INSTITUTE OF HISTORY & PHILOLOGY

ACADEMIA SINICA

Special Publications No. 94

AN ACOUSTIC PHONETIC STUDY ON TONES IN MANDARIN CHINESE

Chiu-yu Tseng

Nankang, Taipei, Taiwan, ROC
1990

Forward

This monograph is based essentially on the work done for my doctoral dissertation under the same title "An Acoustic Phonetic Study on Tones in Mandarin Chinese". Needless to say, relevant theories of the field have gone through quantal changes. The main issues addressed in this work, i.e. acoustic measurements of tones in Mandarin Chinese, perceptual saliencies of the fundermental frequency patterns, and the interaction between lexical tone and intonation, especially in spontaneous fluent speech, would still offer some guidelines to research in Mandarin speech production and perception. What I see as impending and related issues include at least the following: (1) more detailed acoustic analysis on Mandarin intonation, read speech and spontaneous fluent speech alike; (2) more understanding and description on the role of stress patterns of Mandarin Chinese and (3) a global model of tones, stress and intonation for Mandarin. Some of the issues have already been investigated by me in subsequent studies and reported elsewhere. However, in the meantime, more recent works on intonation theories by other researchers such as Liberman and Pierrehumbert (1984) and Thorsen (1978, 1984) have also offered enlightening insights and would definitely be considered in my future pursuit of the study of Mandarin tones and intonation. I have, in later time, done some work on stress related features in Mandarin Chinese (see Tseng, forthcoming, 1989), and will continue to work in that area. Thus I would present this piece of work as it is, and leave the other issues in other possible forms of presentation.

During the initial processes of this study, I owe much gratitude to Professor Philip Lieberman, my dissertation director, whose genuine interest in a language that was entirely unfamiliar to him gave inspiration as well as encouragement to a then graduate student, and whose challenging ideas and never ending enthusiasm for basic research prove to be exemplary years later. John Mertus offered immeasurable help by making computer implemented analysis and tests possible. His famous chocolate chip cookies and freshly brewed coffee constituted the fondest part of my memories associated with my years at Brown University. My appreciation also goes to Professor Shiela Blumstien who as my thesis committee member spent much time commenting the written form of this work, and whose insights provided guidelines for subsequent works. I am also indebted to Professor James J. Wrenn, who offered much support over my years at Brown. The list of people I grew with during my eight years of stay in the U.S. could go on and on, and I thank my then lab mates Karen Landahl, Carol Chapin Ringo, Kathy Kubaska Ball, Jack Raylles, Pat Keating and many others for sharing that part of my experience with me. My subsequent affiliation with the Institute of History and Philology, Academia Sinica, Taipei, which has since been my home institute, supported my interest in speech related research in Mandarin and provided me the chance to establish a research lab from scratch. I would also like to express my thanks to the Institute for publishing this monograph. In addition, my family have also always been extremely supportive. I, of course, am solely responsible for errors and would most appreciate for response from readers.

Chiu-yu Tseng, 1989,Taipei

Table of Contents

Chapter I

INTRODUCTION

The major aim of the present project is an experimental study focusing on both the production and perception of tones in Mandarin and the general problem of tone and intonation. The target language of the study is Mandarin, the official language of China originated from the dialect of the Peking area, which is currently spoken by over 500 million people.

It is generally assumed that the principal phonetic feature of tone is found in the domain of pitch, and its primary acoustic correlate is fundamental frequency. Other phonetic parameters such as duration, amplitude, vowel quality, and manner of tone offset, etc., are also involved in the overall production as well as the perception of tones. However, in the present study, we try to limit ourselves to the study of duration, the fundamental frequency pattern and amplitude as phonetic features of tone. Acoustic phonetic analysis of the production of Mandarin tones as well as native speakers' perception of these tones are investigated. Our data consist of both citation forms and spontaneous running speech. An initial study of the interaction between tones and intonation is also included.

In Chapter 2, the role of duration regarding the production and perception of Mandarin tones is investigated. The duration of vowels with respect to tones is measured both from speech produced in citation forms as well as in spontaneous running speech. Experiments are designed to test the percption of tones regarding different duration conditions. In Chapter 3, experiments are designed to test the production of Mandarin tones in spontaneous fluent speech. Acoustic analysis of the data is made in order to test how in fact the phonetic output of actual speech corresponds to the prediction of phonetic output made by phonology. Chapter 4 is a study on the perception of Mandarin tones from spontaneous speech with all the interacting contextual information removed. Chaper 5 addresses the question of tones and intonation. Production data of Mandarin declarative sentences from spontaneous speech are used to test two recent intonation theories, namely, the breath–group theory (Lieberman, 1976) and the declination theory (Maeda, 1967; Pierrehumbert and Liberman, 1979; Sorensen and Cooper, 1980).

As a background for the present study, a brief summary of the phonetic inventory of Mandarin will be presented, following the work of Howie (1976). The Mandarin Phonetic Symbols II (MPS II), as well as its diacritics for tones, are employed throughout the study for transliteration. No special device will be used to indicate the tones. We shall use the terms level, rising, falling–rising and falling tones which are abbreviations for high–level, mid–rising, mid–falling–rising and high–falling tones respectively in the text as manifestation of the tones. More detailed description of the phonological predictions of tones will be presented in Chapter 3. Each chapter is written with its own introductory section for the convenience of the reader. Instrumentation and experimental techniques involved are described in the Methodology section of each chapter.

1.1 Phonological inventory of Mandarin

Following the discussion by Howie (1976:4–8), the phonological inventory of Mandarin can be summarized as follows:

1.1.1 Vowels

The vowel phonemes in Mandarin are presented in Table 1.1. For the inventory of phonemes with their pricipal allophones, see Appendix 1.1.

Table 1.1. Vowel phonemes in Mandarin

		front	central	back
high	rounded	y		u
	unrounded	i		ɯ
mid	rounded			ɔ
	unrounded	ɛ		ɤ
low	rounded			
	unrounded		a	

1.1.2 Consonants

Table 1.2 presents the consonants in mandarin.

Table 1.2. Consonats in Mandarin

place of articullation / manner of articulation			bilabial	labio–dental	dental–alveolur	alveolar	retroflex	palatal	velar
Stop	voiceless	unaspirated	P			t			k
		aspirated	p^h			t^h			k^h
Affricates	voiceless	unaspirated			ts		tʂ	tɕ	
		aspirated			ts^h		tʂ^h	tɕ^h	
fricatives	voiceless			f	s̺		ʂ	ɕ	X
	voiced			(V)			ʐ		
lateral	voiced								
nasal	voiced		m			n		(ȵ)	(ŋ)

1.1.3 Tones

Following Y.R. Chao, the tonal contours are presented schematically by "time–pitch" graphs attached to the left of a vertical reference line divided into four intervals by five points (Chao, 1968 : xxiv, note 4). Table 1.3 summarizes all the existing descriptions for Mandarin tones, including the Mandarin Phonetic Symbols II (MPS II).

Table 1.3. Tones in Mandarin

Sample makes	Tone	Chinese name	Description	Pitch	Graph
ma	1st Tone	yinping	high–level (level tone)	55:	˥
ma	2nd Tone	yangping	high–rising (rising tone)	35:	˧˥
ma	3rd Tone	shang sheng	low–dipping (falling–rising tone)	214:	˨˩˦
ma	4th Tone	chiu sheng	high–falling (falling tone)	51:	˥˩

Figures 1.1 – 1.4 are examples from the present study, showing the actual contour vowel [i] in these four lexical tones.

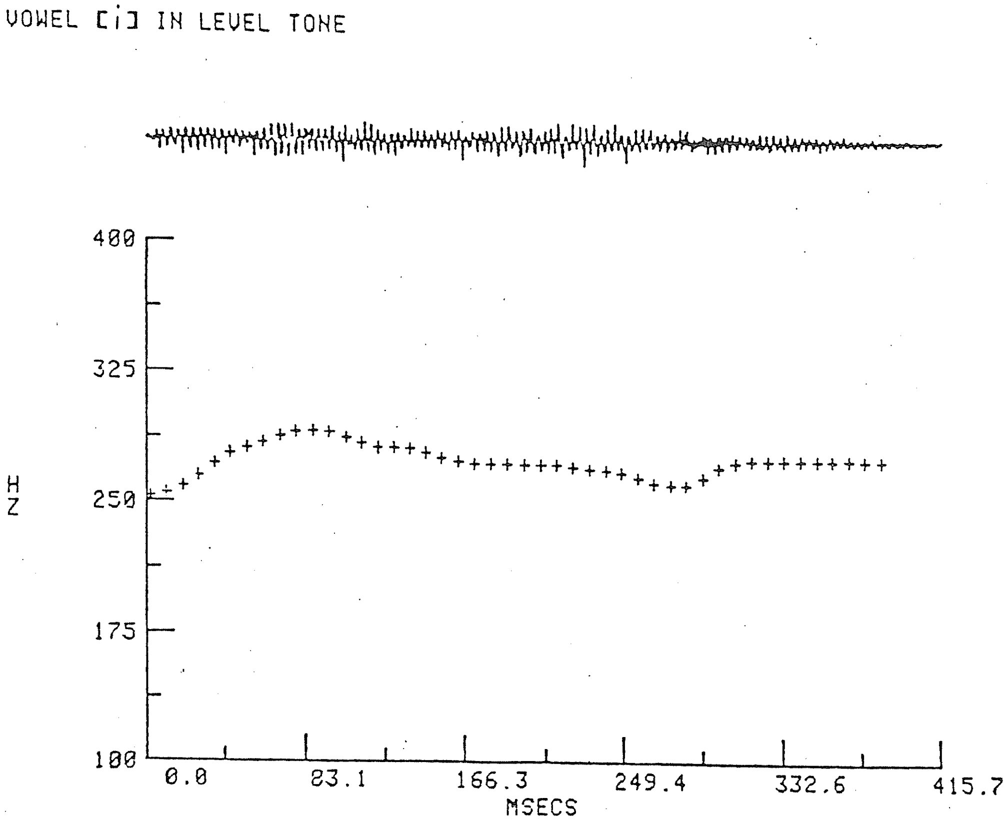


Figure 1.1. Fundamental frequency plotting of Mandarin vowel [i] in level tone, and the waveform of the vowel.

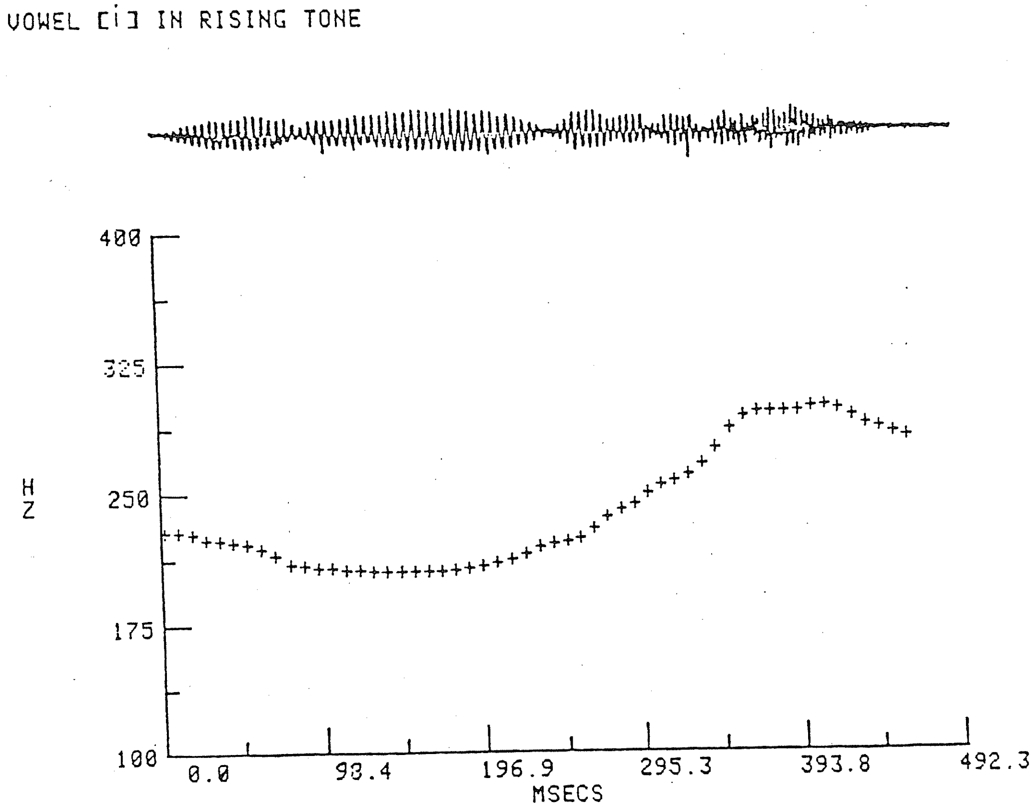


Figure 1.2. The waveform and the fundamental frequency contour of Mandarin vowel [i] in rising tone.

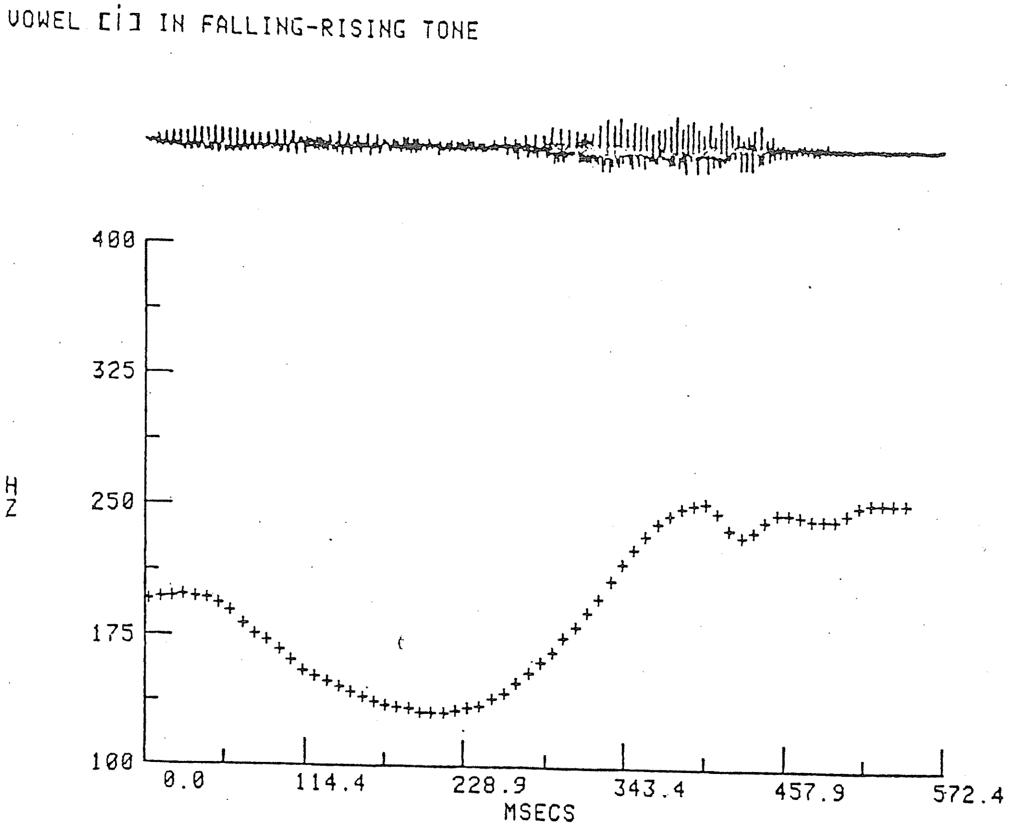


Figure 1.3. The waveform and the fundamental frequency contour of Mandarin vowel [i] in falling-rising tone.

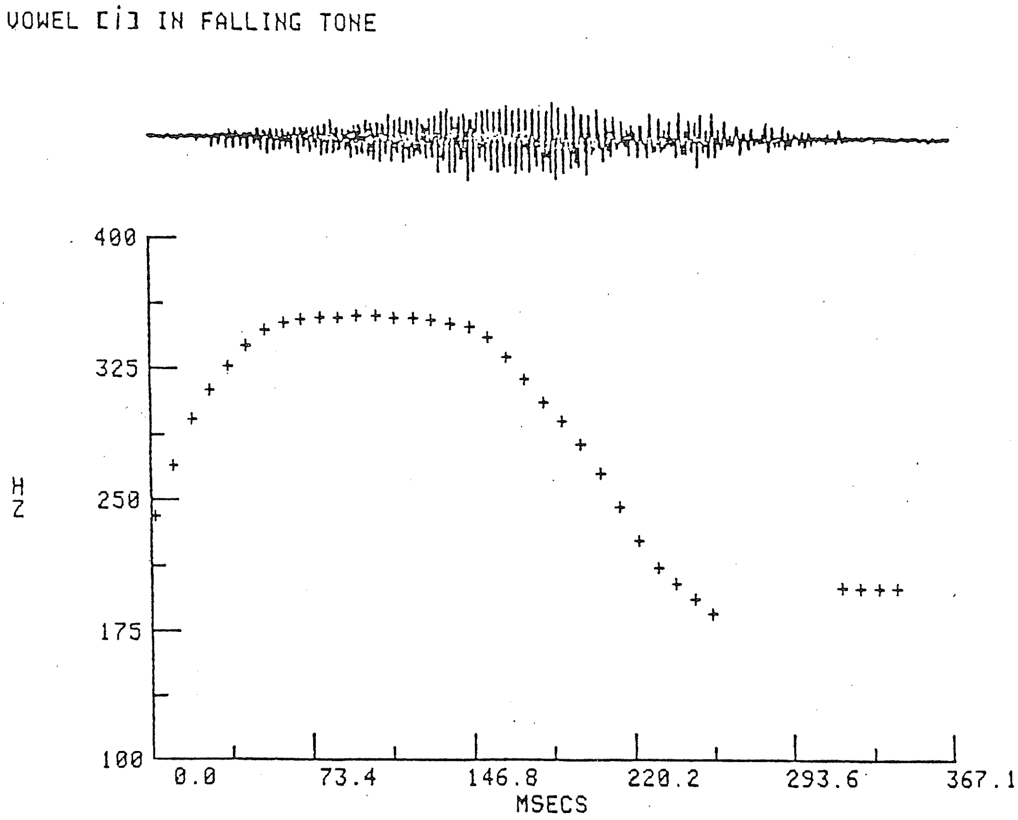


Figure 1.4. The waveform and the fundamental frequency contour of Mandarin vowel [i] in falling tone.

1.1.4 The syllables

The syllable in Mandarin (except for very rare dissyllabic morphemes) corresponds to the morpheme or to two morphemes when the morpheme /r/ [ɚ] er is suffixed to another morpheme. Homophonous morphemes are quite numerous in Mandarin because the constraints on the combination of segments in the syllable admit only around 418 different syllables (without the /r/ suffix). With the accidental gaps in the distribution of the four tones, fewer than 1200 tonally differentiated syllables occur as unit morphemes. The syllable is an important unit in the phonology of Mandarin because each syllable as a dictionary entry is assigned with a distinct tone (although this is not altogether true from a strictly phonetic viewpoint). The basic syllable structure of a syllable is (C)V with a tone assigned to it. The vowel portion of a syllable can be a nuclear vowel (or diphthong), or vowel with a preceding glide, or a vowel plus a dental nasal –n or a velar nasal –ng. No other consonants except the dental and velar nasals occur in the syllable–final position. And for this reason, Mandarin is often said to have open syllable only. The retroflex /r/ can be suffixed to a syllable to form a morphemically complex monosyllable (Chao, 1968: 19; Howie, 1976: 18). Detailed discussion of the syllable will not be included in the present study. Besides the vowel phonemes and their allophones (See Appendix 1.1), diphthongs include [ay, e, au, ou].

Thus in Mandarin, there are six vowel phonemes with their allophones, four diphthongs, and 19 consonants, as well as four lexical tones with the syllable as the unit for tone assignment.

