acoustic correlates of word stress in American English. The Journal of the Acoustical Society of America 32, 451-454.

Lieberman, P. 1965. On the acoustic basis of the perception of intonation by linguists. Word 21, 40-54.

Lightbown, P. and Spada, N. 1990. Focus-on-form and corrective feedback in communicative language teaching. Studies in Second Language Acquisition 12, 429-446.

Lin, Mao-can and Yan, Jing-zhu. 1980. Beijinghua qingsheng de shengxue xingzhi (The acoustic characteristics of neutral tone in Beijing Mandarin). Fang Yan 3, 166-178.

Lin, William C.J. 1985. Teaching Mandarin tones to adult English speakers: Analysis of difficulties with suggested remedies. Regional English Language Center Journal 16, 31-47.

Lindström, O. 1978. Aspects of English Intonation. ACTA Universitatis Gothoburgensis.

Lisker, L. 1970. On learning a new contrast. Haskins Laboratory Status Report on Speech Research No. 24, 1-15. New Haven, CT.

Lisker, L. and Abramson, A. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word 20, 384-422.

Lisker, L. and Abramson, A. 1967. Some effects of context on voice onset time in English stops. Language and Speech 10, 1-28.

Liu, Xiun, Guo, En-Ming, and Liu, She-Hui (eds.), 1988 Practical Chinese Reader, Book I. Beijing: the Commercial Press.

Locke, J. L. 1970. The value of repetition in articulation learning. International Review of Applied Linguistics 8, 147-154.

Locke, J. L. 1983. Phonological Acquisition and Change. New York: Academic Press.

Loewenthal, K. and Bull, D. 1984. Imitation of foreign sounds: What is the effect of age? Language and Speech 27, 95-97.

Lyovin, A. 1978. Review of Tone and Intonation in Mandarin Chinese, by Rumjancev, M. Moscow: Izdatel'stvo Moskovskogo Universiteta. Journal of Chinese Linguistics 6, 120-168.

MacKain, K. 1988. Filling the gap between speech and language. The Emergent Lexicon. M. D. Smith and J. L. Locke (eds.), 51-74. New York: Academic Press.

MacKain, K., Best, C. and Strange, W. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. Applied Psycholinguistics 2, 369-390.

Major, R. C. 1987. Phonological similarity, markedness, and rate of L2 acquisition. Studies in Second Language Acquisition 9, 63-82.

Massaro, D., Cohen, M. and Tseng, Chiu-yu. 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.

McClasky, C., Pisoni, D. and Carrell, T. 1983. Transfer of training of a new linguistic contrast in voicing. Perception and psychophysics 34, 323-330.

Miller, G. A., Heise, G. and Lichten, W. 1953. The intelligibibity of speech as a function of the context of the test materials. Journal of Experimental Psychology 41, 329-335.

Miracle, W. C. 1989. Tone production of America students of Chinese: a preliminary acoustic study. Journal of Chinese Language Teachers Association 14:3, 49-65.

Mochizuki, M. 1981. The identification of /r/ and /l/ in natural and synthesized speech. Journal of Phonetics 9, 283-303.

Morton, J. and Jassem, W. 1965. Acoustic correlates of stress. Language and Speech 8, 159-181.

Mueller, T. H. and Niedzielski, H. 1968. The influence of discrimination training on pronunciation. Modern Language Journal 52, 410-416.

Neufeld, G. 1978. On the acquisition of prosodic and articulatory features in adult language learning. The Canadian Modern Language Review 34, 163-174.

Neufeld, G. 1979. Towards a theory of language learning ability. Language Learning 29, 227-241.

Neufeld, G. 1980. On the adult's ability to acquire phonology. TESOL Quarterly 14, 285-298.

Neufeld, G. and Schneiderman, E. 1980. Prosodic and articulatory features in adult language learning. Research in Second Language Acquisition, R. Scarcella, and S. Krashen (eds.), 105-109. Cambridge: Newbury House Publishers.

Obler, L. 1982. The Parsimonious Bilingual. Exceptional Language and Linguistic Theory. L. Obler and L. Menn (eds.), 339-346. New York: Academic Press.

Odlin, T. 1989. Language Transfer. New York: Cambridge University Press.

Olson, L. and Samuels, S. J. 1973. The relationship between age and accuracy of foreign language pronunciation. Journal of Educational Research 66, 263-267.