Appendix 1.1. Inventory of Vowel Phonemes with their Principal Allophones

a. Vowel

/i/ [ʅ] (high central retroflex, unrounded, often with continued friction) shr [sʅ]

[ɿ] (high back, unrounded, often with continued friction) after /ts tsʰ s/ only, e.g., /si/ [s ɿ] sz

[ɪ] after /ə ɑ /, e.g., /ai/ [ɑɪ] ai

[ɪ] after / ə / on the 1st and 2nd tones when preceded by initial consonant and /u/, e.g. /xuəi/ [xᵘɪ] huei before /ən əŋ /, e.g., /iən/[ɪn] yin

[ⁱ] before all other occurrences of / ə ɑ /, e.g., /ia/ [ⁱa] ya

[i] elsewhere, e.g., /i/ [i] yi

/ü/ [ʊ̈] before / ə n/ when preceded by /l/, e.g., /lüən/ [lʊ̈n] luen

[ᵘ̈] before all other occurrences of /ə ɑ /, e.g., /üə/ [ᵘ̈ ɛ] yue

[ü] elsewhere, e.g., /ü/ [ü] yu

/u/ [ʊ] after /ə ɑ /, e.g., /iɑu/ [ⁱɑʊ] yau

[ʊ] after / ə / on the 1st and 2nd tones when preceded by /i/, e.g., /iəu/ [ⁱʊ] yóu, /xiəu/ [xⁱʊ] shioū before / ə ŋ / when preceded by an initial consonant, e.g., /xu ə ŋ / [x ʊŋ] hung before /ən/ on the 1st and 2nd tones when preceded by an initial consonant, e.g., /xuən/ [xʊn] huen

[ᵘ] before all other occurrences of /ə ɑ /, e.g., /uɑ/ [ᵘɑ] wa

[u] elsewhere, e.g., /u/ [u] wu

/ə/ zero after /i/ on the 1st and 2nd tones when followed by /u/, e.g., /i u/ [ⁱʊ] you, /xi u/ [xⁱ ʊ] shiou after /i/ when followed by /n ŋ /, e.g., /i n/ [ɪ n] yin after initial consonant and /u/ on the 1st and 2nd tones when followed by /i/, e.g., /xu i/ [xᵘ ɪ] hueī after initial consonant and /u/ on the 1st and 2nd tones when followed by /n/, e.g., /xuən/ [xʊn] huen after initial consonant and /u/ when followed by /ŋ/, e.g., /xu n/ [xuəŋ] hung after /lü/ when followed by /n/, e.g., /lüən/ [lʊ̈n] luen

[ö] after all other occurrences of /ü/ when followed by /n/, e.g., /üən/ [üön] yun

[o] after /ü/ when followed by /ŋ/, e.g., /üəŋ/ [uoŋ] yung before all other occurrences of /u/, e.g., /iəu/ [iŏ] you, /xiəu/ [xiŏ] shiou

[e] before all other occurrences of /i/, e.g., /u i/ [ueɪ] wei, /xuəi/ [xueɪ] huei

[ɔ] finally, after /u/, e.g., /u/ [u ɔ] wo

[ɛ] finally, after /i ü/, e.g., /i / [i ɛ] ye

[ɤʌ] finally, after a consonant, e.g., /s̀ / [s̀ɤʌ] she

[ɿʌ] elsewhere finally, e.g., /ə/ [ɿʌ] e

[ə] elsewhere, e.g., /uən/ [uən] wen, /xu n/ [xuən] huen

r/ [ə˞] as the sole segment of the syllable, e.g., /r/ [ə˞] er when suffixed to another syllable, see list under the heading vowels in the section on the distribution of phonemes

/a/ [æ] after /i ü/ when followed by /n/, e.g., /ian/ [ɪ æ n] yan

[a] before all other occurrences of /n/, e.g., /an/ [an] an

[ɑ] elsewhere, e.g., /xɑ / [x ɑ] ha

Chaper II

ACOUSTIC CORRELATES OF TONE

2.1 Production measurement of Mandarin vowels

2.1.1 Objectives

The objectives of this part of the study are to investigate the following questions:

(1) is the fundamental frequency contour the primary acoustic correlate for the lexical tones of Mandarin Chinese?

(2) is the fundamental frequency contour the primary cue for the perception of the lexcal tones of Mandarin Chinese?

(3) does the duration play a role in limiting the tones?

There appears to be a mixed usage of terminology in the literature of describing speech or speech–like sounds in acoustic and/or psycholingustic studies. "Read speech" has been used interchangeably with "citation forms", "natural speech" and "running speech" as contrasted with "synthetic speech"; whereas "running speech" and/or "natural speech" are sometimes used to mean "spontaneous speech". To clarify and prevent any possible confusion, we will use the terms "citation forms" to mean read, elicited speech; and "spontaneous speech" to mean free running speech (fluent speech).

2.1.2 Methodology

The experiments designed here are to obtain tokens from both citation forms and spontaneous speech for the purpose of duration studies. The subject COL is an adult female speaker of the standard Peking dialect. The citation forms were recorded in a sound treated booth at the Linguistics Lab at Brown University, using a Nagra type 42 tape recorder and a Sony electric condenser ECM 220 microphone. The spontaneous speech portion was obtained during classes on Modern Chinese Literature taught by the subject. A Sony TC–800B reel–to–reel tape recorder and a ECM–16 wireless microphone were used for recording.

2.1.2.1 Recorded Mandarin vowels in all 4 tones in citation forms

The subject was required to read a list of the 6 Mandarin vowels–phonemes so that in all, there were 24 tokens (Table 2.1).

Table 2.1. Mandarin vowel phonemes in 4 lexical tones

vowel	level tone	rising tone	falling–rising tone	falling tone
/i/	衣	移	椅	意
/u/	迂	於	雨	玉
/u/	屋	吳	摀	誤
/ə/	痾	鵝	嘔	餓
/ɑ/	哈 [Xa]	嗄	Y ˇ [a]	啊
/r/	儿 [ə̌]	兒	耳	二

Notice that these six vowels presented are phonemes in the phonological inventory of Mandarin. Complete discussion of Mandarin phonology, especially regarding vowels and tones is available in Howie (1976, Acoustic Study of Mandarin Vowels and Tones, see also 1.1). Chinese characters were used to represent each vowel–tone combination in the table, and this was the way the list was presented to the subject. For these vowel–tone combinations that do not occur as morphemes in the phonological inventory, a character that contains such vowels and tones with an initial velar fricative [x] was then used. For example, the vowel [a] in level tone does not occur in the system, but the syllable [xa] 'ha–laughing sound' does. the morpheme ha 'Ha' was then used. The velar–fricative and vowel combination was used as the fricative does not contain any transition from the consonant to vowel and thus imposes no difference of duration for the vowel portion of the token. If the above described CV construction does not occur in Chinese characters, then the Chinese phonetic alphabet, the jùyīnfúhàu (注音符號) was used to transcribe the nonsense token. Among the 24 vowel phonemes, only token ha 'ha–laughing sound' was presented with a velar fricative, and two more tokens were presented in the Chinese Phonetic alphabet. These two token are vowel – a – in falling–rising tone, and retroflex vowel –r– in level tone. Thus there were two nonsense tokens (8.33%) of the phonemes. However, as most native speakers of Mandarin are well aware of the existence of the four lexical tones, the subject did not produce these two nonsense syllables in any different manners. In other words, these two nonsense syllables did not cause the subject to shift the range of frequency regarding tones.

To avoid inconsistency in pitch range (Wang Li, 1955 : 28), the subject was asked to read each vowel first in level tone; then the level tone and the rising tone; the level tone, the rising tone, and the falling–rising tone; and finally all four tones.

This kind of reading results in the target vowel–tone combination occurring in breath–group final position. The subject then was given another list of the tokens in Table 2.1 in random order. Each vowel–tone combination appears three times and the list contains a total of 72 tokens. The subject was then asked to read every token as if the token was the last word of a sentence. Each token was read and followed by a pause before reading the next token. This kind of reading also resulted each token to occur at the breath–group final position. These two different readings, i.e., tokens read in sequence of the four tones and tokens from a radomized list, in fact resulted in very little variation regarding fundamental frequency range, contrary to what had suspected.

2.1.2.2 Recorded spontaneous speech

The experimenter, equipped with a Sony type TC8W–B reel–to–reel tape recorder and a wireless microphone clipped on the coller of the subject, sat in and recorded three sessions (six hours) of classes on Modern Chinese Literature taught by the subject. The settings were lecture/seminar, and only the spontaneous speech of the subject was transcribed for later experiments.

2.1.2.3 Inclusion of tokens

For the duration measurements of vowels in citation forms, tokens consisting of target vowel–tone combinations from the reading of Table 2.1 were used, as well as all tokens from the randomized list. In all, duration was measured for 96 tokens from the two different types of reading. For the measurements of spontaneous speech, only vowel [i] was used in comparison with the same vowel produced in citation forms. In Mandarin, the vowel [i] with all four tones are monosyllabic words. Thus the selection of this particular vowel with any of the four tones would result in real words in the language. Thirty–six utterances containing vowel [i] in both stressed and unstressed positions were selected for this study.

2.1.2.4 Analysis

In Howie's (1976) study of Mandarin vowels and tones, the average curves of each of the four tones were derived from the separate plots of thirty–four syllables from two male native speakers (Howie, 1976; 218–225). On the frequency scale, the level tone occurs at approximately 150 Hz. The rising tone starts at approximately 115 Hz and gradually rises to about 150 Hz. In other words, the actual rising occurs within a range of approximately 35 Hz. The falling–rising tone begins at approximately 113 Hz, moves downward to about 90 Hz and then rises back to about 113 Hz again. That is, the actual "dip" for the falling–rising tone occurs in a range of approximately 23 Hz. The falling tone begins at approximately 157 Hz and falls to about 105 Hz. That is, the drop of the fundamental frequency is around 52 Hz. The data for the present study, produced by a female native speaker, yields a somewhat diferent result. On the frequency scale, the level tone occurs approximately around the neighborhood of 215 Hz. The rising tone begins at about 150 Hz and rises to about 250 Hz, which means that the actual rising occurs within a range of approximately 100Hz. The falling–rising tone begins at approximately 135 Hz, moves downmard to about 95 Hz before moving upward to about 180 Hz. In other words, the "dip" for the falling–rising tone is around 40 Hz from the onset and subsequently moves up another 95 Hz before the offset. The falling tone begins at approximately 245 Hz and then drops to around 150 Hz, giving a drop of 95 Hz.

The above two studies illustrate fairly different results regarding Mandarin tones. First of all, we observe that our female subject for the present study has an overall higher pitch range as compared with Howie's male subjects. Moreover, our femal subject appears to have a wider range of tonal contour. Table 2.2 illustrates the comparison. The first difference is expected between male and female speakers. The second difference, nontheless, maybe attributed to individual styles of speech rather than the difference between sexes.

Table 2.2 Comparison of andarin tone production between Howie's (1976, hence H's) and the present (T's) studies

Tone	Results from	Fo pattern in N's			
		beginning Fo	Fo rise	Fo drop	ending Fo
levels	H's subjects	150	—	—	150
	T's subjects	215	—	—	215
rising	H's subjects	115	35	—	150
	T's subjects	150	100	—	250
falling–rising	H's subjects	113	40	40	113
	T's subjects	135	90	45	180
falling	H's subjects	157	—	52	105
	T's subjects	245	—	95	150

2.1.2.4.a Duration measurement of Mandarin vowels

All token of vowel [i], both in citation forms and in spontaneous speech were analyzed using a Kay Sono–Graph 6061B Sound Spectrograph with a scale magnifier that was set to a full scale display of approximately 4.8KHz. Both wide band and narrow band spectrograms were made for each token. Measurements of duration were then made according to the criteria described by Peterson and Lehiste (1960). All of the vowels in citation forms were analyzed using a PDP11–34 computer for both duration and tonal pattern analysis. Each token was analyzed by a pitch extracter program and the pitch profile for the duration of the syllable was plotted on a Tektronix storage scope. Hard copies of the graphic display from the storage scope of each token were then made. (For the description of the pitch extracter program, see 3.3; also Mertus, 1980.) Figure 2.1 is an example of the fundamental frequency plotting.

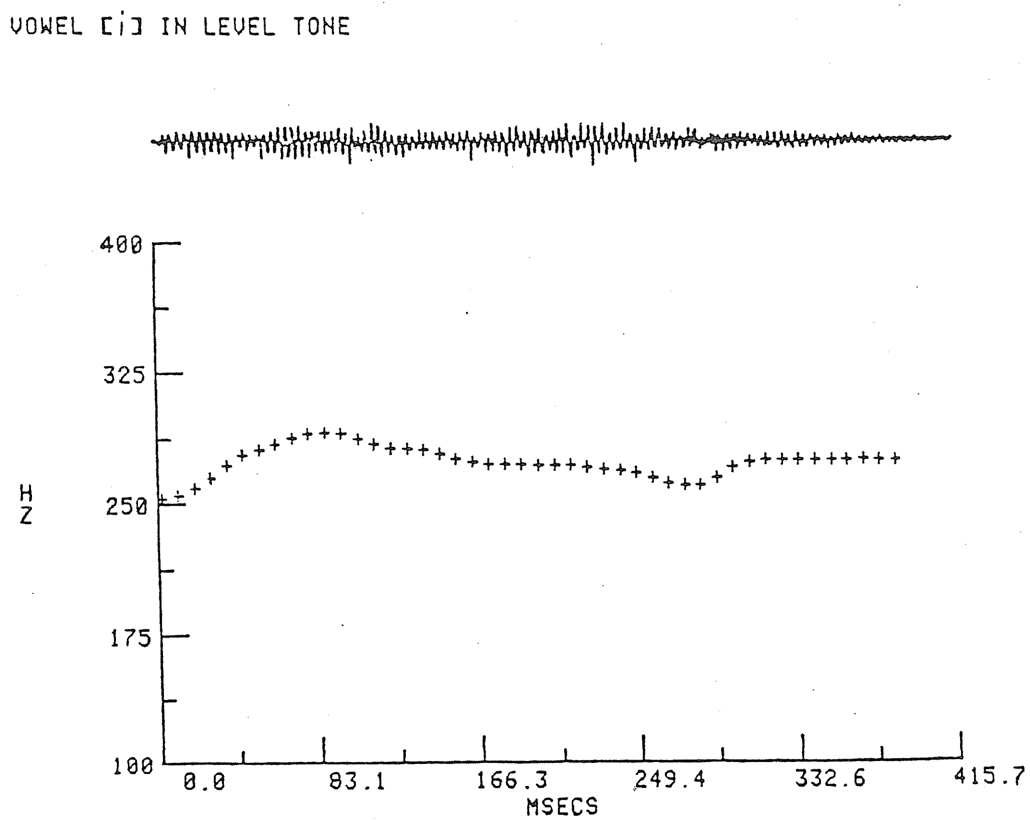


Figure 2.1. The waveform and the fundamental frequency contour of Mandarin vowel [i] in level tone.

The measurement of vowel duration in citation forms was based on the graphic display of the fundamental frequency contour as well as the amplitude. It was decided that anything below 100 unit energy frequency (See 3.3) was sufficiently weak to not consider it as part of the vowel. (Figure 2.2)

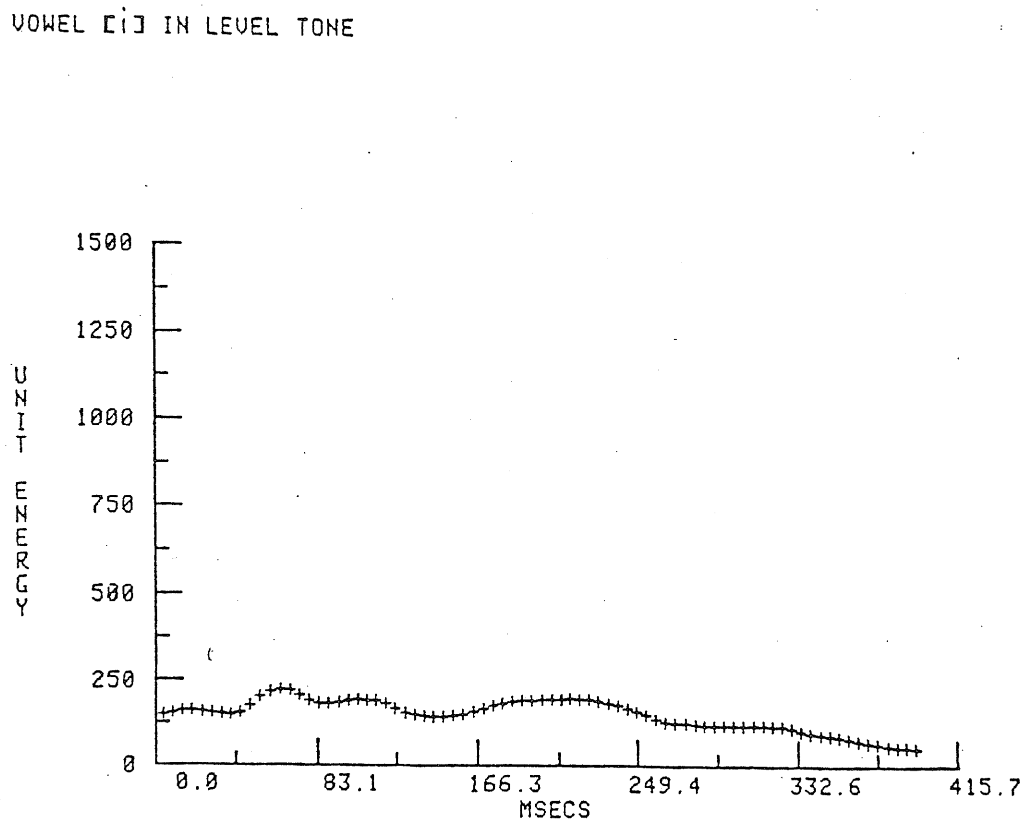


Figure 2.2. Unit energy measurement of Mandarin vowel in level tone as shown in Figure 2.1.

Using this amplitude information in relation to the extracted fundamental frequency, the total vowel duration was derived. (Figure 2.3)

VOWEL [i] IN LEVEL TONE

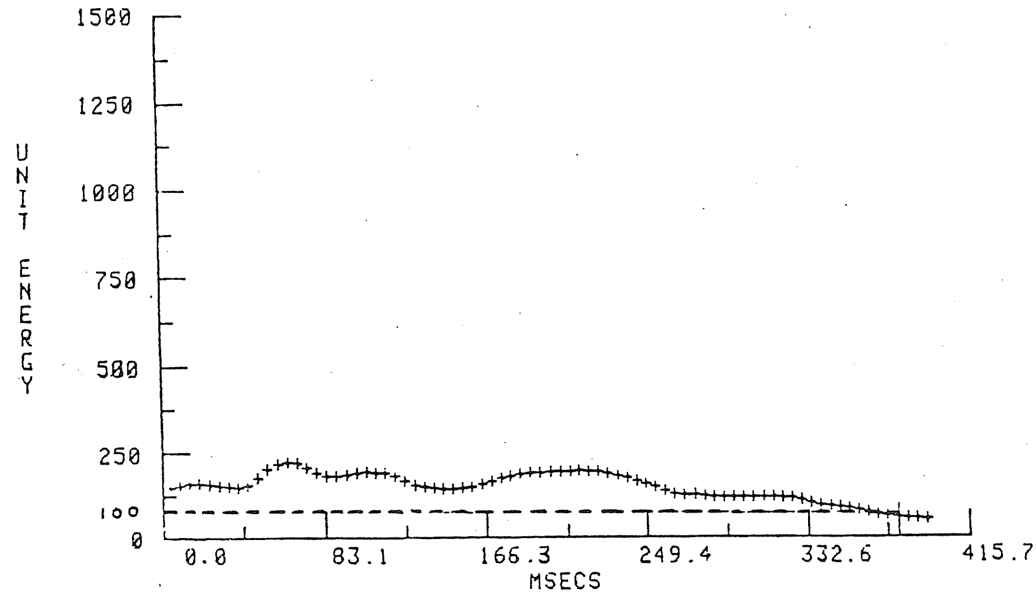


Figure 2.3. Unit energy measurement of Mandarin vowel [i] in level tone. Note that the dotted horizontal line indicates where the energy measurement for the token beings to fall below 100.

VOWEL [i] IN LEVEL TONE

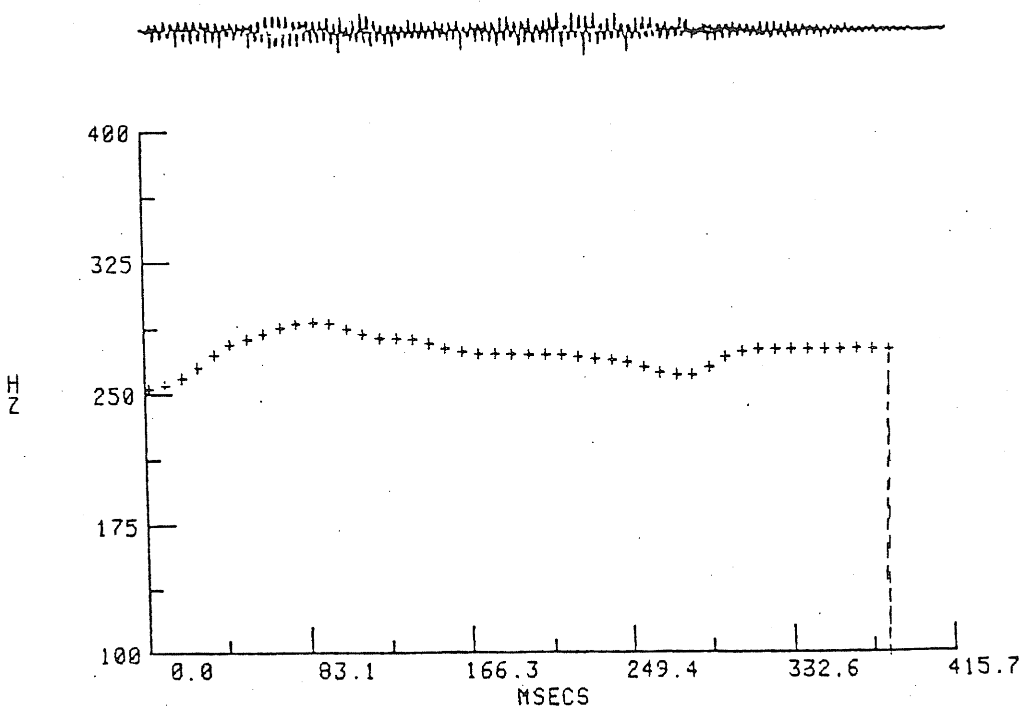


Figure 2.3.a. Fundamental frequency contour of Mandarin vowel [i] in level tone. Note that the horizontal dotted line indicates where the offset of the vowel is decided, which corresponds to the unit energy measurement made on Figure 2.3.

2.1.2.4.b Duration measurement of vowel [i] in spontaneous speech

Among the selected 36 Mandarin phrases/sentences, each one contained at least one occurrence of a Chinese word or morpheme with the [i] in one of the four lexical tones. We have selected such [i]'s in both morphemically–dertermined stressed and unstressed positions. (For more detailed discussion, see Tseng 1988, forthcoming.) For example sentences like: "tā huéiyì... 'He recalls...' where yì 'recall' is stressed, and "jège piényí... 'This inexpensive ...' where yí 'inexpensive' is unstressed. Both wide band and narrow band spectrograms were made for each utterance and the duration for each occurrence of vowel [i] was measured.

2.1.3 Results

2.1.3.1 Results of duration measurements of Mandarin vowel phonemes

The measurement of vowel duration with respect to different tones can be see in Table 2.3

Table 2.3. Duration Measurements of Mandarin vowel phonemes in citation forms from computer plotting of F0 pattern as well as amplitude in msec.

		level tone	rising tone	falling–rising tone	falling tone
vowel phonemes	i	364.1	471.5	544.2	328.05
	u	329.0	358.02	463.04	215.9
	u	288.7	363.1	452.19	268.72
	ə	332.65	356.04	418.04	193.89
	a	355.4	318.44	412.0	211.14
	r	440.7	390.6	454.38	199.1

2.1.3.2 Results of duration measurement of vowel [i] in citation forms vs. in spontaneous in Table 2.4:

Table2.4.

	in citation forms at breath–group final position		in spontaneous speech at non–breath–group final position	
	range	mean	range	mean
i	312.5 – 415.7 msec	364.1 msec	– – –	– – –
i	450 – 512.3 msec	417.15 msec	90 – 260 msec	175 msec
i	512 – 572.4 msec	542.2 msec	180 – 262 msec	221 msec
i	289 – 367.1 msec	328.05 msec	60 – 350 msec	205 msec

Among the 16 tokens of vowel [i] produced in citation forms embracing 4 occurrences of every vowel–tone combination, the range as mean duration was determined. for the 36 utterances of spontaneous speech containing words with [i], none of the tokens occured in breath–group final position. Also, there was no occurrence with vowel [i] in the level tone. Therefore it was only possible to select the tokens from spontaneous speech that occured at the non–greath–group final position. Again the range for such occurrences was found and the mean duration was then derived.

2.1.4 Observations and discussins

2.1.4.1 Intrinsic vowel duration in different tones

From the results of the production data of all Mandarin vowel phonemes in citation forms (Table 2.3), it is observed that within each vowel, different tonal patterns have different effects on the duration of the same vowel. The falling–rising tone, traditionally called the dipping tone, which has been phonologically described as a contour tone, maintains the longest duration. The falling tone has the shortest duration. No consistent pattern in terms of duration is found between the level tone and the rising tone. In short, the longest tone is the falling–rising, the shortest the falling, while level tone and rising tone fall randomly in the middle of the continuum. Out of the 6 vowel phonemes, the rising tone of 4 vowels (66.67%) have longer duration than the level tone; whereas 2 vowels (33.33%) illustrate the reverse pattern.. That is, level tone being longer than rising tone. These results also correspond with previous studies in the literature; namely the more complicated the tonal contour is, the longer the duration becomes, (Chao, 1968; Abramson, 1970). However, our results do not correspond to Howie's study (1976) in which the mean duration of the voiced part of the syllables, presumably the vowel portions, are: level tone, 237 msec; rising tone, 271 msec; falling–rising tone, 277 msec and falling tone 251 msec (Howie, 1976; 225). The hierarchy is that falling–rising tone is the longest, followed by the rising tone, then the falling tone, and finally the level tone. The data presented in this study illustrate that the mean duration of vowels are: level tone, 351.76 msec; rising tone, 376.28 msec; falling–rising tone 457.04 msec; and falling tone 236.12 msec. This result could be interpreted as the fact that all the tokens in citation forms for the present study were produced in the breath–group final position, whereas Howie's tokens were all derived from a carrier sentence from a non–breath–group final position. This suggests that the occurence of the target token regarding its position in an utterance has a definite effect on the duration. However, the falling–rising tone is the most complicated tonal pattern in Mandarin, and both studies support the fact that across all vowels, the falling–rising tone has a direct effect on vowel duration. Thus we conclude at this point that when produced in citation forms, the intrinsic vowel durations vary with respect to different tonal patterns.

2.1.4.2 Comparison of vowel duration in citation forms vs. spontaneous speech

The production data of Mandarin syllables produced in citation forms demonstrate that vowel durations vary according to the tonal patterns imposed on them. The question then is, are such durational differences maintained in spontaneous speech? The different results between the present study and Howie's study with respect to citation forms already suggests some variation. Table 2.4 gives the results of duration measurements of the same vowel, i.e., vowel [i], in citation form as contrasted to spontaneous speech. It shows that the duration difference in isolated production does occur when the same vowel with the same tonal pattern occurs in spontaneous speech. For vowel [i] in level tone, the duration in citation forms varries from 415.7 msec to 312.5 msec (m=364.1 msec). The range of duration is 103.2 msec. As indicated above, the subject produced no occurrence of vowel [i] in level tone in the available data to make any comparison. Nevertheless, it was possible to extract tokens of other tonal patterns. for vowel [i] in rising tone, the range of duration in citation form is 42.3 msec and 450 msec (m=471.15 msec). The range of duration for vowel [i] in falling–rising tone is 60.4 msec (between 572.4 msec and 512 msec m=542.2 msec), and the range for the same vowel in falling tone is 78.1 msec (between 367.1 msec and 289 msec m=328.05 msec).

The data for spontaneous speech containing vowel [i] with different tones produced at non–breath–group–final position offer a different result. Due to the fact that speakers do not produce every word as discrete units in actual spontaneous speech, the duration of words produced in actual running speech is expected to be different from when produced in citation forms. However, what is interesting is whether the hierarchy of the intrinsic duration of the four tones is still maintained in spontaneous speech and in addition, and whether the range of duration remains somewhat similar. Examining the production data of spontaneous speech yields the results already presented in Table 2.4. Although the falling–rising tone, the longest duration in isolated forms, still holds the longest duration in spontaneous speech (m=221 msec), the falling tone (m=205 msec) is longer than the rising tone (m=175 msec), which is the reverse from the results for the citation forms. Though the corpus of utterances is small and no tokens of level tone were included, the data still show that the shortest tone in connected speech is not the shortest tone on citation form. Moreover, if we further examine the range of duration under each condition, i.e., citation forms vs. spontaneous speech, no proportional relationship can be found. That is, the rising tone ranges from 90 msec to 260 msec, thus giving a differnce of 170 msec in spontaneous speech as contrasted to 42.3 msec in citation forms. The falling–rising tone ranges from 180 msec to 262 msec, giving a difference in duration of 83 msec in spontaneous speech as contrasted to the 60.4 msec difference in citation forms. The falling tone, surprisingly has the widest range in duration. Its duration ranges between 60 msec to 350 msec, giving a difference of duration in 290 msec as contrasted to the 78.1 msec difference in citation forms.

If we examine the data across conditions, we find that in actual running speech, the intrinsic difference of duration does not necessarily have to be maintained. Such results suggest that any tone can be produced with different duration depending on where a word occurs in an utterance. The fact that our results do not completely correspond to those of Howie's serves as a good illustration. It is expected that many factors are involved in the actual production of speech in verbal communication, such as whether a particular morpheme occurs in the stressed position vs. in the unstressed position (see 3.1.1.3), which syntactic property this particular morpheme bears, and the role the speaker's pragmatic knowledge plays during the production of a particular utterance, etc. To elaborate the involved factors is at this point beyond the scope of the present study, however we acknowledge such interacting factors and suggest that there must be a complicated interaction between different linguistic levels of representation, which results in the actual production of speech. We realize that the data presented here were produced by a single speaker and are also fully aware of the likehood of performance variations regarding different speakers. However, such study requires a larger body of data samples as well as speakers, and we suspect that more variations to vowel duration regarding tones might exist.

2.1.5 Summary and conclusion

The data reported here suggest that in the production of Mandarin vowels with different tonal patterns, the intrinsic vowel duration varies according to different tones and is most clearly demonstrated when the syllables are produced in citation forms at breath–group final position. The falling–rising tone appears to have the longest duration, the falling tone has the shortest duration; and level tone and rising tone do not demonstrate apparent distinctions. However, such an intrinsic difference of duration is not maintained in connected spontaneous speech. The results suggest that duration is not a primary phonetic parameter in the production of Mandarin tones.

2.2 Perception data of Mandarin vowels

The aim of the perception study here is to test the following:

(1) The primary cues for perceiving Mandarin tones, in particular, the role of vowel duration with respect to the fundamental frequency in tone perception.

(2) Whether there is any relationship between tone production and perception.

2.2.1 Methodology

2.2.1.1 Stimuli and test procedures

Vowel [i] in four Mandarin tones produced in citation forms from the production data were used as source of stimuli. Computer editing techniques were employed to edit the vowel from the end of phonation at various points. The tokens consisted of: (1) the vowel in each tone produced in full, including the fry portion toward the end of the vowel where no regular fundamental frequency period would be found; (2) the fry portion edited out with the exception of one case. This coincided with approximately 25% of the entire vowel length; (3) with 50% of the vowel edited from the offset (see above description); and (4) with 75% of the vowel edited out from the offset of phonation, leaving the initial 25% of the vowel only. Figure 2.4 illustrates the editing.

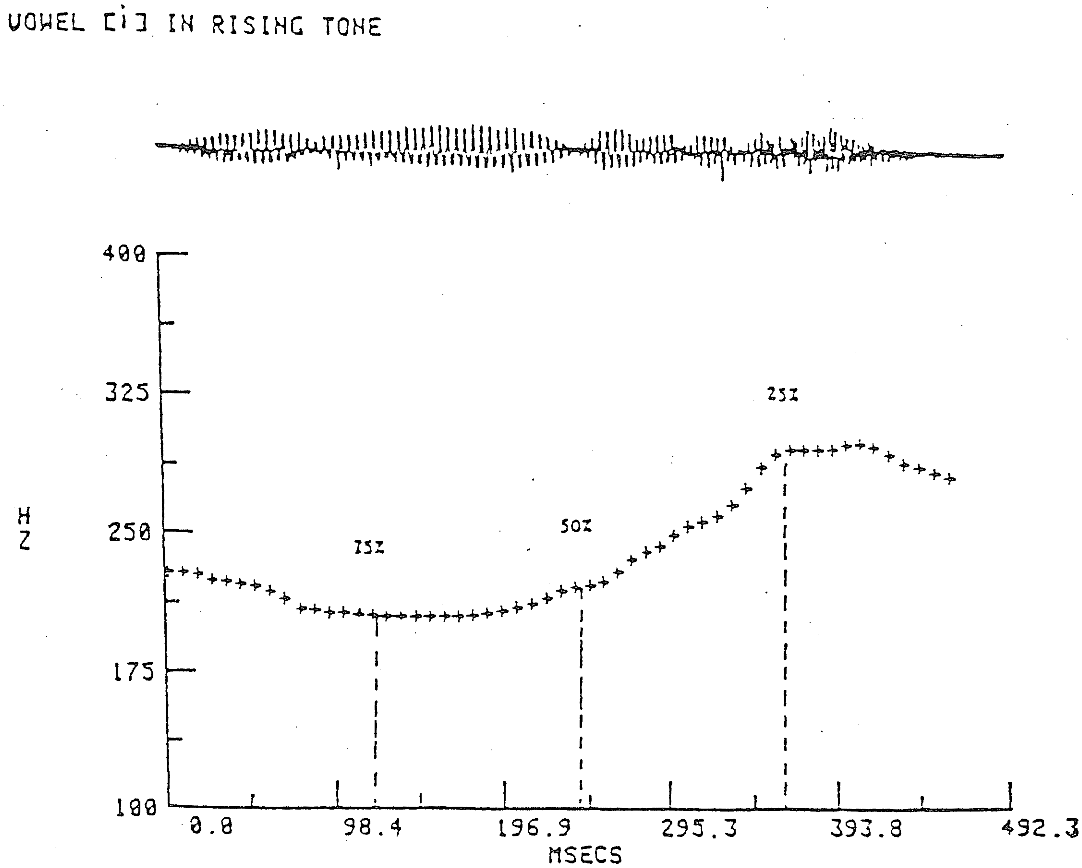


Figure 2.4. Mandarin vowel [i] in level tone, showing the points where the vowel is edited off from the offset of the vowel.

Although all the editing was done on zero crossing of the waveforms, a smoothing technique was still used for the last 20 msec of each edited stimulus to remove the abrupt ending which introduced a stop–like artifact (Mertus, 1980). Figure 2.4 is an example of the fundamental frequency of vowel [i] in rising tone showing where the editing was exercised.

Each token appeared four times in random order and altogether 68 tokens were generated and recorded on an audio tape. A 5–second interval was given between each token and subjects were asked to identify the tones of each token. The tape was presented to two female and two male native speakers of Mandarin Chinese through KOSS PRC/4AA headphones.

2.2.2 Results from perception test of vowel [i]

The performance of each subject was scored and the results from the perception test was tallied. Table 2.5 shows the overall results.

Table 2.5. Results of perception test of vowel [i] in four Mandarin tones

	whole				25% off			
Perceive tone test–tone	L	R	F–R	F	L	R	F–R	F
level	100%				100%			
rising		100%				100%		
falling–rising			100%				100%	
falling				100%				100%

	50% off				75% off			
Perceive tone test–tone	L	R	F–R	F	L	R	F–R	F
level	100%				100%			
rising		100%			100%			
falling–rising			100%		25%			75%
falling	100%				100%			

L = level
R = rising
F–R = falling–rising
F = falling

In general, the subjects behaved consistenly throughout the test, and demonstrated that they had no difficulty in identifying the tones produced in citation forms. For [i] in level tone, no mistake was made by any subjects across all conditions. For [i] in rising tone, all subjects perceived the shortest token, i.e., 75% off the vowel duration, as level tones. As for the rest of the conditions, i.e., at 50% and 25% off the vowel, all tokens were identified correctly. for [i] falling–rising tone, subjects had no difficulties except with the shortest condition, i.e. 75% off the vowel from the offset of phonation. This condition was perceived as either falling or level tones. for [i] in falling tone, subjects began to perceive the tokens from 50% off the duration as level tones. That is, the two shorter conditions of [i] were misidentified.

2.2.3 Discussion

From the fact that no listener failed to identify tones when produced in isolation, it clearly demonstrates that regardless of duration, native speakers of tone languages are able to identify the tones in real speech monosyllabic word. This result is consistent with previous perceptual studies on other languages, for example, Abramson's (1962:128, 1975:3–4) studies on Thai. It was also found by Abramson (1976a) that the confusion between register tones was virtually eliminated when the utterances of only one speaker was used for identification.

Since Mandarin Chinese does not have register tones, such phenomena need not be considered. However, to test whether Thai listeners could identify each of the tones on the basis of fundamental frequency alone, Abramson (1962:131–134) superimposed synthetic average fundamental contours on real–speech monosyllabic words, and 10 Thai subjects easily identified the tones with near–perfect accuracy. Abramson thus suggested that this finding provides strong evidence that fundamental frequency variation overrides any concomitant phonetic features such as duration or relative amplitude.

The duration study reported here, exhibits similar properties of tone. For the level tone, no subject made any error in identifying the tokens even when 75% of the vowel duration was sliced off, leaving only the initial 104 msec of a total 416 msec. In other words, all conditions for the level tone were correctly labeled. Because the fundamental frequency contour of the level tone is maintained regardless of duration, subjects perform perfectly under all duration conditions. Figure 2.5 illustrates as an example of this.

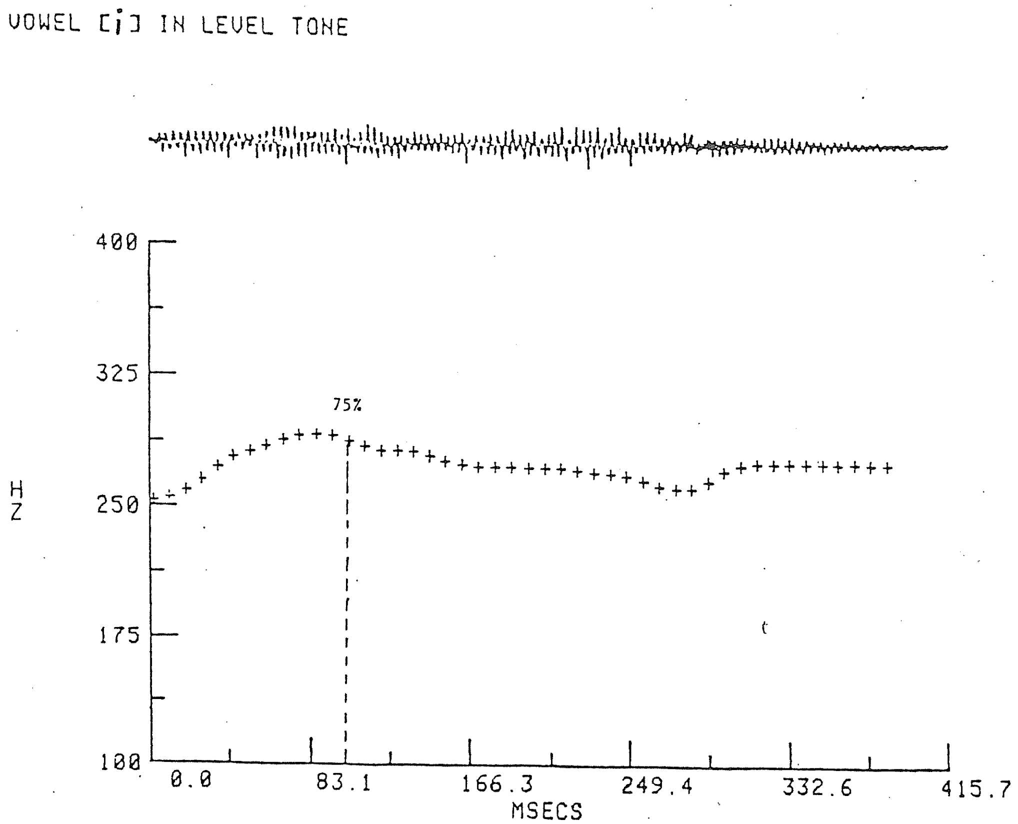


Figure 2.5. Fundamental frequency contour of Mandarin vowel [i] in level tone. Note the fundamental frequency contour when 75% from the offset of the vowel is edited off.

For the rising tone, all subjects perceived tokens at 75% off as level tones, and labeled the rest of the conditions correctly. The pitch plotting (Figure 2.6) show that fundamental frequency of a rising tone does not begin to rise until the middle of the vowel, and cutting off 75% of the vowel from the end leaves no trace of the required fundamental frequency pattern.

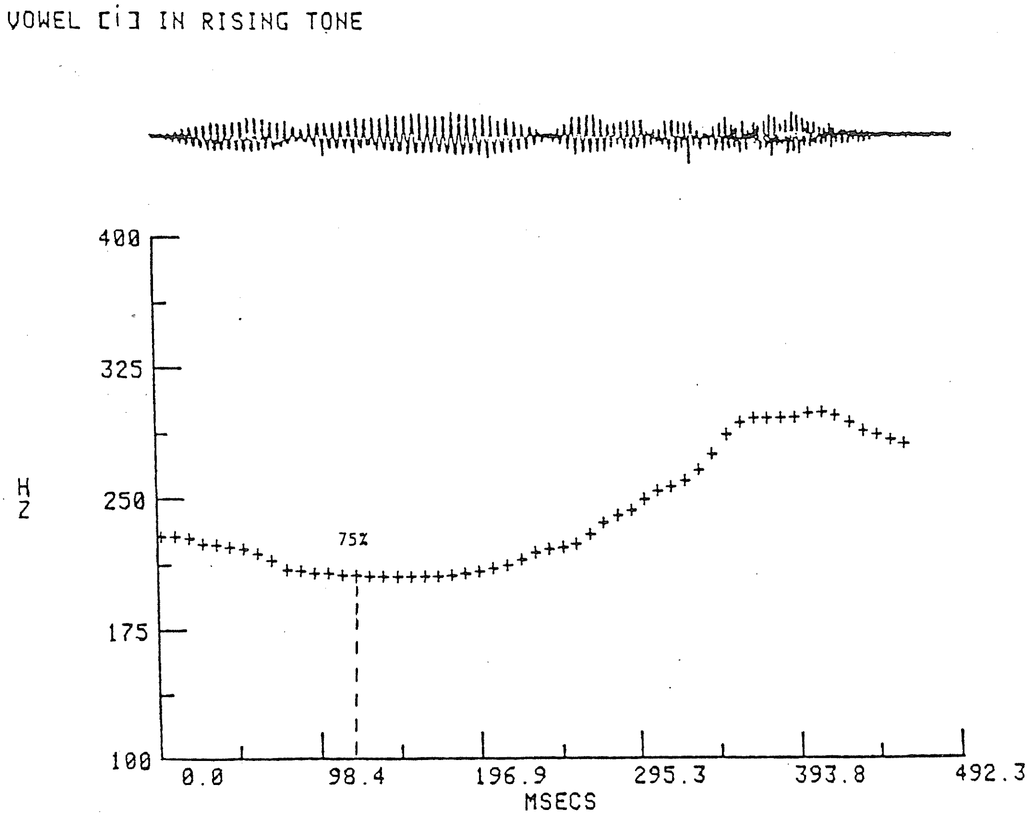


Figure 2.6. Fundamental frequency contour of Mandarin vowel [i] in rising tone. Note that when 75% of the vowel from the offset was edited off,the fundamental frequency contour does not contain any of the rising pattern.

For the falling–rising tone, subjects performed perfectly across the first three conditions, namely, the full vowel, 25% off the vowel, and 50% off the vowel respectively. As for the shortest condition, i.e., 75% off the vowel from the offset, one subject perceived it as the level tone, and three other subjects perceived it as the falling tone. Figure 2.7, i.e., the fundamental frequency plotting of vowel [i] in falling rising tone shows that at 75% off the vowel length, the display shape of the fundamental frequency was edited off, leaving no cues for the particular pitch contour.