Oyama, S. 1976. A sensitive period for the acquisition of a nonnative phonological system. Journal of Psycholinguistic Research 5, 261-283.

Pike, E. V. 1959. A test for predicting phonetic ability. Language Learning 9, 35-43.

Pike, K. L. 1972. General characteristics of Intonation. D. L. Bolinger (ed.), 53-82. Baltimore: Penguin Books Inc.

Pimsleur, P. 1963. Discrimination training in the teaching of French pronunciation. Modern Language Journal 47, 199-203.

Piper, T. and Cansin, D. 1988. Factors influencing the foreign accent. The Canadian Modern Language Review 44, 335-342.

Pisoni, D., Aslin, R., Perey, A. and Hennessy, B. 1982. Some effects of laboratory training on identification and discrimination of voicing contrasts. Journal of Experimental Psychology: Human Perception and Performance 8, 297-314.

Pisoni, D. and Luce, P. 1986. Speech perception: research, theory, and the principle issues. Pattern Recognition by Humans and Machines, vol. 1, Speech Perception. Schwab and Nusbaum (eds.), 1-50. Orlando: Academic Press.

Purcell, E. T. and Suter, R. W. 1980. Predictors of pronunciation accuracy: a reexamination. Language Learning 30, 271-287.

Rivers, W. 1979. Learning a sixth language: an adult learner's diary. Canadian Modern Language Review 36, 67-82.

Ross, E. D., Edmondson, J.A., and Seibert, G.B. 1986. The effect of affect on various acoustic measures of prosody in tone and non-tone languages. Journal of Phonetics 14, 283-302.

Rutherford, W. 1987. Second Language Grammar: Learning and Teaching. London: Longman.

Rutherford, W. and Sharwood-Smith, M. (eds.) 1988. Grammar and Second Language Teaching. New York: Newbury House.

Sato, C. 1985. Task variation in interlanguage phonology. Input in Second Language Acquisition. S. Gass and C. Madden (eds.), 181-196. Rowley, Mass.: Newbury House.

Schneiderman, E., Bourdages, J. and Champagne, C. 1988. Second-language accent: the relationship between

discrimination and perception. Language Learning 38, 1-19.

Scott, D. 1987. Prosody. New Horizons in Linguistics 2. J. Lyons, R. Coates, M. Deuchar and G. Gazdar (eds.), 82-102. Baltimore: Penguin Books.

Scovel, T. 1988. A Time to Speak. New York: Newbury House Publishers.

Segalowitz, N. and Gatbonton, E. 1977. Studies of the non-fluent bilingual. Bilingualism: Psychological, Sociological and Educational Implications, P. Hornby (ed.), 77-89. New York: Academic.

Seliger, H., Krashen, S. and Ladefoged, P. 1975. Maturational constraints in the acquisition of second language accent. Language Sciences, Aug. 1975, No. 36, 20-22.

Sheldon, A. and Strange, W. 1982. The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. Applied Psycholinguistics 3, 243-261.

Shen, Xiaonan S. 1985. A contrastive study of Mandarin Chinese and French interrogative intonologies: A contribution to the prosody of Chinese. Ph.D. dissertation, University of California, Berkeley.

Shen, Xiaonan S. 1989. Toward a register approach in teaching Mandarin tones. Journal of Chinese Language Teachers Association 14:3, 27-47.

Shen, Xiaonan S. 1990. The prosody of Mandarin Chinese. University of California Publications in Linguistics, vol. 118.

Shimizu, K., and Dantsuji, M. 1983. A study of the perception of /r/ and /l/ in natural and synthetic speech sounds. Studia Phonologica 17, 1-14.

Singleton, D. and Little, D. 1984. A first encounter with Dutch: Perceived language distance and language transfer as factors in comprehension. Language Across Cultures. L. MacMathuna and D. Singleton (eds.). Dublin: Irish Association for Applied Linguistics.

Smith, M. D. and Locke, J. L. (eds.) 1988. The Emergent Lexicon. New York: Academic Press.

Snow, C. E. and Hoefnagel-Höhle, M. 1977. Age differences in the pronunciation of foreign sounds. Language and Speech 20, 357-365.

Snow, C. E. and Hoefnagel-Höhle, M. 1978. The critical period for language acquisition: evidence from second language learning. Child Development 49, 1114-1128.

Snow, C. E. and Hoefnagel-Höhle, M. 1979. Individual differences in second-language ability: a factor-analytic study. Language and Speech 22, 151-162.

Strange, W. and Jenkins, J. 1978. The role of linguistic experience in the perception of speech. Perception and Experience. R. Walk and H. Pick, Jr. (eds.), 125-169. New York: Plenum.

Suter, R. 1976. Predictors of pronunciation accuracy in second language learning. Language Learning 26, 233-253.

Tahta, S. 1980. An investigation into second language learning. Unpublished M. Phil. thesis. City University, London.

Tahta, S., Wood, M. and Loewenthal, K. 1981a. Foreign accents: factors relating to transfer of accent from the first language to a second language. Language and Speech 24, 265-272.

Tahta, S., Wood, M. and Loewenthal, K. 1981b. Age changes in the ability to replicate foreign pronunciation and intonation. Language and Speech 24, 363-372.

Tees, P. C. and Werker, J. J. 1984. Perceptual flexibility: Maintainance or recovery of the ability to discriminate nonnative speech sounds. Canadian Journal of Psychology 38, 579-590.

Thompson, I. 1984. An experimental study of foreign accents. Ph.D. Dissertation, George Washington University.

Tseng, chiu-yu 1981. An Acoustic Phonetic Study on Tones in Mandarin Chinese. Ph.D. Dissertation, Brown University.

Van Els, T. and De Bot, K. 1987. The role of intonation in foreign accent. The Modern Language Journal 71, 147-155.

Walz, J. 1975. A longitudinal study of the acquisition of French pronunciation. Ph.D. dissertation, Indiana University.

Wang, W. S-Y. and Li, Kung-Pu. 1967. Tone 3 in Pekinese. Journal of Speech and Hearing Research 10, 629-636.

White, C. 1980. Mandarin Tone and English Intonation: A Contrastive Analysis. M.A. thesis, the University of Arizona.

White, C. 1981. Tonal pronunciation errors and interference from English intonation. Journal of Chinese Language Teachers Association 16:2, 27-56.

Williams, L. 1977. The perception of stop consonant voicing by Spanish-English bilinguals. Perception and Psychophysics 21, 289-297.

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

Williams, L. 1980. Phonetic variation as a function of second-language learning. Child Phonology, Vol. 2, Perception. G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (eds.), 185-216. New York: Academic.

Winitz, H. 1981. Input considerations in the comprehension of first and second language. Native Language and Foreign Language Acquisition. H. Winitz (ed.), 296-308. New York: the New York Academy of Sciences.

Wu, Zong-ji. 1982. Putonghua yuju zhong de shengdiao bianhua (Tonal variations in Mandarin sentences). Zhongguo Yuwen 6, 439-449.

Yamada, J., Takatsuka, S., Kotake, N. and Kurusu, J. 1980. On the optimum age for teaching foreign vocabulary to children. International Review of Applied Linguistics 18, 245-247.

Zhao, J-M. 1987. Cong Yixie Shengdiao Yuyan de Shengdiao Shuodao Hanyu Shengdiao. Paper presented at the 2nd International Coference on Teaching Chinese. August 10-14, 1988, Beijing.

APPENDICES

APPENDIX A

Tone Perception

Part (A): 1-syllable words. Please listen to each word, then write down the tone number or tone mark on your answer sheet. Each word will be spoken twice. Please listen carefully.

1. ku4	2. ba3	3. lu2	4. dı1
5. peı2	6. pa4	7. mei3	8. paoı
9. ni2	10. ful	11. mi3	12. da2
13. bu4	14. gueı3	15. hao4	16. feiı
17. fa3	18. laı	19. daı4	20. pao2

Part (B): 2-syllable words. Please listen to each word, then write down the tone numbers or tone marks on your answer sheet. Each word will be spoken three times. Please listen carefully.

1. faxianl4	2. pingfang22	3. youbang31	4. huamaolı
5. dianji4l	6. honghaı23	7. diyu44	8. naipıng32
9. daoda42	10. malu34	11. xınqıngı2	12. tıkuzl
13. pıdaı24	14. moxıe43	15. naızuı23	16. xiangshuiı3

Part (C): Short sentences. You will hear an English sentence first, followed by four choices in Chinese. Only one Chinese sentence has the correct tones. Please choose the correct one by circling the letter a, b, c, or d. For example: How are you? (a) Nıl hao ma? (b) Nı2 hao ma? (c) Nı3 hao ma? (d) Nı4 hao ma? The correct answer is (b), so you should circle the letter b.