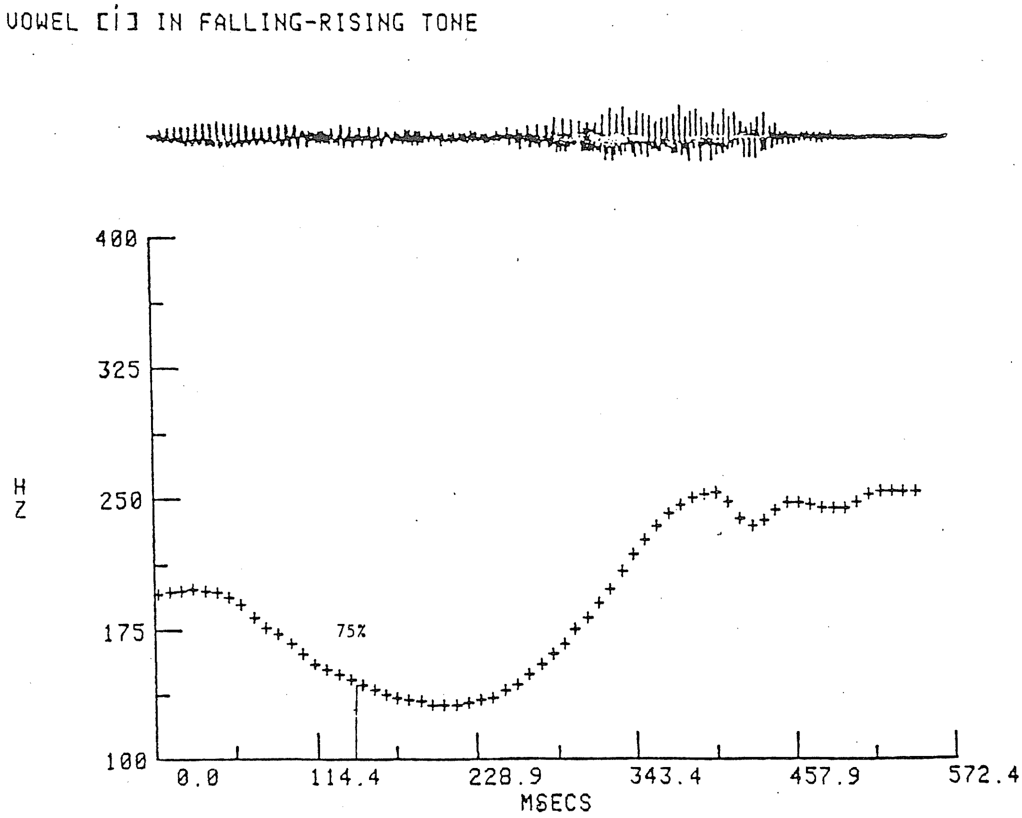


Figure 2.7. The fundamental frequency contour of Mandarin vowel [i] in falling-rising tone. Note that when 75% of the vowel is edited off from the offset, the fundamental fewquency controu does contain enough acoustic information of the falling-rising pattern.

To perceive the falling–rising at 75% off from the offset as falling tone indicates the subjects' awareness of the falling, i.e., the direction of the fundamental frequency pattern; whereas to perceive it as level tone suggests that the level tone might be the unmarked tone of the system, since it carries no information of direction of the fundamental frequency, nor a pitch change of register, but rather a constant shape across the entire tone. The Chinese Phonetic Alphabet system also handles the level tone as unmarked, since no tonal diacritic was assigned to represent the level tone.

For the falling tone, all subjects begin to perceive it at 50% off as level tones. These results again correspond to the fundamental frequency pattern of this particular token (Figure 2.8), since the fundamental frequency does not begin to fall until almost after 50% into the vowel length. Without the falling contour, subjects could only perceive the fundamental freqquency pattern as the level tone. This result, of course, is partially contributed by the fact that relative frequency height is not phonemic in Mandarin Chinese, and may not be applied to dialects which employ relative frequency height in tones (Massaro, Tseng & Cohen, 1983; Massaro, Cohen & Tseng, 1985; Tseng, Massaro & Cohen, 1986).

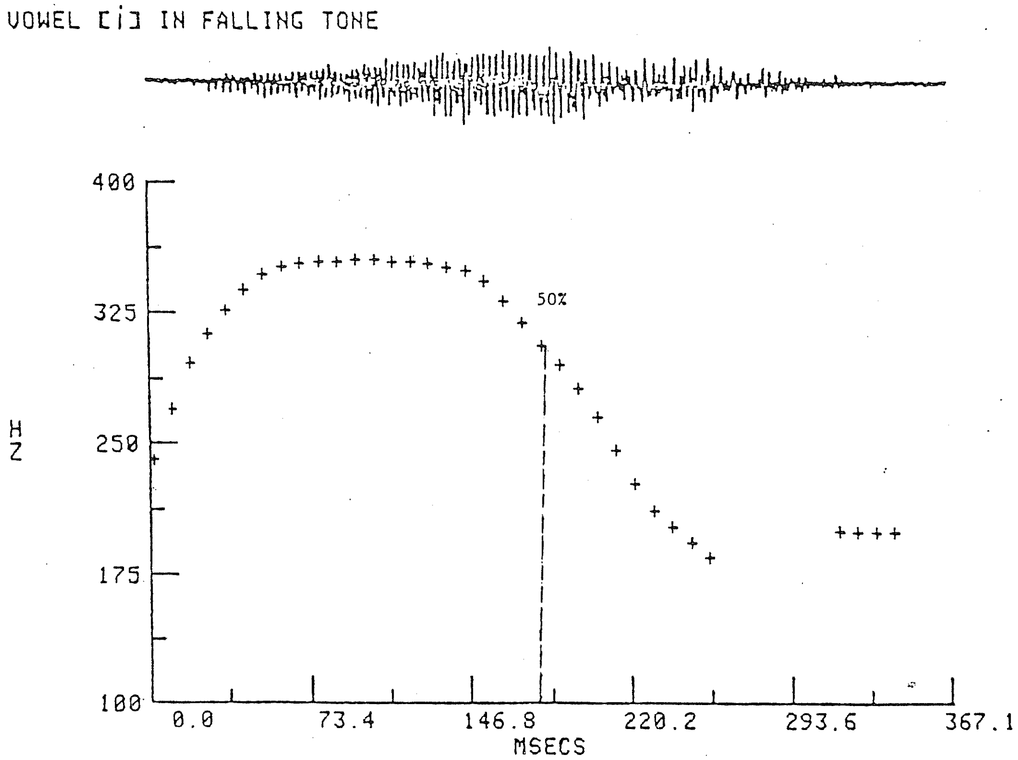


Figure 2.3. The fundamental frequency contour of Mandarin vowel [i] in falling tone. Note that when 50% of the vowel is edited off from the offset, the fundamental frequency contour hardly consists any falling of the vowel.

The uniform performance obtained suggests that again either the subjects are aware of the fundamental frequency pattern of the test tokens, or that the level tone is the unmarked tone of the system.

2.2.3.1 Fundamental frequency pattern as the primary cue for tone perception

The data on perception of tone discussed here are thus consistent with the hypothesis that in Mandarin Chinese the fundamental frequency pattern is the primary cue for the perception of tones. These data were consistent with data for other tone langauges such as Thai (Abramson, 1962, 1975). The perception study on vowel [i] also suggests that vowel duration is not crucial for tone perception. For tone identification, duration might be treated as a concomitant cue.

In the data reported, we also noticed that the subject had a tendency to produce a substantial portion of fry register toward the end of articulating each token in isolation (Figure 2.9).

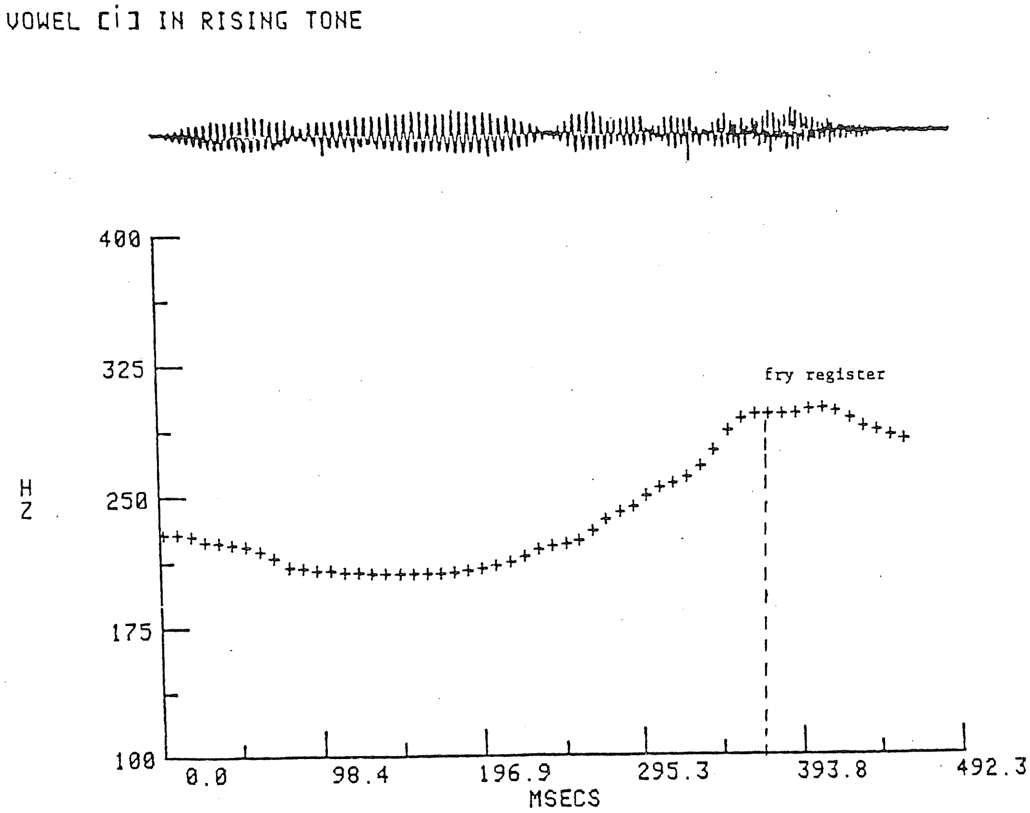


Figure 2.9. Fundamental frequency contour of Mandarin vowel [i] in rising tone. Note that the fry register is approximately 1/4 into the vowel from the offset.

The fry portion, which in most cases was approximately one fourth of the duration of the entire token, did not seem to affect the subjects' perception of tones. As in Figure 2.9 vowel [i] in rising tone demonstrates that the fry register was about 115 msec. The fundamental frequency of this protion clearly falls for about 20.83 Hz. However the perception of this token was not affected by the presence of the fry register at all. This result suggests that fry register at the end of a morpheme, in most cases the downdrift effect of the fundamental frequency, seems to carry across no significant tonal information for perception as long as the fundamental frequency pattern of that particular morpheme is present. Nonetheless, we realized that this could be a performance facter not necessarily present for all speakers, due to the previously mentioned fact that the data for this section came exclusively from one speaker.

2.2.4 Summary and conclusion

The aim of the perception study was in part to see whether vowel duration is a crucial cue in the perception of Mandarin tones, and also to determine the relationship perception might bear with production. By using computer implemented techniques, it was possible to edit portions of Mandarin Chinese morphemes, usually CV in structure, for spontaneous speech in order to test listeners' perception of such tokens. The results of the perception test leads to the following conclusion: First, native speakers of Mandarin, like native speakers of Thai (Abramson, 1963, 1975) are able to identify lexical tones in real speech monosyllabic words produced in isolation. Second, the duration study here shows that different durations do not affect the perception of tones. Mis–identification of the test tokens is due to the absence of the particular fundamental frequency contour, not purely due to the vowel length. We recognize that the presented perception study only includes vowel [i] produced by one speaker and no generalization is attempted for some speaker's production, duration <u>might</u> play a more important role. We suspect that a speaker might choose to use duration rather than the fundamental frequency pattern in producting tones. Therefore, there could be some trading relationship between the fundamental frequency and duration.

However, from previous studies of other tone languages, it appears to be safe to assume the although more complicated phenomenon might be expected with respect to other vowels in Mandarin, the fundamental frequency pattern plays the primary role in both tone production and perception. Moreover, they also suggest that duration does play a role, at least in citation forms. The production data presented in this chapter also demonstrate consistent results. For the time being, it seems to be quite clear that when monosyllabic words are produced in citation forms, the fundamental frequency contour is the primary acoustic correlate. The same fundamental frequency pattern is required for the perception of tones, too.

Chaper III

CHAPTER 3 PRODUCTION OF MANDARIN TONES IN SPONTANEOUS SPEECH

3.1 Introduction and objectives

In this chaper, production data are untilized to demonstrate the acoustic characteristics of tones. The main issue in this part of the study is an endeavor to see how native Chinese actually produce tones and intonation in real life communication. However, the focus here is on natural spontaeous speech (fluent speech), rather than formally read tokens from a list recorded in a sound treated room, as we have discussed in Chapter 2. Chiba (1935:66) discussed word tones (lexical tones), as well as sentence intonatiion and tones, on the basis of acoustic data (See discussion, 3.2.4.1). His work appears to be one of the earliest instrumental studies designed to invêstigate the acoustic nature of tones and intonation. Most other linguistists, acoustic phoneticians in particular, seem to be only concerned with lexical tones appearing in isolation (Lehiste, 1970; Abrambsom, 1975). When phonologists discuss about tones, their concern is often with lexical tones that function to distinguish meaning in a language, and thus they deal with lexical tones on the word level only. It is of course not their task to test either instrumentally or psychophysically the actual physical properties of tones, thus their judgement of lexical tones and all kinds of sandhi rules naturally falls on intuition and/or impression of those phenomena. (Leban, 1973, Goldsmith, 1975, 1976a, 1976b; Larson 1971; Schacter, 1969; Hsieh, 1976b; Fromkin 1972; McCawley, 1973a Yip, 1980).

To my mind, in the domain of linguistic investigation, if one wishes to investigate how speech is articulated and perceived, one would certainly wish to look at speech not only in elicited forms, but also in real–life situations. In other words, one would be interested to find out how the speech code really functions in verbal communication. Later studies on the production and perception of fluent speech (Cole ed., 1980) show that lingustists and others have begun research directed toward this area. Unfortunately, no tonal language was studied in this latest attempt.

The aim of this chapter is precisely such an endeavor, i.e., the actual production of tones in spontaneous speech. It is of course understood that when dealing with actual spontaneous speech, many factors are involved, namely, lingustic and/or extra–lingustic. However, the domain of the present study only covers the possible lingustic factors, with emphasis on what instrumental studies can offer. We will compare the data with observations and findings from the previous chapter, i.e., results from read speech in isolated forms. Production data of Mandarin will be used to show how spontaneous speech is produced differently from speech produced in citation forms. We would like to find out: (1) how much acoustic information is maintained in spontaneous speech regarding tones, (2) what are the possible criteria on which lexical tones are reduced? (For more detailed discussion of stress related tone reduction, see Tseng (1988, forthcoming).)

In Mandarin, theoretically each lexical entry in the dictionary has been assigned tones. Certain phonological and morphophonemic rules would predict tone reduction up to the phrase level. That is, the interaction among sequences of tones is phonologically predictable at both the word and phrase levels. In short, the phonology predicts the reduction of lexical tones under certain specified circumstanes. The issue here is whether it is really the case that acoustically these tones are reduced as they are phonologically specified? The present study attempts to investigate whether a one–to–one correlation of the acoustic information obtained from spontaneous speech and the phonological predictions can be found. In other words, when phonological rules predict the preservation as well as the reduction of tones, are these tones consistently produced accordingly at the acoustic level?

Before presenting the production data to test the above–raised questions, it is necessary to discuss the related theoretical implications regarding the interaction of tones on the phonological level as a background for later experiments.

3.1.1 Mandarin phonology: tonally ralated phenomena

Studies of English (Crystal, 1968) have shown that in an utterance, usually the content words, namely, nouns, verbs, adjectives, adverbs, receive more stress and are more clearly produced than function words, i.e., prepositions, articles, etc. That is to say, it has been stated (Cole, 1980) that in fluent speech, not every word is equally clearly produced. What should we or do we expect from tones? Is the same true for Mandarin Chinese in the area of the production of tones as well?

In order to follow the later discussion of our data on Mandarin Chinese, some background knowledge of the phonology would be essential. As mentioned in Chaper 1, the phonetic inventory of Mandarin consists of 19 consonants, 6 vowel phonemes and their allophones, 4 diphthongs (ay, e, au, ou), and 4 lexical tones. (See Chapter 1.2)

3.1.1.1 Lexical tones

A brief description of some important tonally related phenomena will be presented. We shall begin by repeating the phonemic pitch patterns in Mandarin. Recall that there are four distinct phonemic tones in mandarin, as was well documented and verfied in the literature, i.e.

level	(1st)	55	:	mā	'mother'
rising	(2nd)	35	:	má	'kemp'
falling–rising (mid–falling–rising)	(3rd)	214	:	mǎ	'horse'
falling	(4th)	41	:	mà	'to scold'

we have presented here three different notations for characterizing the four lexical tones. (For a more detailed description of the lexical tones, see Table 1.3 located in Chapter 1.) The first type is by verbally describing the shape of the contour, namely level, rising, falling–rising and falling. The second type is by the Chinese numerical covention, i.e., of the first, second, third and fourth tone. The third type is Chao's (1968, 25) well–know 4–equal–interval, 5–point scale system. That is, one meaning low, two half–low, three middle, four half–high, and five high. Also recall that a syllable in Mandarin is the bassic unit for tone (See 1.2.4.). That is, as dictionary entries, almost all syllables are assigned a phonemic tonal pattern, as Chiba (1935 : 64) noted: "In Chinese, on the other hand, each word has a stereotyped rise and fall."

3.1.1.2 Neutral tones

Although those linguists who work with Mandarin are familiar with the four kinds of distinct tones, they also describe the existence of a few toneless morphemes and the important role they play within the system. As Chao (1968 : 38) pointed out, a small group of morphemes never show up with any of these four lexical tones, and are always called "neutral tone". The neutral tone has been described as "toneless" in Chinese phonology for the following reasons: All such morphemes are clitic–like suffixes or particles which never occur in initial position of a phrase, and whose pitch is predictable phonetically from the tone of the preceding syllable.

The phenomena can be better illustrated if we utilize Chao's five–scale terminology. That is, given the lexical tone of the preceding syllable, a simplistic representation the neutral tones are produced as follows:

Preceding morpheme		fallowing neutral tone morpheme
level	55	3
rising	35	3
falling–rising	214 — 21	4
falling	41	1

For example, the subordinate particle le can be attached to almost anythinng to make it a modifier, or an agent in the case of verbs. Thus in theory it could be added on to any of the above mentioned well–known examples to generate the following:

māde	'mother's'
máde	'hemp's'
mǎde	'horse's'
màde	'the one who is scolding'

Other examples include aspect marker le 'perfective', classifier measure, such as, ge, of any other classifier in the structure of 'a + classifier + of + noun', etc. Phonologically, it has been rejected that these morphemes be regarded as carrying a fifth tone of their own that undergo some complicated sandhi rules. The reason for rejecting such a proposal as implausible is mainly because there are a large number of morphemes that appear in the system with the following characteristics: these morphemes alternate between having either of the four phonemic tones or a neutral one (Cheng, 1973). For example:

<u>jāngluo</u> 'open–sieve (to set up a sieve with a stick with grain underneath as a device to catch birds) — to make arrangement, to arrange'

<u>luómiàn</u> 'sieve–flour — to sieve flour'

The morpheme luo 'sieve' in the phrase <u>jāngluo</u> 'to arrange' contains neutral tone. The same morpheme <u>luó</u> 'sieve' in the phrase <u>luó</u>mian 'to sieve flour' has the rising tone. To produce the morpheme <u>luó</u> in <u>jāngluo</u> 'to arrange' with the rising tone is acceptable, but somewhat not as Standard Mandarin. To produce the same morpheme <u>luó</u> 'sieve' in <u>luó</u>mian 'to sieve flour' with the neutral tone is, nonetheless, definitely unacceptable. In other words, the morpheme <u>luó</u> 'sieve' can have either a neutral tone as a noun or a rising tone as a verb depending on the syntactic category it is in. Similar examples can be found in English, only the difference is in stress rather that the tone. For instance, the word <u>in</u>crease has its stress on the first syllable <u>in</u>– when used as a noun, whereas when used as a verb, the stress then is on the second syllable –<u>crease</u>. There are, in fact, many morphemes in Chinese with this type of characteristics as illustrated by the morpheme <u>luó</u> 'sieve'. They may show up in one combination, i.e., a word or phrase, as one of the four tones; and in some other combinations as the neutral tone. If we accept the proposal that the neutral tone is the fifth tone the system, we must find constraints to specify the fact that the neutral tone can alternate between the fifth tone and any of the four other tones, but no such alternation is allowed among the other four tones. Otherwise, we see no reason why the other four tones can not alternate between each other, for instance, the level tone with any of the three other tones. If we postulate the neutral tone as a toneless morpheme, all there is to say would be that lexical tones can be neutralized to neutral tone under certain conditions. More detailed discussion of the phonological status of each tone in Mandarin is, at this point, outside the scope of the present study, and thus will not be addressed. The reason that we brought up the neutral tone is mainly because there seemed to have been a lot of discussion regarding the issue of neutral in the literature, but not much common ground achieved. For instance, Chiba's (1935) study on accent made no mention of the neutral tone, which in consequence has caused some unnecessary misunderstaning of the system. However, it is important to be aware of the fact that there is a fifth tone phonetically, which is toneless in the sense that acoustically it does not have a distinct fundamental frequency pattern of its own and is referred to as the neutral tone. If we bear in mind the existence of the toneless neutral tone, whose fundamental frequency contour is simply a low flat plateau (Figure 3.1) when forced to produced it in elicited forms without preceding or following syllables. We shall subsequently discuss some other important tonally relevant phonemena. Most importantly, how is it produced in connected speech? Is there any variation in read speech vs. in spontaneous running speech?