1. How is he? Ta <u>hao</u> ma?	(a) 2 (b) 4 (c) 1 (d) 3
2. He is fine. Ta hen <u>hao</u>.	(a) 1 (b) 4 (c) 3 (d) 2
3. Are you busy? Ni <u>mang</u> ma?	(a) 4 (b) 3 (c) 2 (d) 1
4. I'm not busy. Wo bu <u>mang</u>.	(a) 3 (b) 2 (c) 1 (d) 4
5. They're all busy.Tamen <u>dou</u> mang.	(a) 4 (b) 1 (c) 2 (d) 3
6. We are also busy.Women <u>ye</u> mang.	(a) 2 (b) 4 (c) 1 (d) 3
7. What is this? <u>Zhe</u> shi shenme?	(a) 3 (b) 4 (c) 1 (d) 2
8. Who is he? <u>Ta</u> shi sheı?	(a) 4 (b) 2 (c) 1 (d) 3

Pronunciation

There are three parts in this test: 1-syllable words, 2-syllable words, and short sentences. For each part, first you will hear the words or the sentences on the tape. Please repeat each word or sentence right after you hear it. Then, please read the word or sentence from the following list after you hear the number.

(A) 1-syllable words

1. mǎ	2. fù	3. fēi	4. pí	5. tà
6. dāi	7. nǐ	8. gǎi	9. dì	10. lǎo

(B) 2-syllable words

1. Měiguó	2. mǐlù	3. Húang hé	4. dà hǎi
5. lǎoshī	6. gāoyào	7. dìtú	8. páomǎ
9. gōngpíng	10. Huáng hǎi	11. hēi māo	12. pǎobù
13. wài yī	14. nígū	15. fābiǎo	16. lùdì

(C) short sentences

1. Tā hǎo ma? (How is he?)
2. Tā hén hǎo. (He is fine.)
3. Tā yě máng. (He is also busy.)
4. Nǐ máng ma? (Are you busy?)
5. Nà shì shū. (That is a book.)
6. Zhè shì bào. (This is newspaper.)
7. Tā shì shéi? (Who is he?)
8. Shéi shì lǎoshī? (Who is a teacher?)

APPENDIX B

Perception Scores (% correct)

Subject	Sex	L2	Test 1			Test 2		
			(A)	(B)	(C)	(A)	(B)	(C)
1	M	Y	90	75.00	62.5	*	*	*
2	M	Y	95	68.75	62.5	90	96.88	75.0
3	F	Y	95	75.00	37.5	90	96.88	100.0
4	F	Y	100	96.80	62.5	95	100.00	100.0
5	F	Y	30	25.00	37.5	65	40.63	62.5
6	F	Y	80	43.75	25.0	100	100.00	75.0
7	F	Y	*	*	37.5	100	93.75	87.5
8	M	Y	100	78.13	87.5	100	100.00	100.0
9	F	N	60	50.00	37.5	85	78.13	87.5
10	M	N	95	87.50	62.5	85	100.00	100.0
11	F	Y	90	62.50	62.5	85	100.00	75.0
12	M	Y	70	65.63	50.0	100	78.13	87.5
13	M	N	60	46.88	37.5	95	81.25	87.5
14	F	M	*	62.50	25.0	90	84.38	87.5
15	M	M	90	90.63	50.0	95	100.00	100.0
16	M	M	85	40.63	25.0	100	75.00	87.5
17	M	M	85	68.75	75.0	*	*	*
18	M	Y	95	65.63	75.0	100	96.88	75.0
19	F	Y	85	65.63	37.5	85	65.63	75.0
20	F	Y	90	62.50	37.5	95	90.63	75.0
21	M	Y	100	81.25	87.5	100	96.88	87.5
22	F	Y	55	37.50	37.5	*	*	*
23	M	Y	100	78.13	62.5	95	87.50	100.0
24	F	Y	100	68.75	50.0	95	100.00	100.0
25	F	Y	100	68.75	62.5	95	93.75	100.0
26	M	Y	90	62.50	12.5	95	90.63	87.5
27	M	Y	90	59.38	62.5	*	*	*
28	M	N	90	56.25	50.0	90	84.38	75.0
29	F	M	85	59.38	37.5	95	71.88	75.0
30	F	M	80	62.50	37.5	90	78.13	50.0
31	F	M	95	75.00	75.0	95	93.75	87.5