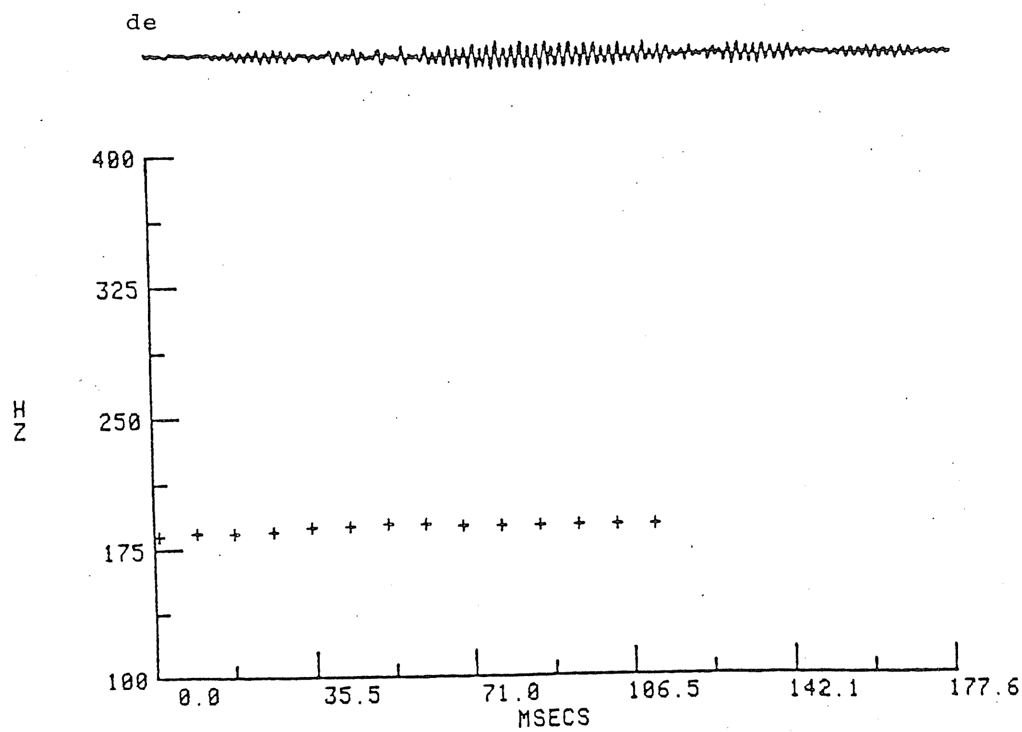


Figure 3.1. Fundamental frequency contour of netural tone morpheme de 'possessive marker' produced in citation form and in isolation.

A Mandarin word can be either monosyllabic, dissyllabic, trisyllabic, or quadrasyllabic. For example:

mà	'to scold'
jāngluo	'to arrang'
túshūguǎn	'library'
jūnghuámínguó	'Republic of China'

Except for the toneless morphemes, every syllable in Mandarin is 'born with' a phonemic tone. Such tones do not always have to be present when a morpheme is combined with other morpheme or morphemes to form a word that is more than one syllable in structure. In other words, tones can be deleted under certain circumstances. This phemomenon can be the result of either of the following two conditions: (1) tones and stress within words; (2) neutralization of tones across word boundaries.

3.1.1.3 Tones and stress

Except for morphemes which do not possess a distinct tone pattern and whose tone can be predicted according to the tones of the preceding syllable and are thus called 'toneless morphemes' or ' morphemes with neutral tone' (See 3.1.1.2) it is important to note that there are a great number of morphemes that appear in some words with one of the four tones, i.e., their 'born' tones; whereas in other cases, they apear in neutral tone. That is, instead of having a distinct fundamental frequency pattern intrinsic to that morpheme, the lexical tone of the unstressed morpheme becomes a neutral tone at first, the ultimate phonetic output of this reduced morpheme is then modulated by the tones of the preceding syllable. In order to understand such a phenomenon, it becomes necessary to describe, in a simplified manner, the word structure of Mandarin. As Chao (1968 : 35; 147) has explicitly described, Mandarin has final stress in both words and phrases consisting of full–toned morphemes. The secondary stress falls on the initial syllable, with the intervening syllables carrying lower levels of stress. For example:

jiéhūen	'to get married'
fànguǎn	'restaurant'
diànyǐngyuàn	'cinema, movie theatre'
jūnghuámínguó	'Republic of China'

However, there are also many words and phrases with initial stress, for instance:

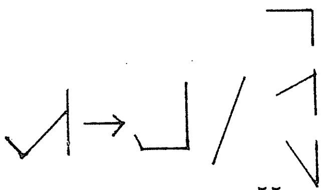
jāngluo	'to arrange'
shiuéshengmen	'students'
páchilaile	'to have come back on one's feet'

This group of examples differ from the preceding group not only in the position of the main stress, but in the flatting of the tones on the other syllables. In all such cases the non–initial syllables are completely unstressed, causing the tones to be reduced and to become what is traditionally called toneless. (For detailed phonological discussion via auto–segmental phonology, see Yip, 1980, 144–156). That is in Mandarin, whenever there is a word with more than one syllable (morpheme) in structure, it is in the lexicon that the word has been assigned initial stress, and subsequently the rest of syllables become either a secondary stress or neutral tone, i.e., toneless.

3.1.1.4 Neutralization (deletion) of tones

We shall also mention briefly another tonally related phenomenon here for the case of later discussion of our acoustic phonetic data: the variation or deletion of tones, better known as the phenomenon termed "tone sandhi" which in nature is neutralization of tones. This type of neutralization can occur across word boundaries. Tone sandhi in Mandarin includes the following:

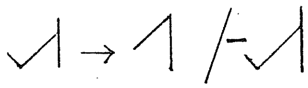
(1) The falling–rising tone lacks the final rise in non–final position and becomes a plain low. It can be represented as:



Using Chao's notation:

$$214 \longrightarrow 21 / \text{---} \begin{matrix} 55 \\ 35 \\ 51 \end{matrix}$$

(2) The falling–rising tone becomes a rising tone when it is before another falling–rising tone. That is:



or 214 —> 35 / — 214

for example:

mǎi mǎ —> mái mǎ 'buy a horse'

mái mǎ 'bury a horse'

The two phrases above are phonetically indistinguishable, as Wang Li (1963) and Wang, William (1967) noted. This sandhi phenomena is the best–known and most discussed in the literature, and we shall not particularly address it in the scope of the present study.

(3) The falling tone does not fall all the way through the tonal range when it is followed by another falling tone.

or 51 —> 53 / — 51

In short, the falling–rising tone has three allotones: the full range falling–rising (214) at phrase or sentence–final position, or rather, in citation forms as presented in Ch. 2; an allotone without the rising; and the rising tone. The falling tone has two allotones: the full range falling (51), and half–falling (53) when followed by another falling tone.

Moreover, pronouns as objects of verbs are unstressed and thus become reduced in tone; similarly, the reduplication of verbs also results in neutralization in tone. Both phenomena can be attributed to as deletion of tones due to the syntactic imformation superimposed on the lexical items involved. Of course we expect that the tone of voice and the speaker's intention to contrast or emphasize should also affect the actual production of the tones.

From the discussion above, it becomes quite clear that one cannot and should not deal with Mandarin Chinese as having four distinct phonemic tones only. Nor should one assume that the lexical tones of morphemes are always present. In Chiba's early work on accent (See 3.1.), there had been some discussion on the acoustic correlates of the four lexical tones in Mandarin. However, he did not mention the neutral tone morphemes, nor the interaction of neutral tones morphemes with the other four lexical tones. No mention was made about word structure and stress either. Instead, Chiba expected every morpheme's tone, by which he meant any one of the four tones only, to be physically present in actual speech. He further concluded that "the accent is by no means modified by the union of the syllables. (p.65)". Although he did recognize that syllables become squeezed together in "easy utterance," he unfortunately claimed that the fact that his subject's production which did not of course demonstrate every morpheme tone in connected speech, was due to 'dialectal variety'. In fact, the neutral tone, as Chao (1968:38) noted, has a much heavier phonological load in Mandarin than in other dialects. In dialects such as Cantonese, there is no neutral tone in which the stress is completely weak. It is therefore very important that we do not overlook such crucial facts.

3.2 Methodology

3.2.1 Production data of spontaneous speech

It has been mentioned in the previous chapter (Ch. 2) that the production data of speech samples were obtained form only one subject. The first subject COL was an adult female native speaker raised around the neighborhood of Peking. She went to college there, and left China in 1949. The new subject, LH, is also a female native speaker. In her late thirties, LH was born in Peking and educated there. In fact she had barely left Peking before being sent over to Brown University as a visiting scholar in applied mathematics in the fall of 1979. Thus the two subjects speak a nearly identical dialect. The data were collected in the following two settings: The first was during several lecture/seminar classes on Modern Chinese Literature taught by COL, the same data used in Chapter 2. The other consisted of free conversation between COL and LH. The recording session took place at the experimenter's apartment over afternoon tea and cookies. A Nagra tape recorder and a Sony ECM–50PS omni–directional microphone were used. The experimenter did not monitor either the content or the style of the conversation. Since COL and LH had recently become acquainted, no effort was necessary to make them converse with each other. A total of 68 minutes of speech was obtained. The tapes were later transcribed orthographically by the experimenter.

3.2.1.1 Inclusion of tokens

Twenty–five sentences were selected from each speaker's production. For COL, the sentences came from both of the previously mentioned recording sessions. As for LH, since only one recording session was available, her speech was selected from free conversation only. A total pool of 50 sentences, 25 from each subject constitute the data body of this part of the study. All of the selected utterances were simple declarative sentences in their syntactic structure. By simple declaratives we mean a sentence that roughly has the structure of NP + (VP) + NP. If we are dealing with languages such as English, the constraint would be much simpler. We could specify sentences as declarative as long as it has a Subject–Predicate Structure. Nonotheless, we would like to bring the reader's attention to some related facts of Mandarin Chinese. As Li (1975 : 457–480) discussed in detail, a number of languages in the world were investigated for the most common structures in existence. For declarative sentences, there can be two basic structures. That is, besides the most familiar Subject–Predicate structure, there also exists another structure that is actively used in some Asian and African languages, namely, the Topic–comment Structere. This structure, resembling topicalized sentences in English, is such that the theme of a sentence is always located at the beginning of a sentence. Consider sentences such as "I want to see Tom." versus "It is Tom that I want to see". The features of Topic–Comment structures can be summarized as:

(1) The initial NP of a sentence does not have to bear any agent or dative relationship with the verb. Consequently, no morphological agreement between nouns and verbs are required. For example:

Nèige	tsài	tzuòde	jēn hǎu
that–classifier	dish	cook–particle	very good

'That dish was really well made'

The NP <u>tsài</u> 'dish' can not be the agent of the verb <u>tzuo</u> 'cook'.

(2) A sentence can be verbless in the strict sense, but rather with so–called stative verbs which correspond to adjectivals in English. For example:

nèi	jǔng	huā	jēn	piàuliang
That	kind	flower	very	pretty

'That kind of flower is really pretty'.

(3) A sentence can have a so-called double-subject[1]. For example:

(3)

jèige shiuéshiàu	shiàuyuán	hěndà
this-classifier school (NP)	campus(NP)	very big

'The campus of this school is very big, as constrasted to not the buildings'.

Languages having Topic-Comment structures as an active syntactic device are known to have no agreement between nouns and verbs. Also, passivization is very restricted in use (Li, 1975).

The language used for this study, Mandarin Chinese, happens to be a language in which both Subject-Predicate and Topic-Comment structures occur. To restrict "simple sentences" as having subject-Predicate structure only would not account for the language. That is why we incorporated both Subject-Predicate and Topic-Comment structures in our data of "simple declarative" sentences. The following two sentences are samples from the data. Sentence (4) is Subject-Predicate structure, whereas sentence (5) is Topic-Comment.

(4) $_S$[Géng-ěr $_{VP}$[diǎnle $_{NP}$[jǐpán nióu yáng ròu]$_{NP}$]$_{VP}$]$_S$

Geng-er order-asp a few dish beef lamb meat

---

1 The misleading term "double subject" requires some clarification. It does not mean that a sentence can have two subject NP's It also does not mean that the initial two NP's found in a sentence are in the form of one NP modifying the other. It actually mans that a sentence begins with the topic-NP followed by a comment, and the comment begins with an NP. Consequently the sentence has two independent NP's before the stative verb. The term "double subject" would be less confusing if it is called "double NP's." These types of sentences are usually produced with equal stress on both NP's, as well as a pause inserted between the two NP's. If the two initial NP's are compounds, the first NP usually receives more stress than the second one, and no pause can be inserted between the two NP's.

'Geng–er ordered a few dishes of beef and lamb.'

(5) $[_S$ $[_{Top}$ nèijǔng	pùtzi$]_{Top}$	$[_{Com}$ jī	yā	kéyǐ	jū	ròu	
that kind	store	chicken	duck	all right	pork	meat	
bù	shíng	$]_{Com}$	$]_S$				
Neg.	all right						

'It's all right to buy chicken or duck at stores like that, but not pork.'

Thus we included both Subject–Predicate and Topic–Comment structures for our data of simple declarative sentences.

3.3 Results and analysis from the production data

Each utterance was sampled onto RKO–5 disc of a PDP–11/34 computer at the sampling rate of 20 KHz. A PDP–11/34 computer system adapted for convenient analysis of speech was developed. The system makes use of a VT–11 scope with light pen as its basic visual display. Speech waveforms for this study sampled at a rate of 20,000 samples per second with 10 bit quantization. This preserves frequency information at 10KHz. The system can store a sample of up to two minutes of speech at a time in digital form. The WAVE system allows us to display waveforms, listen to them, edit them and perform various acoustic analyses. The signal that is displayed on the scope is also repeated through a loudspeaker or headphones. The duration of the segment displayed can be varied between 8 and 941 msec. The display can be held stationary or it can be moved "forwards": or "backwards" in time through the signal. A joystick controls the rate at which the display moves. The amplitude of the waveform displayed can also be adjusted. Two cursors, a left one, LC and a right one, RC, can be positioned on the waveform by using a light pen. The duration of the interval between the cursors is always displayed. The computer also can provide a computation and plot of fundamental frequency, called the pitch extracting program. A number of parameters must be entered to achieve optimum accuracy in the autocorrelation program that is used. The display that is usually used is such that the speech waveform is display above the computed fundamental frequency contour. These plots were displayed on a tektronix storage scope and a copy of such display can be obtained.

The strength of a digitized speech signal can also be represented graphically. Because the waveform on the computer is composed of discrete values between 0 and 1023 (See for example Mertus, 1980), a window can be described as a finite number of these points. For example, assuming 20 KHz sampling rate, a 25.6 millisecond window consists of 513 points. Given a window of size N + 1, the energy, E, is defined as

$$E = \sqrt{\frac{N}{K=0 \;\; / \, s\,p}} \quad (W(K)-ZL) \qquad \text{ZL is 512}$$

For the present study, the fundamental frequency contour, as well as the energy measurement of each sentence in the production data were plotted and a copy of each was subsequently maintained.

For speaker COL, the duration of her 25 utterances ranges between 1158.5 msec and 3071.9 msec, and for speaker LH between 722.1 msec and 2629.2 msec. The fundamental frequency range, i.e., the highest vs. the lowest Fo produced throughout those sentences for speaker COL is between 461.5 Hz and 102.0 Hz, for speaker LH between 478.5 Hz and 123.04 Hz. The range of the fall of fundamental frequency within an utterances varies from 69.32 Hz to 315.4 Hz for speaker COL, and from 85.05 Hz to 279.7 Hz for speaker LH.

In summary, the duration of the selected 50 utterances ranges between 722.1 msec and 3071.9 msec, the fundamental frequency ranges between 478.6 Hz to 102.0 Hz, and finally the fall of the fundamental frequency within an utterance ranges from 69.32 Hz to 315.4 Hz. (See Ch. 5 for more detailed discussion).

Given the knowledge we have about Mandarin phonetics and phonology, that is, if we know how the lower level phonetics are manifested due to certain phonological constrains, and given the syntactic structure as well, we should then be able to predict what a given sentence might sound like. The production data would then suffice to justify how valid the above mentioned hypothesis is by comparing the physical property of a sentence with its predicted phonetic output. The task here was to see whether a one–to–one correlation can be found between the phonological prediction and the actual phonetic output. In order to make the comparison, we first transcribed the sentences according to the phonological prediction of word tones, making the tones of all syllables as they were supposed to be produced. However, we would like to point out that the 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 (See 3.1.1.1.–4.). We are fully aware of the fact that sentence intonation interacts with lexical tones as well as word stress; thus causing modulation of tones as well as intonation. Unfortunately due to the lack of phonological rules beyond the phrase level, we thus had to limit the phonological predictions to the phrase level. We then examined the actual phonetic output of each sentence, picking out those syllables which contained sufficient information. In other words, we tallied those syllables whose fundamental frequency patterns can be identified visually as having the distinct tonal contours specified by the phonology.

Correlating fundamental frequency contours with lexical tones is by no means an easy task. Figures 3.2 and 3.3 are examples of the less difficult kind, due to the short duration as well as small numbers of syllables.

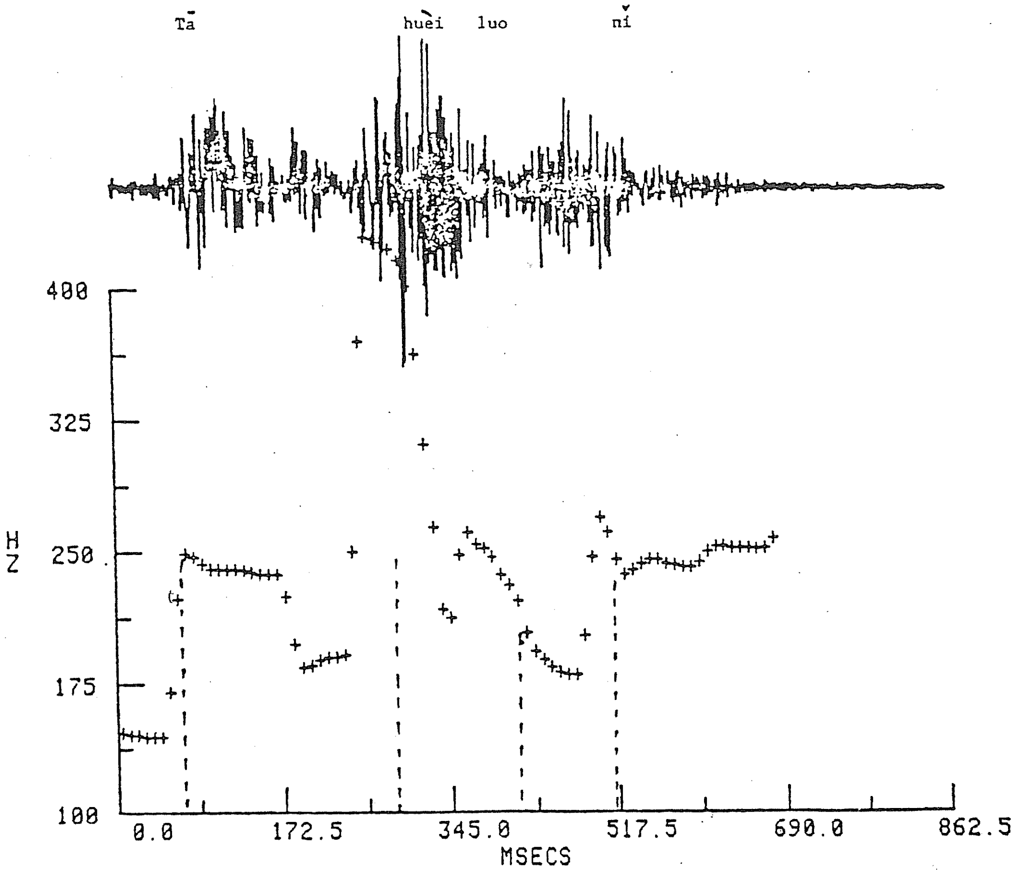


Figure 3.2. Fundamental frequency contour of Mandarin sentence Tā huèiluo nǐ 'He bribes you.'

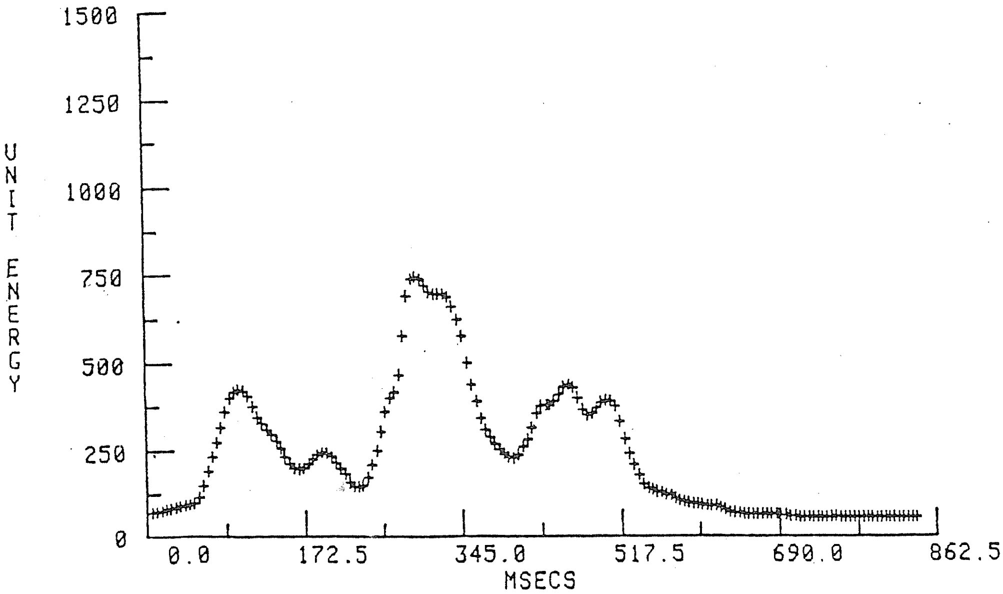


Figure 3.3. Energy measurement of Mandarin sentence Tā huèiluo nǐ 'He bribes you.' as shown in Figure 3.2.

The PDP–11/34 pitch extractor program made it possible to derive the fundamental frequency contour, as well as the speech wave forms of the selected speech sample to be both present on the same graphic display. However, since we are no longer dealing with isolated Mandarin syllables, no discrete tonal patterns are expected from fluent spontaneous speech. Especially when the unit energy of a sentence is somewhat even throughout the entire sentence, the identification of each individual lexical tones in such a sentence is almost impossible. In the sentence in Figure 3.2., Ta hueiluo ni 'He bribes you' The dotted line drawn over the graphic display of the pitch pattern of the sentence shows approximately where each syllable begins. Figure 3.3 showed that the unit energy measurement of the same sentence is rather flat. This sentence is not too long in duration. Sentences that are longer in duration usually are more difficult for the identification of the lexical tones for the contained words. Fortunately, since it was possiple to move the waveforms of the selected speech sample back and forth on the refresher screen, together with the aid of the lightpen to move the cursors, i.e., the left cursor (LC) and the cursor (RC), any section of the speech samples can be isolated, listened to, and finally analyzed.

The overall fundamental frequency contour of each of the fifty sentences was then examined on the refresher screen to determine where the words occurred. Both orthographic representation in Chinese and transliteration in MPS II (Mandarin Phonetic Symbols II) were made for each sentence on the original copies. As we noted before, we should not expect every word tone to be present in connected speech, especially in fluent conversational style. As a result, the focus was then on the following:

(1) Which are those syllable and/or words in a sentence whose word tones are clearly produced, such that the fundamental frequency contour of these words still possess the distinct pattern and thus can be clearly identified on the graphic dispaly?

(2) Which are those words in a sentence produced in fluent spontaneous speech whose lexical tones are not present? The toneless morphemes? The unstressed morphemes? Or those morphemes whose tones are deleted due to certain sandhi rules?

The analysis was done by first breaking down each sentence into syllables, which are the units for tones in Mandarin. The total number of syllables in the data were calculated. Since each sentence had been transcribed according to the tones predicted by the phonolgy, the next task was to check how the actual phonetic output from the production data correlated to the respective phonological prediction. For example, in case of the sentence in Figure 3.2, tā huèiluo nǐ 'He bribes you', the dictionary first assigns each syllable with a distinct tone, i.e., the syllable tā 'he' in level tone, the two syllables huèiluò 'bribe' both in falling tone, and the syllable nǐ 'you' in falling–rising tone. Then the stress rule assigns the stress for the only di–syllable word huèiluo 'bribe' on the first syllable, resulting with the syllable huèi remaining in its original falling tone, while reducing the tone of the syllable luò from the falling tone into the neutral tone. Thus the word huèiluo 'bribe' should be produced with a falling tone followed by a neutral tone. Moreover, the neutral tone of the syllable luò is predictable, because the rule (See 3.1.1.2.) specifies the neutral tone following a falling tone, in this case huèi followed by luò, be produced as a low level tone. The ultimate phonological prediction of the sentence tā huèiluo nǐ 'He bribes you', should result in tā 'he' in level tone, huèiluo 'bribe' a falling tone followed by a neutral tone in the form of low level tone, and nǐ 'you' in falling–rising tone. Some interaction between the intonation contour and the individual lexical tones were expected, but we had to exclude this factor for the lack of rule specifying the interaction. Figure 3.2, the actual phonetic output of the sentence shows that the predicted tonal pattern for tā 'he' and huèi 'bribe' and nǐ 'you' were present, whereas luo lacks the low level tone. Therefore for this sentence, it was tallied that three out of the four syllables had their tones corresponding to the phonological prediction.

The analysis of the production data is summarized in Table 3.1.

Table 3.1. Acoustic analysis of production data from spontaneous speech

Number of syllables in that data	Number of syllables that a one–to–one phonological–phonetic correlation is found	Percentage of correlation
465	168	36.13%

A further analysis of the production data is summarized in Table 3.2.

Table 3.2. Further acoustic analysis of production data from spontaneous speech

phonological prediction		actual phonetic output that correlates the phonological prediction	percentage of correlation
syllable whose lexical tones should present in production	361	145	40.17%
syllables that are neutral–tone suffixes	44	13	29.55%
syllables whose lexical tones are reduced	60	10	16.67%

For the 50 sentences constituting the data of the present study, there are 465 syllables altogether. Among these 465 syllables, 44 syllables are neutral–tone suffixes, whose tonal patterns are predictable according to the tones of the preceding syllables. 60 syllables should be neutralized to neutral tone according to the stress rules or neutralization rules. 361 syllables are supposed to be produced with their lexical tones in unreduced forms. However, the result of the production data reveals that among the 465 syllables, only 168 syllables were produced with their predicted tonal patterns. In other words, 36.13% of the data can be said to have a one–to–one correlation between the phonological prediction and the actual phonetic output. A further break–down of these correlated syllables presents a more detailed picture of the kinds of syllables maintaining a higher rate of the phonological–phonetic correlation. Among the 361 syllables whose tones should be unreduced by phonolgical prediction, 145 syllables were actually produced as predicted. That is, the percentage of phonological–phonetic correlation of the unreduced syllables were 40.17%. Moreover, 134 syllables out of the 44 neutral–tone syllables, i.e., 29.55%, were produced as the tones of the preceding syllable predicted; and 10 syllables out of the 60 neutralized syllables, i.e., 16.67% were actually produced in neutral–tone pattern as the rules predicted[2].

3.4 Discussion: the physical property of tones in spontaneous speech

The major finding of the production data is that we were unable to locate a high correlation between the phonological prediction and the actual phonetic

---

2 For the two subjects in this study, the so–called neutral tones were produced at the frequency scale around 175 Hz as a flat plateau, as contrasted to the level tone which possesses the same fundamental frequency pattern at a higher pitch of at least 250 Hz.

output. In other words, only 36.13% of the production data corresponded to the prediction by the phonological rules. Aomong the syllables whose lexical tones should remain unreduced according to the phonological specification, 40.17% were actually produced as they were predicted. Among those neutral–tone suffixes whose tonal patterns can be predicted according to the tonal pattern of the preceding syllable, only 29.55% correlated to the prediction. Moreover, among those syllables whose lexical tones should be reduced to neutral tone due to the rules of tone and stress, only a mere 16.67% of correlation was found. The result of the production data provided strong evidence against the phonological predictions. In other words, the actual physical properties of tones in spontaneous speech demonstrates that the acoustic information regarding tones, mainly the fundamental frequency pattern, was only partially present in the production of natural spontaneous speech. We did not find a one–to–one correspondence between the phonolgical prediction and the actual phonetic output, which suggests that in spontaneous connected speech, speakers only produce partial acoustic information regarding tones.

Our next attempt was to find out what types of words were actually produced with sufficient acoustic information, as well as what types of words were produced with less specific acoustic information. By this we mean the syllables produced with identifiable tonal patterns as contrasted to syllables with unidentifiable tonal patterns.

Our production data do suggest a higher rate of the phonolgical–phonetic correlation among those syllables whose lexical tones were not reduced. If we list those among which the lexical tones were produced as the phonological rules specify, we found that such words include pratically every part of speech. They are: proper nouns, nouns, pronouns, auxiliaries, verbs, adjectives, adverbs, adjectivals, demonstratives, conjunctions, and negative markers "<u>bù</u>" and "<u>méi</u>". Obviously such kind of listing does not lead us anywhere. However, instead of just looking at the phonetic realization, if we also take into account the syntactic structure of these utterances, a clearer picture began to emerge.

A close examination of sentences (4) and (5) (see Appendix A), that is, looking for both reduced and unreduced elements in the sentences, reveals some interesting phonomena. Figures 3.4 and 3.5 are graphic displays of the fundamental frequency contours for sentences (4) and (5) respectively.

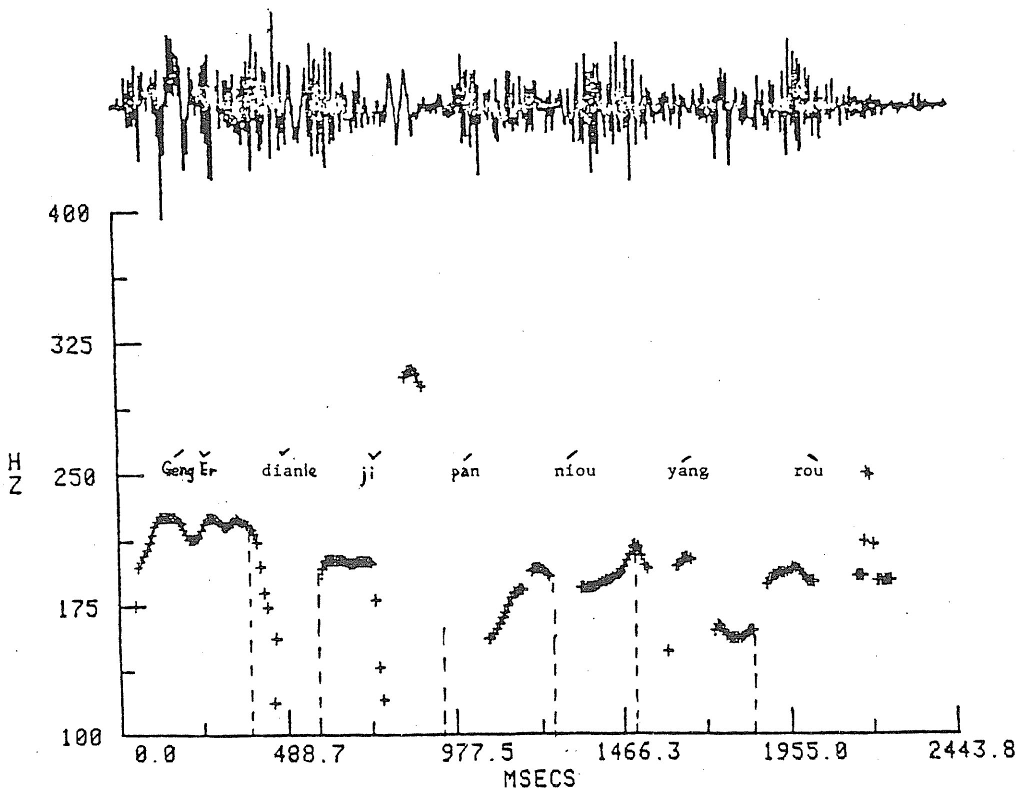


Figure 3.4. Waveform and fundamental frequency of Mandarin sentence <u>Géng-Ěr diǎnle jǐpán níou yáng ròu</u> 'Geng-Er ordered a few dishes of beef and lamb.'

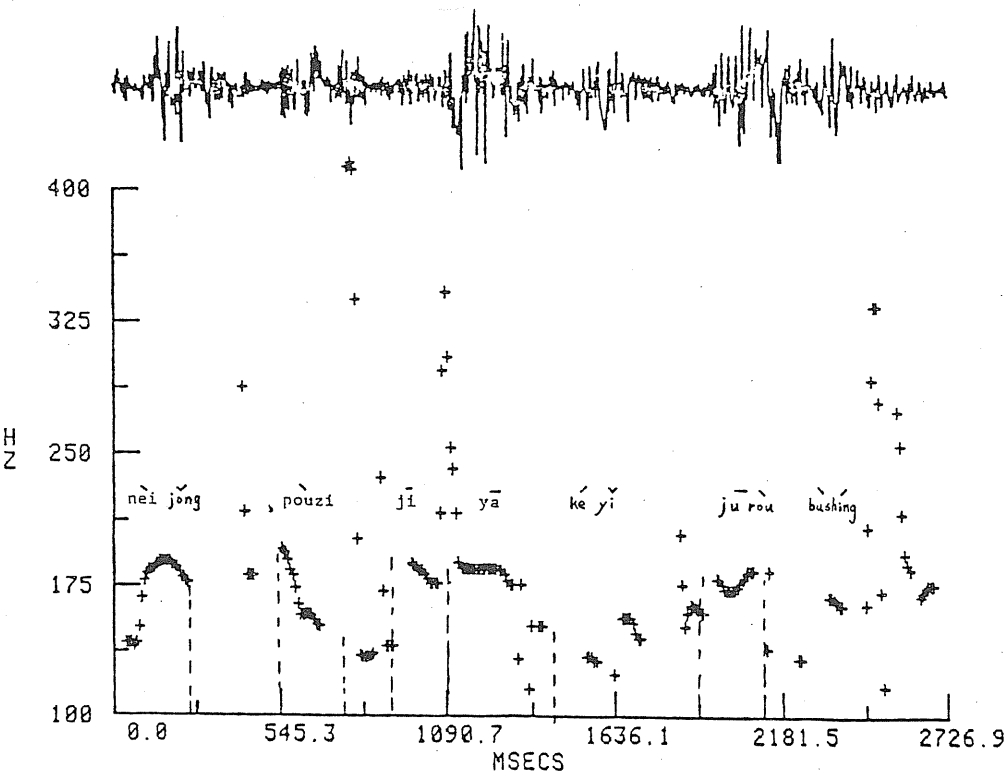


Figure 3.5. Waveform and fundamental frequency of Mandarin sentence <u>Nèi jǒng pòuzi jī yā kéyǐ jūròu bù shíng</u> 'that-king store chicken duck OK pork not OK--It is all right to buy chicken or duck in that kind of store, but not pork.'

The following presentation of the sentences may offer us some clue of what these sentences might denote beyond their literal meaning. First, let us put in parenthesis those syllables that were produced with partial acoustic information regarding tones, and no transliteration beneath. What is left afterwards are those syllables that were produced with sufficient acoustic information as the phonological rules sometimes would specify and presumably were more salient cues. We then only transliterated these syllables without the parenthesis. The transliteration becomes quite suggestive in terms of the meaning of the sentences.

(4)	(Géng–Ěr)	diǎn(le)	(jǐ)	pán	nióu	yáng	ròu
		order		dish	beef	lamb	meat

(5) nèi(jǔng) pùtzi, jī yā (kéyǐ), jūròu (bushing)
that store chicken duck pork

What we get, is what one might call a telegraphic kind of sentence, which might be said to contain only the key words that carry out not necessarily the meaning of the entire sentence, but what probably the speaker considers the primary information he intends to convey. Of course each of the above sentences can be emphasized in other ways, and thus the words whose tones are more fully produced might be other words in the sentence. For instance, when uttering sentence (5), instead of saying that in stores run by Muslems, it is perfectly all right to ask for chicken or duck, but absolutely not pork, the speaker might fully articultate the words kéyǐ, 'all right', as contrasted to bùshíng 'not all right', but still using the same sentence.

This kind of phenomenon occurred for both subjects. Findings of this sort lead to the following generalizations: when we as linguists take a sentence into consideration, we examine the structure first, break it down to constituents, figure out the word structure and apply the word stress rules, single out the neutral–tone morphemes, apply the necessary stress and neutralization rules, then we have what the phonology would predict as output, at least in comparison with those morphemes whose lexical tones are reduced. Then we try to consider the overall intonation contour as well and thus derive what we presume the most likely way of producing that particular sentence. All of these analyses would be performed without considering the contextual information that might involve. On the other hand, what a speaker may have in mind could be something entirely different. What he is more concerned is how to get the meaning across most efficiently in oral communication. The speaker, as well as the hearer, both share some common knowledge of their native language. Thus, while a substantial portion of a sentence might be produced with partial acoustic information, the listener appears to have no difficulty to derive the meaning of such sentences, intergrating his knowledge beyond the phonetic level in the process of decoding the incoming acoustic signal.

The speaker on the other hand appears not to watch very closely whether he has followed the word formation rules, and/or the sandhi rules faithfully, but rather to concentrate on what he considers of primary communicative significance. For example, in the present data, we found two instances in which the subject produced the supposed–to–be reduced syllables with their lexical tones without applying the word–stress rule. One example is in the sentence nèige shŕhou méiyǒu shénme jr̄shi 'that time not have much knowledge —— During those days people didn't have much knowledge (of nutrition).' In this sentence, the syllable hou 'time' in shrhou 'time' should be reduced from the falling tone which is its lexical tone to the neutral tone, because the stress falls on the syllable shŕ 'time'. However, the subject produced it with a clear falling tone which although violating the word–stress rule, was nonetheless unambiguous in terms of its meaning. Thus, this production could be regarded as an acceptable pronunciation, whereas to produce this syllable with any other of the three lexical tones would certainly cause ambiguity. This further suggests that the reduced tone of a syllable can be produced with its original lexical tone, should the speaker intend to clarify or emphasize. In other words, all the rules do exist, but certain degrees of violations appear to be tolerable, and the constraints of these rules vary as well. Thus we should certainly expect linguistic communication to be a much more complicated phenomenon. Similar observations have been made by Bolinger (1955). Tone neutralization, especially the sandhi rules may probably partly be the consequence of certain physiological constraints. For instance, two falling–rising tones sequence become a rising + falling–rising tone sequence.

As far as how sentences are actually produced, we would like to demonstrate and offer one interpretation of what in fact is going on. If we accept the explanation that speech production is a complicated interaction involving phonetics, phonology, semantics, syntax, and above all, the speaker's knowledge of the language and furthermore and most importantly, his intention regarding what should be conveyed, we might then be able to look at speech communication in a more productive manner. The data presented here provide us with incidences of fully produced tones in almost every part of speech in the system, but overall at a very low rate of correlation between the phonological prediction and actual phonetic output. The result supports the argument that not all acoustic information, in our present study with respect to tones in particular, has to be physically present during the production of spontaneous speech.

Another attempt in this study was to find out what types of words were produced with less acoustic information. Our production data already demonstrates an even lower correlation between the phonolgical prediction and the actual phonetic output among the neutral–tone suffixes (29.55%) as well as those syllables whose lexical tones were supposed to be reduced to neutral tones (16.67%). These results suggest that as far as neutral–tone suffixes are concerned, their tonal pattern are not alway governed by the tonal pattern of the preceding syllable, as the phonology specifies. On the contray, they were produced with more variations. As for the reduced syllables, they appeared to be even less consistent, which suggest that the reduction of tones might be only predictable at the phrase level, as the rules were so derived, but definitely not at the sentence level. In other words, producing polysyllabic words involving reduction of tones in elicited forms outside the framework of a sentence would probably offer a higher correlation between the phonological prediction and the phnetic output.

Another example will be presented here to illustrate a case in which the lexical tones of certain unreduced syllables were consistently produced with less acoustic information then the existing phonolgical rules specify. For example, the word wǒ 'I' or 'me', is lexically assigned with a falling–rising tone. It appeared 13 times throughout the data, and the existing phonological rule only specifies that at the non–final position in a sentence, the falling–rising tone lacks the rising portion (See 3.1.1.4.). The 13 occurrences of the particular word wǒ 'I' appear both in subject and object positions in sentences, and for four times it occurred in sentence–initial position. However, out of the 13 occurrences, wǒ 'I' was never once produced with fully identifiable tonal contour. Fig. 3.6 and 3.7 are two examples of such sentences.

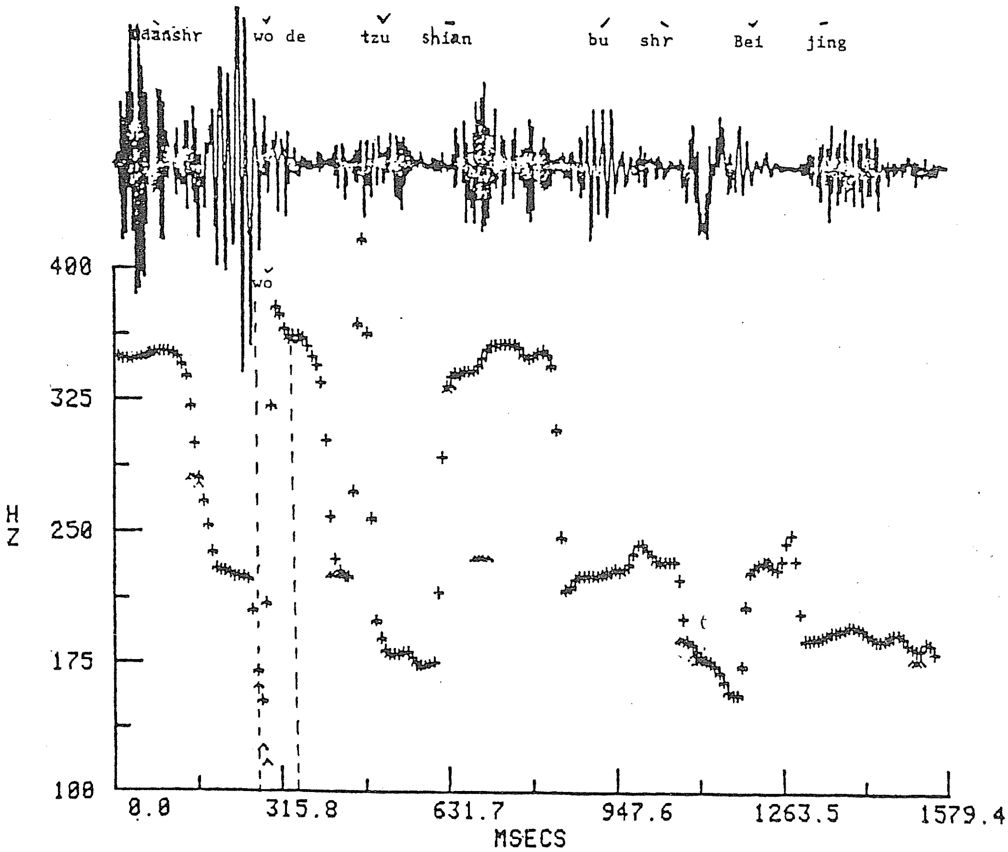


Figure 3.6. Fundamental frequency contour of Mandarin sentence Dànshr wǒde tzǔshian bú shr Běijīng 'But I-possessive ancestor not copular Beijing· But my ancestor was not from Beijing.' Note the F0 contour for the syllable wo 'I'.