N=No L2 experience
Y=Other L2 experience
M=Mandarin experience

Reading Scores (% correct)

	Test 1			Test 2		
Subject	(A)	(B)	(C)	(A)	(B)	(C)
1	91.67	92.71	93.33	*	*	*
2	100.00	75.00	94.67	91.67	84.38	80.00
3	100.00	79.17	88.00	91.67	81.25	77.33
4	87.50	93.75	100.00	91.67	95.83	90.67
5	79.17	53.13	77.33	54.17	61.46	76.00
6	100.00	72.92	77.33	95.83	67.71	85.33
7	100.00	81.25	84.00	100.00	82.29	80.00
8	91.67	90.63	94.67	95.83	97.92	89.33
9	75.00	59.38	72.00	79.17	56.25	76.00
10	91.67	100.00	94.67	87.50	92.71	94.67
11	91.67	69.79	66.67	100.00	87.50	56.00
12	95.83	40.63	53.33	79.17	62.50	48.00
13	91.67	58.33	65.33	75.00	60.42	62.67
14	91.67	63.54	70.67	95.83	63.54	77.33
15	100.00	78.13	76.00	87.50	78.13	89.33
16	100.00	71.88	60.00	100.00	64.58	58.67
17	87.50	84.38	96.00	*	*	*
18	100.00	86.46	86.67	95.83	88.54	86.67
19	95.83	65.63	61.33	70.83	61.46	57.33
20	100.00	67.71	81.33	100.00	67.71	93.33
21	100.00	85.42	73.33	91.67	*	*
22	75.00	62.50	61.33	*	*	*
23	100.00	65.63	64.00	100.00	63.54	77.33
24	95.83	83.33	94.67	79.17	86.46	97.33
25	100.00	85.42	88.00	100.00	87.50	93.33
26	79.17	81.25	66.67	75.00	73.96	80.00
27	91.67	44.79	68.00	*	*	*
28	95.83	65.63	69.33	91.67	52.08	85.33
29	91.67	62.50	60.00	87.50	62.50	74.67
30	87.50	60.42	49.33	*	*	*
31	100.00	60.42	69.33	*	*	*

Imitation Scores (% correct)

	Test 1			Test 2		
Subject	(A)	(B)	(C)	(A)	(B)	(C)
1	95.83	93.75	96.00	*	*	*
2	95.83	94.79	100.00	100.00	96.88	98.67
3	100.00	94.79	97.33	95.83	89.58	94.67
4	100.00	97.92	100.00	100.00	100.00	100.00
5	91.67	97.92	97.33	62.50	97.92	96.00
6	100.00	97.92	96.00	91.67	86.46	93.33
7	100.00	97.92	100.00	100.00	96.88	94.67
8	100.00	98.96	96.00	100.00	100.00	98.67
9	95.83	98.96	97.33	87.50	100.00	98.67
10	100.00	100.00	100.00	100.00	97.92	96.00
11	100.00	97.92	96.00	100.00	97.92	90.67
12	100.00	95.83	98.67	100.00	97.92	93.33
13	100.00	94.79	96.00	100.00	96.88	96.00
14	100.00	94.79	88.00	100.00	98.96	88.00
15	100.00	100.00	96.00	100.00	97.92	97.33
16	100.00	98.96	97.33	100.00	97.92	88.00
17	91.67	95.83	93.33	*	*	*
18	100.00	96.88	93.33	100.00	98.96	96.00
19	100.00	89.58	92.00	100.00	95.83	97.33
20	100.00	96.88	100.00	100.00	96.88	98.67
21	100.00	95.83	94.67	100.00	100.00	*
22	100.00	89.58	93.33	*	*	*
23	91.67	95.83	90.67	100.00	92.71	94.67
24	100.00	100.00	96.00	100.00	98.96	100.00
25	95.83	100.00	100.00	100.00	100.00	97.33
26	100.00	98.96	81.33	91.67	95.83	94.67
27	100.00	96.88	96.00	*	*	*
28	100.00	98.96	93.33	100.00	98.96	96.00
29	100.00	95.83	93.33	100.00	97.92	82.67
30	100.00	97.92	96.00	*	*	*
31	95.83	98.96	97.33	*	*	*

MICHIGAN STATE UNIV. LIBRARIES
31293008897948