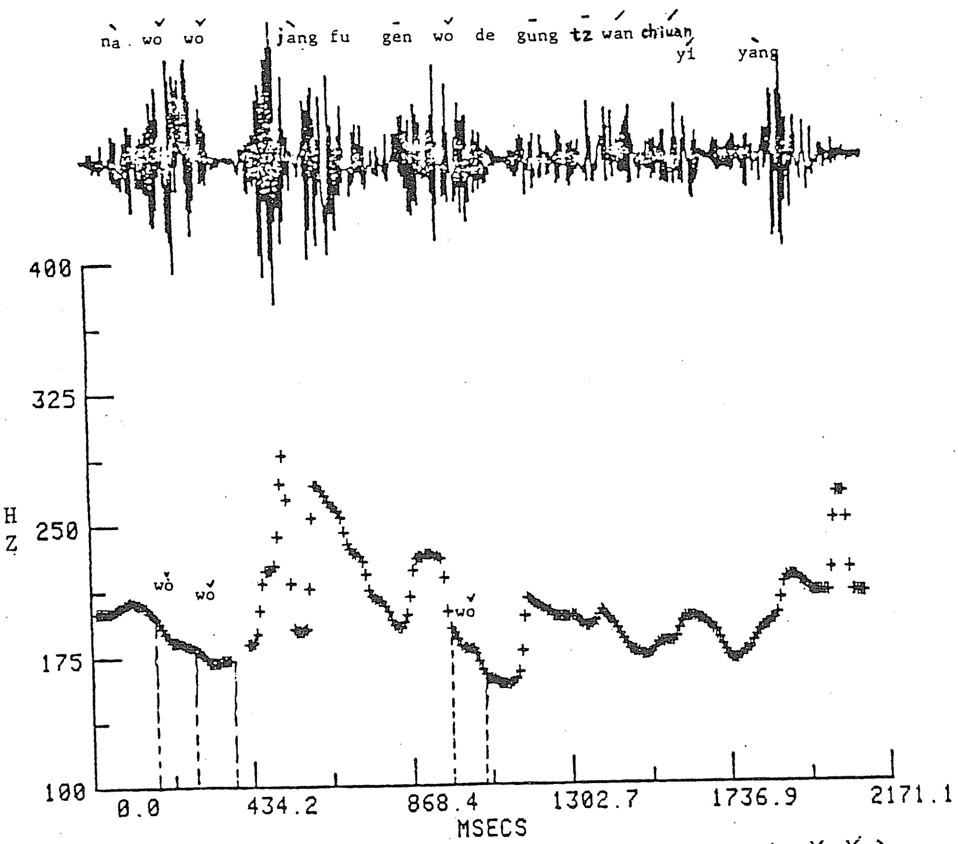


Figure 3.7. Fundamental frequency contour of Mandarin sentence Nà wǒ wǒ jàngfu gēn wǒde gūng tz wánchiúan yíyàng 'Then my, my husband and my salary exactly same--Well then, my husband's salary is exactly the as mine.' Note the F0 contour for the three wo's 'I' in this sentence.

In this sentence, wǒ is combined with de, here as a possessive marker thus making wǒde the possessive pronound 'my' or 'mine'. Since it is followed by a neutral–tone suffix, no tonal interaction is expected on the part of wǒ. However, the figures demonstrate no distinctive falling–rising pattern at all. Instead, it was produced with a rather flat fundamental frequency contour at fairly low pitch. The reason why wǒ is never fully articulated in terms of its tonal contour in all occasions present in the data here requires us to examine this particular morpheme and its status in the system. Contrary to our best known examples mā 'mother', má 'hemp', mǎ 'horse', and mà 'to scold', the wo syllable presents a different case. When level tone is assigned, wō means 'nest' or 'to nest', and is usually accompanied by a preceding modifier, for example *jūwō* 'pig sty', whereas the combination of *wo* and the rising tone *wó* does not exist in the lexicon and thus is a nonsense syllable. When the falling–rising tone is assigned, *wǒ* can only mean 'I' or 'me', the first–person pronoun, whereas when a falling tone is assigned *wò* means 'to lie down' and would almost always combine with another syllable, either preceding or following, to form a disyllabic word, such as *wòshia* 'to lie down' or *tzuòwò* 'to sit or lie down', instead of appearing in monosyllabic form. Since the speaker knows that when *wo* occurs at either subject or object position, it could mean nothing else but 'I' or 'me', as a result he does not have to bother to maneuvre the articulatory gestures to specify a particular fundamental frequency contour, because the absence of the specified tonal pattern causes no ambiguity. This example also further illustrates what a speaker *can* do in actual speech production.

3.5 Summary and conclusion

In this section, we have presented production data for 50 simple declarative sentences from two female speakers in spontaneous conversation. The data demonstrate that within a sentence, unless it is intened that every syllable be fully articulated, only a small number of syllables were produced with their lexical tones as the phonology predicts. However, we recognize the fact that the phonological prediction only involves the phrase level. Phrases read in citation forms may serve as better sources to test the specified prediction. The absence of specific rules at the sentence level, incorporating sentence intonation and the interaction between tones and intonation might further explain the fact that only partial acoustic information is required. The actual phonetic output is only predicatable when all interacting information involved is accounted for, namely, phonetics, phonology, syntax, semantics, and moreover, the speaker's knowledge of the language as well as the speaker's intention as to what elements of that particular sentence should constitute the overall meaning of the sentence. Thus we conclude that the production of tones in spontaneous speech in terms of the required acoustic information is only partial, and is very likely the consequence of a complicated interaction among all these above mentioned levles; rather than just a low level phonetic phonomenon. This is expected in every language, tonal or non–tonal. In other words, as far as tone is concerned, it might be a very different linguistic experience to produce it and perceive it, but its role in the overall sense of speech communication is no different from other segmental components.

Chapter IV

CHAPTER 4 THE PERCEPTION OF MANDARIN TONES IN SPONTAEOUS SPEECH

4.1 Objectives

Recall that in Chapter 2, we demonstrated that native speakers of tone language are capable of identifying their native tones presented isolation, when the tones are produced in citation forms. Duration appeared to play only a concomitant role in tone perception, as long as the necessary fundamental frequency contour was present. Thus it was concluded that the fundamental frequency contour is the primary cue in tone perception. However, the production data in Chapter 3 provided evidence that in actual spontaneous speech, the production of tones is no longer as distinct as the way they are produced in citation forms.

As Cole and Jakimik (Cole, 1980 : 133) have pointed out "words do not exist as discrete physical events in fluent speech. If we remove the sounds which accompany a word from their fluent speech context and listen to them in isolation, we often cannot recognize the word. The sounds which accompany a word may be insufficient by themselves to produce its recognition. Words, then, are the product of speech perception. They exist in the mind of the perceiver, and not in the physical stimulus."

We suspect that similar situations happen to tones, whose primary acoustic correlates are the fundamental frequency pattern. When presented with fluent speech, listeners often will "hear" or rather "reconstruct" all the lexical tones on the basis of partial acoustic information. In this chapter, we will investigate how such spontaneously produced tones are perceived under the following condition. If we eliminate the context and interacting factors involved in spontaneous connected speech, namely phonology, semantics, syntax, and the speaker's linguistic knowledge and communicative intentions, given what is left, that is, the phonetic output of related syllables extracted from spontaneous speech alone, what then do native speakers perceive? In other words, can native speakers of a tone language judge the relative fundamental frequency value when other factors, linguistics and/or pragmatic, are absent? What role does the actual fundamental frequency contour in spontaneous speech play in tone perception? How much of the fundamental frequency pattern is required for perceiving tones without any context? How far can a speaker's linguistic competence of tones allow him to go? In summary, we are interested in finding out how listeners perceive tones without the support of context.

4.2 Methodology

4.2.1 The experiment: how lexical tones from spontaneous speech are perceived in isolation

The purpose of the perception experiment in this section is to see how word tones from fluent spontaneous speech can be perceived without context. Using the production data from Chapter 3 as a source of stimuli, tokens were extracted from spontaneous speech. The syllable is used as the unit of study, since the smallest unit for tones in Mandarin is the syllable. Computer implemented techniques were employed to generate the stimuli. Through a PDP–11/34 computer, the waveforms of the recorded sentences were displayed on a Nell refresher scope. By moving two cursors RC and LC (see 3.2), following not only the displayed waveform on the refresher scope, but also listening to the auditory output through a headphone, the segment between the cursors were edited out from a waveform and stored on a disc. The pitch extracting program and the energy frequency program were then used to derive the fundamental frequency contour as well as the unit energy measurement of the edited stimuli. Hard copies were made of the plotting for each token. Because the source of stimuli was recorded during naturalistic settings, background noise was inevitable. When a (C)V syllable was followed by another vowel with no intermediate glottal stop in between, the exact division of the two vowel sequences became impossible. These types of syllables were not included in the experiment. Each token was labelled with reference to the speaker who produced it, and whether it was a reduced or unreduced syllable within a word according to the phonology (3.1.1.3). The selected unreduced tokens were all produced with at least 250 unit energy measurement (3.2.1), whereas the reduced tokens were produced below 150 unit measurement. Thus there existed a considerable difference in amplitude among these tokens, The unreduced category includes either monosyllabic words or the stressed syllable from a polysyllabic word, which were the syllables that were produced with more sufficient acoustic information regarding their fundamental frequency pattern, correlating to their lexical tones from the phonology. The reduced category, on the other hand, included both neutral tone suffixes as well as syllables whose tones should be reduced according to the phonology (3.1.1).

4.2.2 Stimuli and Test Procedures

The stimuli: One hundred reduced tokens as well as one hundred unreduced tokens were chosen for the perception test. Out of the 100 unreduced tokens, 50 tokens were from speaker COL, 50 other tokens were from speaker LH. The same conditions applied to reduced tokens, too, so we could maintain an equal distribution of the test tokens from both subjects. Since both COL and LH speak standard Mandarin, no dialectal variation should exist. However, because the test stimuli were extracted from natural spontaneous speech, they varied drastically in amplitude. The duration for the test stimuli also vary. The following measurement provides an illustration. For the unreduced tokens, the longest token produced by COL is duō 'many in hěnduō 'very many', 618.15 msec. The longest unreduced token produced by LH is dàu 'to go', 436.6 mesc. The shortest unreduced token produced by COL is dōu 'all', 108 msec, and by LH is tzǔ in tzǔshiān 'ancester', 113.7 msec. For the reduced tokens, the longest token produced by COL is yuán in fúwùyuán 'service staff', 354.5 msec; by LH is de 'possessive suffix', a neutral tone morpheme, 301.0 msec. The shortest reduced token produced by speaker COL is yi in yíyàng 'the same', 45.6 msec; by LH is shr̀ in dànshr̀ 'but', 106.5 msec. Table 4.1 summarizes the distribution of stimuli in duration.

Table 4.1. Distribution of test stimuli in duration

		reduced tokens	unreduced tokens
duration range	subject COL	618.5—108 msec	354.5—45.6 msec
	subject LH	436.6—113.7 msec	301.0—106.5 msec

These 200 tokens of stimuli were then randomized by the computer and recorded onto a tape. Each stimulus occurs only once. There was a 5–second interval between each stimulus, and a 10–second interval between each block of ten stimuli.

The test procedure: the test required subjects to listen to the 200 stimuli through a headphone and label the tone they thought that they had just heard. Six choices were given to the subject for every stimulus they heard: the four lexical tones in Mandarin, the neutral tone, known to Chinese as chīngshēng, and finally nothing recognizable. Twelve native Chinese, eight males and four females served as subjects for the perception experiment. Except one subject who was strictly a mono–lingual Mandarin speaker, the rest of the subjects were all bi–dialectal. Besides Mandarin, they speak a variety of dialects including Amoy, Cantonese, Nanking, Hokka, and Yangzhou. Each subject was first presented with a 10–token practice session. Those tokens did not appear during the actual test. Subjects then were presented with the test without any interruption.

4.3 Results and analysis

The tokens were broken down into two groups for scoring: the reduced tokens vs. the unreduced tokens; 100 each in number. For the unreduced tokens, the original lexical tone of the morphemes was taken as the correct tone to be identified. As for the reduced tokens, they consisted of the following two types: toneless morphemes, unstressed morphemes in a word whose lexical tones had been deleted to neutral tones by either a word–formation rule or some sort of sandhi rule. Thus all the reduced tokens should first become neutral tones, then the phonetic output of these neutral tones were modulated by the tones of the preceding syllables. Therefore both the reduced and unreduced tokens were scored according to the phonological predictions. The results of the subjects' performance was as follows: The overall performance showed 40.5% correct indentification. For the reduced tokens, 25.4% of the tokens were correctly identified. For the unreduced tokens, 55.66% was correctly identified.

Recall that the analysis of our production data (Chapter 3) yields the result of a fairly low correlation (36.13%) between the phonological prediction and the actual phonetic output of the selected speech samples. We felt at this point that an acoustic analysis of the test stimuli might reveal similar results of the production data in Chapter 3, which could further lead for scoring of the perception test according to the actual acoustic information. For instance, the pitch plotting of the reduced tokens demonstrates that although some of such tokens are indeed toneless ,i.e., others may vary while still carrying a distinct pitch contour which does not necessarily correspond completely to the lexical tone of that morphemes, nor its modulated tone according to the phonology. Figure 4.1 is an example of a morpheme whose lexical tone had been modulated and thus became a phonetically neutral tone.

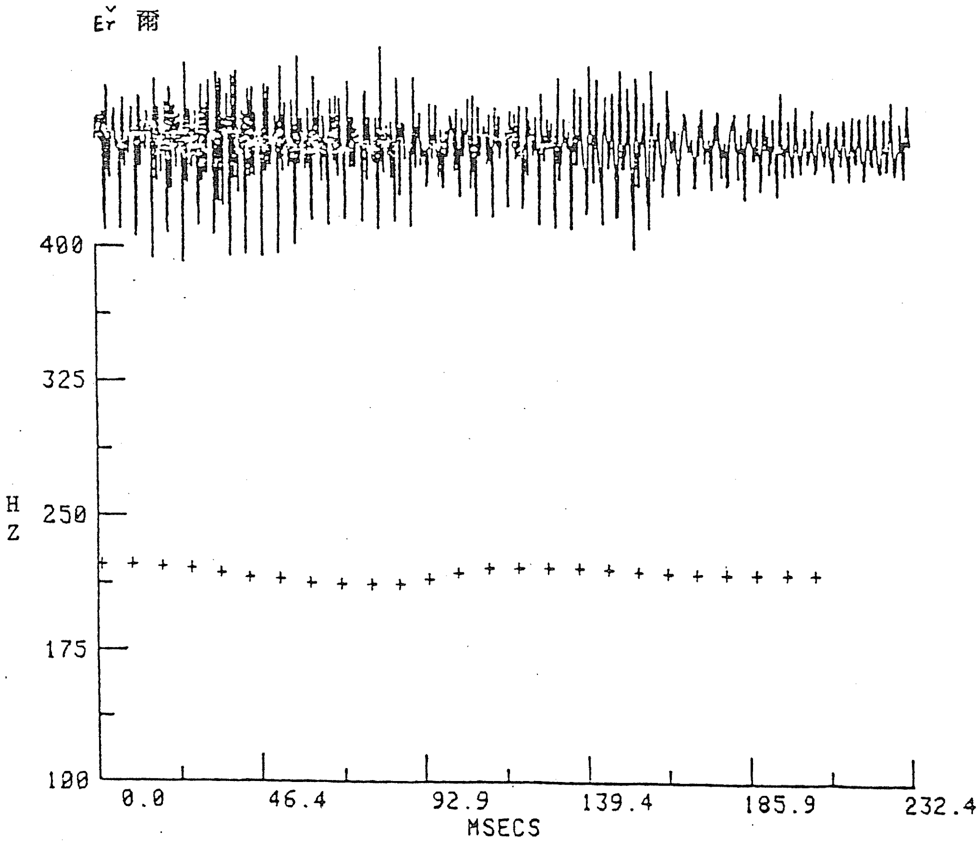


Figure 4.1. Morpheme Ěr 爾 in Mandarin sentence Géng-Ěr diǎnle jǐoán nióu yáng ròu 'Geng-Er ordered a few dishes of beef and lamb. 'Note the lexical tone of Er,which is falling-rising, had been modulated by the sentence intonation. The F0 pattern lacks the soeficied lexical tonal contour.

The sentence was Géng–Ěr diǎnle jǐpán nióu yáng ròu 'Geng–Er ordered a few dishes of beef and lamb.' The token presented in Figure 4.1 is the morpheme Ěr from the proper noun Géng–Ěr, which should have been the falling–rising tone (See 3.1.1.4.). The phonetic output for morpheme Ěr in this case was produced at the frequency range of 225 Hz in a more or less flat contour which is phonologically predictable. In this case, it appears highly unlikely to perceive this particular token as having the falling–rising contour since the specified contour physically does not exist. This could mean that the original lexical tone of the token was absent, but the particular token still demonstrated a distinct tonal pattern, instead of the predicted tone pattern. This is particularly true regarding the phonologically predicted toneless (neutral tone) tokens. Figures 4.2 and 4.3 are examples illustrating this type of test tokens.

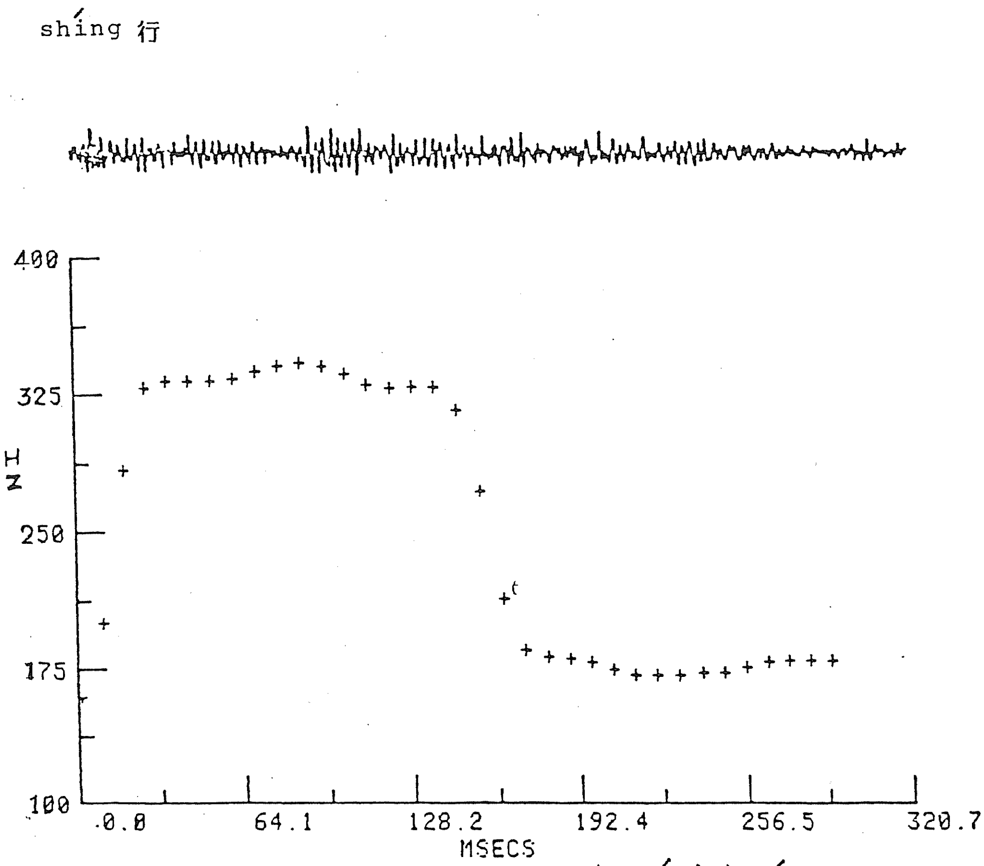


Figure 4.2. Syllable shing 行 in Mandarin sentence Nǐ bù shíwǎn bù shíng 'you not wash dishes not all right-- If you don't wash the dishes, then it's not OK.' The F0 pattern of this syllable exhibits a distince pattern instead of the rising lexical tone.

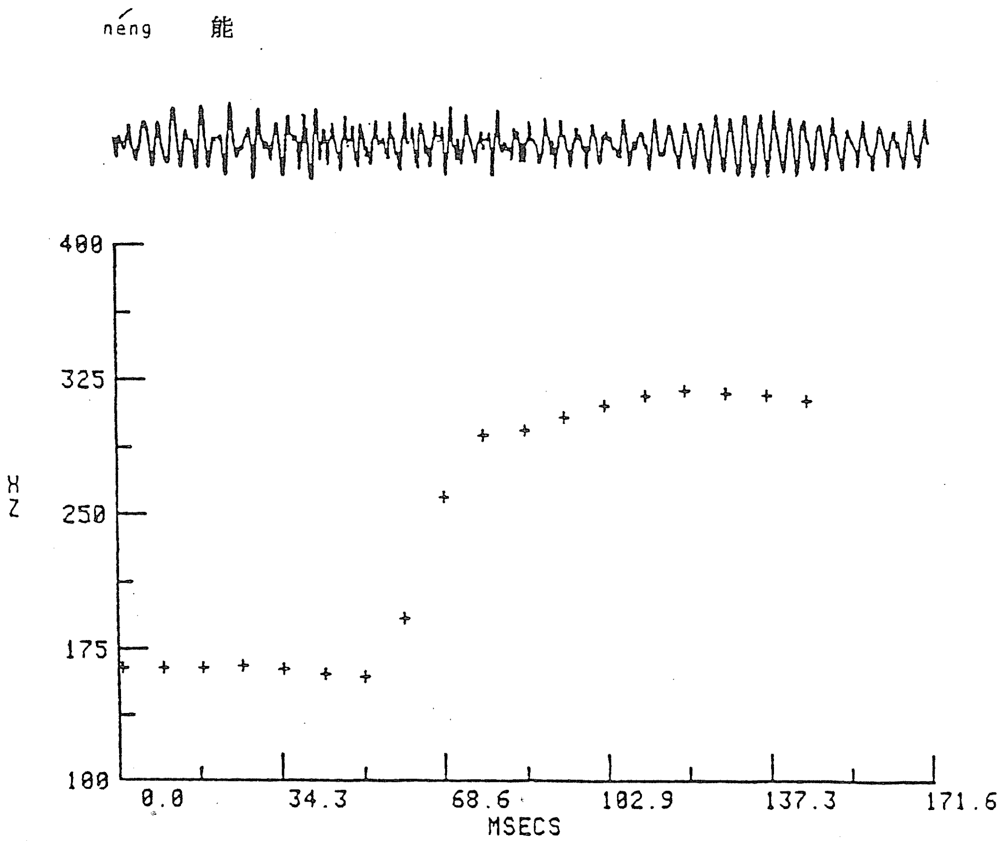


Figure 4.3. syllable néng 能 in Mandarin sentence Bù nénggou suéibiān chr̄ duōshǎu 'not as you wish eat how much--One cannot eat as much as one feels like'.The F0 pattern of this syllable exhibits a distinct pattern instead of the rising lexical, and is somewhat a mirrage image of the F0 pattern of another rising tone syaable shown in Figure 4.2.

Figure 4.2 is the morpheme shíng 'all right' in rising–tone from the sentence Nǐ bù shíwǎn bùshíng 'you don't wash dishes not all right ——If you don't do the dishes, then it's not Ok'. Probably because of its sentence–final and also unstressed position, the morpheme shíng 'all right' which should be stressed according to the stress rule, neither possessed its original rising–tone, nor had been reduced to phonetically toneless. In this case, if the syllable shíng was reduced to neutral tone first due to its position in a sentence, since it was preceded by a falling–tone syllable bù 'not', the phonology predicts that shíng becomes a low flat tone. Instead, it showed a contour consisting of a plateau at about 350 Hz for around 135 msec, then suddenly dropping to another plateau at about 175 Hz for the remaining 165 msec.

Figure 4.3 is an example of a rising–tone morpheme exhibiting a different fundamental frequency pattern from that of Figure 4.2. The morpheme in Figure 4.3 is the morpheme neng in rising–tone from the sentence Bù nénggou suéibiàn chr̄ duōshǎu 'Cannot as you wish eat how much —— One can not each as much as one likes.' Unlike the morpheme shíng in Figure 4.2 which occurs at unstressed sentence–final position, the morpheme néng occurred at unstressed but sentence–initial position. The tonal pattern displayed in Figure 4.3 was almost like a mirror–image of Figure 4.2, a plateau below 175 Hz suddenly going up to another plateau at around 325 Hz, demonstrating another variety of the rising–tone morpheme. A third example of a phonologically specified rising syllable displayed yet another kind of fundamental frequency pattern. From the sentence Tāmen shénme dōu jiàu fúwùyuán 'They everything call service–member —— They call everyone service–members', the final morpheme of the sentence yuán 'member' displayed the contour presented in Figure 4.4.:

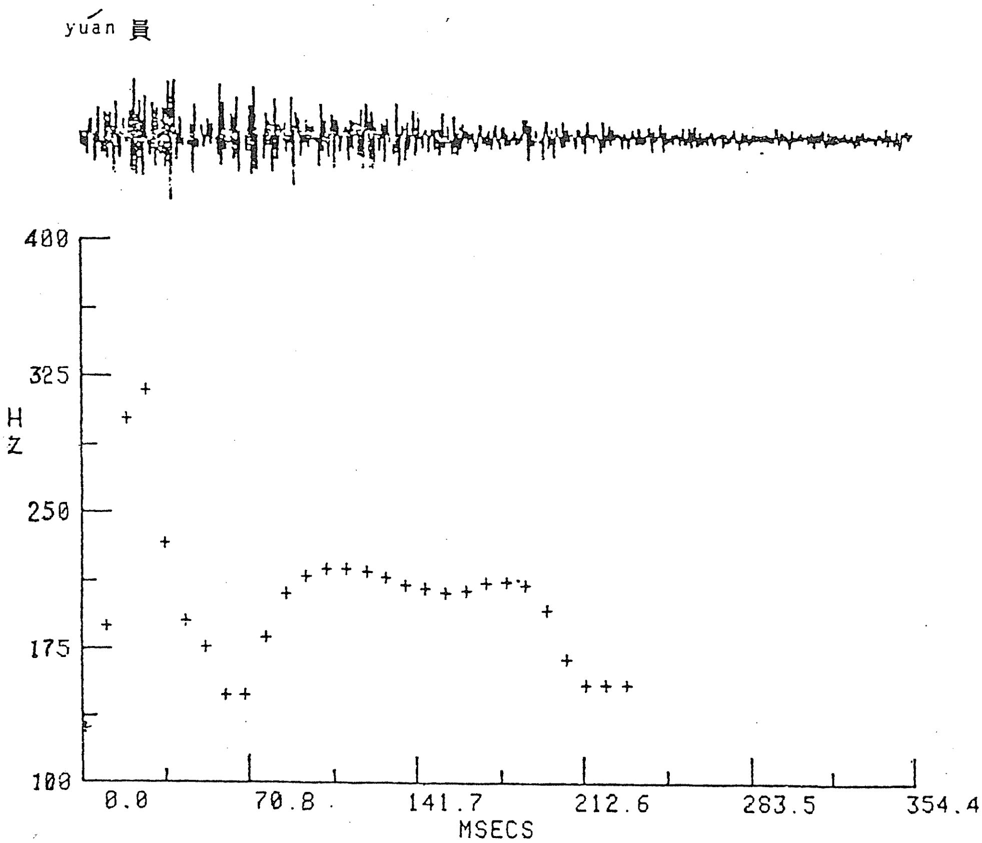


Figure 4.4. Syllable yuan 員 in Mandarin sentence <u>Tāmen shénme dōu jiàu fúwùyuán</u> 'They everyone all call service staff--They call everyone service staff. 'The F0 pattern of this rising-tone morpheme at sentence-final position demonstrates another distinct shape which cannot be identified as any of the tonal patterns in the system.

This particular contour, though not more than 250 msec in duration, displayed a rather complicated and somewhat hat–like shape, which cannot be indetified as having any of the lexical or neutral tone allatones. These examples, i.e., Figures 4.1, 4.2, 4.3, 4.4, have illustrated that acoustically these tokens have demonstrated fundamental frequency patterns that can not be identified as any of the phonologically specified patterns. That is, to score such tokens according to their phonological specification was obviously arbitrary and incorrect.

An acoustic analysis of the test tokens enabled us to look into these test tokens from a different viewpoint. Based on the general tone patterns of each lexical tone and their relative frequency range derived from the production data from the female subjects for the present study (See 2.2.2.4), as well as the neutral tone, we further broke down the test tokens accordingly. For those tokens that demonstrated a distinct fundamental frequency pattern, such as the examples in Figures 4.1 through 4.4, they were classified as not recognizable tones. The analysis yields the following results: Among the 200 test tokens, we found an overall 40.5% correlation between the phonological prediction and the actual phonetic output across the test tokens. A further breakdown of the test tokens yielded that among the unreduced tokens, there was a 55.6% phonological–phonetic correlation. Of the 10% that did not correspond to their phonologically specified lexical tones, on the other hand, 5% exhibited tonal patterns that could be classified as having lexical tones other than their phonologically specified tones and were thus classified by the tones their fundamental frequency demonstrated. And the other 5% exhibited fundamental frequency patterns that could not be specified as any of the distinct tonal patterns and were thus classified as not recognizable tones. Among the 100 reduced tokens, we found a 53% correlation between the phonological prediction and the phonetic output. The analysis showed that 8% of these tokens demonstrated the level tone contour, 3% the rising tone contour, 5% the falling–rising tone contour, 9% the falling tone contour, 50% the neutral tone contour, and finally 25% of not identified contour. We then scored the subjects' performance of the perception test according to the actual phonetic output of each token. The result was as follows: Among the 200 test tokens, we found that 41.7% were correctly identified. For the unreduced tokens, 58.58% was correctly identified, whereas 24.83% of the reduced tokens was correctly identified.

In summary, the results of the perception test were scored under two conditions: the phonological prediction of the tones of each token as well as the acoustic information each token actually possessed.

4.4 Discussion

Results from the perception test manifest the fact that subject's overall performance only varies slightly with respect to different conditions of scoring. In other words, subjects did not behave drastically differently, regardless of whether the scoring condition was by the phonological prediction or by the acoustical analysis of the test stimuli. For the unreduced tokens, since there was a very high correlation (90%) between the phonological prediction and the phonetic output, it was expected that the results of the subjects' identification of tones should not vary drastically under these two different scoring conditions. This might explain why the two different scoring conditions yield similar results, i.e., 55.66% vs. 58.58%. However, the acoustic analysis of the test stimuli among the reduced tokens demonstrated a fairly low correlation (53%) between the phonological prediction and the actual phonetic output. Moreover, 25% of the reduced test tokens did not possess a distinguishable tonal patterns existing within the system. One might expect that the subjects' identification of these tokens would differ due to the gap between the phonological prediction and the phonetic output. On the contrary, the subjects' performance scored by the phonological prediction and the phonetic information was again very similar, 25.42% vs. 24.83%. The general observation here was that subjects performed only slightly above chance level, in this case 16.67%, for the reduced tokens. Nevertheless, we did notice the difference within each condition between the two different kinds of test tokens, namely, the unreduced vs. the reduced tokens (55.66% vs. 25.42% by phonological prediction, 58.58% vs. 24.83% by acoustic information).

Statistical analysis was used to determine whether the subjects' performance was significantly different between the unreduced tokens and the reduced tokens. Two–tailed t–tests for correlated samples were performed between the subjects' performance of the unreduced tokens and reduced tokens for both sets of results. Under the first condition, i.e., the correct identification according to the phonological predication, each subject's performance between the unreduced tokens and the reduced tokens was compared ($t=10.88$, $p<.01$). Under the second condition, i.e., the correct identification according to the acoustic information each token exhibited, $t=8.86$ between the unreduced tokens and the reduced tokens. Results indicated a significant difference between the reduced tokens and unreduced tokens ($t=10.88$, $p<.01$) for phonological comparison and acoustic comparison ($t=8.86$, $p<.01$). In short, the perception test in this study shows that when forced to identify lexical tones of syllables edited from spontaneous speech, that is , when the contextual information had been removed, subjects' identification of unreduced tokens was significantly better than that of the reduced tokens, However, we should note that the subjects' overall performance was relatively poor, compared with the perception study presented in Chapter 2. In other words, the major finding of the perception experiment reported in this chapter is that native speakers' identification of the tones of syllables extracted from spontaneous connected speech is relatively low on the scoring scale.

4.4.1 Fundamental frequency pattern in spontaneous speech

Recall in Chapter 2, the perception test demonstrated that regardless of duration, as long as the required fundamental frequency pattern was present, native Mandarin speakers, as native speakers of other tonal languages, demonstrate a nearly–perfect identification of the lexical tones produced in isolation. However, the perception test presented in this chapter exihibits a different result. When native Mandarin speakers are forced to identify word tones extracted from spontaneous running speech, the performance is far below perfect. Given the effects of intonation contour imposed on a sentence (See Chapter 3 for production data and Chapter 5 for intonation), causing modulation of the individual lexical tones, the fundamental frequency patterns of lexical tones are not maintained in spontaneous speech. Our study on the actual production of tones in spontaneous speech (Chapter 3) confirmed this hypothesis. However, we think that a more detailed discussion of the perception test might illustrate the point more thoroughly.

Consider the unreduced tokens from spontaneous speech first. According to acoustic studies (Chiba, 1936 : 64) and impressionistic observation in earlier linguistic research, the production of such morphemes should, at least to a great extent, maintain and thus manifest the original tonal patterns. Our analysis of the unreduced tokens did exhibit a high correlation (90%) between the phonologically specified tonal patterns and the actual production of such syllables. The phonological specification, as expected, does try to characterize modulation of tones. That is, the falling–rinsing tone might be realized as falling and then somewhat flat instead of rising, in non–final position unless followed by another falling–rising tone, in which case it becomes a rising tone phonetically (See 3.1.1.4). But the results from the perception test presented above definitely did not offer a comparable result. Only 55.66% of such unreduced tokens were correctly identified according to the phonological specification, and only 58.58% were correctly identified even if the scoring was based on the actual phonetic output. Thus there is a sharp contrast between tokens produced in isolation and tokens extracted from fluent spontaneous speech.

The reduced tokens presented even less satisfactory results. By following the predictions of proposed phonological theories strictly, we found a mere 25.42% correct identification. The subjects' performance according to the actual phonetic output did not offer a very different result, namely, 24.83%. Notice that in this case, there existed a rather low correlation (53%) between the phonological prediction and the actual phonetic output. Yet the subjects' performance was equally poor regardless of the scoring condition.

Such results, however, enable us to find a correlation between the production study presented in Chapter 3 and the perception study reported here. The production data demonstrate that when sentences were produced as fluent spontaneous speech in verbal communication, the speakers' production is the consequence of a complicated interaction among lower level phonetics, phonology, and higher levels, namely syntax, semantics, and moreover their knowledge of the language as well as their intention toward such an activity. The phonological specification alone is thus insufficient in characterizing the ultimate production of a sentence. Thus on the surface, it appears that the speakers may very well follow the phonetic and/or phonological rules at one time, and violate some of these rules at other times. However, these phenomena might be a result of the effect of intonation. To what extent is the effect of intonation on tone is not covered within the scope of the present study, but will be pursued in future. The flexibility of people's verbal behavior makes it difficult to predict their linguistic performance, even if we take into consideration all the factors involved. This can be one interpretation of the low phonological–phonetic correlation of the production study. The results of the perception test further support this viewpoint, That is, when forced to identify lexical tones extracted from fluent spontaneous speech, subjects behave rather poorly overall. This may be due largely to the fact that the production of fluent spontaneous speech does not correspond to the production of isolated lexical tones and/or read speech at all, which is not unexpected. The modulation of individual lexical tones in a sentence due to their interaction with the sentence intonation may cause more variation of the fundamental frequency patterns involved. In other words, the fundamental frequency patterns required for lexical tones do not necessarily have to be physically intact in order to maintain their phonemic identity and/or lexicalness. The native speaker makes the decision as to how he intends to produce an utterance, and the native listener extracts information not only from words but also from the context through an integrated knowledge of the linguistic system. Thus we may say that the phonemic/lexical load of tones is reduced in fluent spontaneous verbal communication, due to other interacting linguistic and/or pragmatic factors. In short, there appears to be a trade–off effect among all the information involved in the production of spontaneous speech and the perception of it as well. This further explains that in the production of tones in citation forms, the specified fundamental frequency patterns are always present in a more consistent manner, and these fundamental frequency patterns are in turn essential for the correct perception, too (See Chapter 2). Without the other levels of linguistic/pragmatic information, the phonetics alone has to assume all the responsibility in the encoding processes.

4.5 Summary and conclusion

The perception test demonstrates that without the support of context, native Mandarin speakers identified lexical tones extracted from spontaneous speech much more poorly than their judgment of tones produced in isolation. Comparison of speakers' perception of lexical tones produced in isolation versus in fluent spontaneous speech leads to the following conclusions: In perceiving tones produced in isolation, the fundamental frequency pattern is the primary cue. However, in perceiving individule lexical tones extracted from fluent spontaneous speech, the subjects' rather poor performance might in turn be affected by the absence as well as the distortion of acoustic information. Apparently the context provides enough information for the meaning to get access. Thus the lexical tones in a tonal language do not have to carry the full load of phonemic/lexical contrast, because other linguistic levels of information also contribute to the communication process. The fundamental frequency contour, which was the primary cue both for production and perception of tones in isolated forms, has given way to other interacting linguistic factors in spontaneous running speech. It is still there, but instead of performing the solo, it is now singing in harmony with the chorus.

Chapter V

CHAPTER 5 TONES AND INTONATION

5.1 Introduction

The study of the linguistic consequences of intonation, particularly its sentence delimiting function, traditionally has involved perceptually based transcriptions that are expressed as "pitch " and "stress" patterns. The goal of most of these studies has been the development of a notion for the objective representation of intonational phenomena. The British school phoneticians (Jones, 1909, 1932; Armstrong and Ward, 1926) developed the use of segmental "tunes" to represent intonation[3]. The sequence of tones in an utterance that may rise and fall determines the intonation pattern of the utterance. This group of linguists who prefer to use a tonal representation think that intonation is constrained by the meaning of a sentence and not by syntax (Bolinger, 1972; Crystal, 1969; Stockwell, 1972; Halliday, 1967). Although the intonation system proposed by Halliday (1967) takes into consideration major syntactic constraints, it is meaning, specifically the

---

3 L.e. Armstrong and I. C. Ward, Handbook of English Intonation (1926)
The authors transcribe the over–all intonation pattern by a series of dots and dashes. Each dot marks the relative pitch of an unstressed syllable, and each dash marks the relative pitch of the stressed syllable. The vertical position of the dot or dash reflects the relative pitch of each syllable. The most important characteristic of the system is that two "tunes" are defined. On this pitch with some upward variations on stressed syllables until the end of the sentence, when the pitch falls rapidly. Tune II starts at either a high or a middle pitch and gradually falls, but it ends with rising or level pitch. Sentences are divided into "sense group" and each "sense group" is an intonation group that May have the contour of Tune I or II. In short sentences that have only one "sense group," Tune I is used for statements or imperatives. Tune II is used for some interrogative yes–no questions and for sentences in which the speaker wishes to imply some uncertainly.

information focus rather than syntax, that determines the intonation.

Some analysis of the American "taxonomic" school have applied the segmental techniques of taxonomic phonemics to intonation. The segmental elements have been grouped into suprasegmental "phonemic phrases," "phonemic clauses," "suprasegmental morphemes," as well as phonemic pitch "levels" and "terminals" to describe sentence intonation. Pike (1945) and Trager and Smith (1951) introduced different types of pauses and lexical stress levels, together with the four pitch levels that were used to describe the entire prosody of sentences[4]. This type of representation of intonation has been widely used by American linguisists. In general, attempts have been made to relate these suprasesegmental elements to the constituent structure of the sentence. However, these studies have more or less been occupied with the problem of furnishing "objective" cues for immediate constituent analysis and with providing "morphemes" that will yield a semantic interpretation

---

4 K. L. Pike The Intonation of American English (1945)
Pike describes intonation contours in terms of four pitch levels and two "pauses". He notes that "this number is not an arbitrary one. The four levels are enough to provide for the writing and distinguishing of all the contours that have differences of meaning so far discovered. A description in terms of five or six levels would leave many theoretically possible combinations of pitches unused." (p.26) In Pike's system, number 1 corresponds to the speaker's relative highest pitch level and number 4 to his lowest relative pitch level. There are two "pauses" in this analysis. The "tentative" pause, denoted by [/], "tends to sustain the height of the final pitch of the [preceding] conour" though th[illegible] may be "occasional drift upward." The final pause, [//] "modifies the preceding contour (or contours) by lowering in some sense the normal height of the contour" (p.31). For example: CAMILLE

If Tom goes, 'I will' too
3 – 2–4–3/ 2–4— –4–3 2–4//
'Two, times 'three. 'plus two, is ten
2–4–3/ 4– 2 /– –4–3/ 3– 2–4//

Superficially, the Trager–Smith analysis of the prosodic features resembles Pike's. Trager and Smith use four phonemic pitch levels. Number 1 of a sentence directly from the superficial phrase marker (Lieberman, 1967; 172).

5.1.1 The breth–group theory of intonation

Lieberman (1967) proposed a theory of intonation that claims that the "basic" aspects of intonation are structured by characteristics of breathing. Thus the basic unit for intonation is the "breath–group." The principal point made by Lieberman (1967 : 26) with regard to respiration and speech is that speech is organized in terms of expiratory airflow from the lungs. At the end of each expiration the airflow out of the lungs ceases, and the subglottal air pressure abruptly falls. It has been shown that the fundamental frequency of phonation is directly proportional to the subglottal air pressure. The other parameters that can affect the fundamental frequency of phonation is directly proportional to the subglottal air pressure. The other parameters that can affect the fundamental frequency of

---

corresponds to the speaker's lowest relative pitch and number 4 to his highest relative pitch level. (Notice that numbers are reversed from those of Pike's system.) There are three terminal junctures that correspond to Pike's "two" pauses. The terminal juncture phonetically means a fall in pitch. It corresponds to Pike's "fall pause." The terminal symbols [//] and [/], respectively, correspond to a rise in a pitch and a sustention of pitch. They correspond to Pike's "tentative pause," which either sustained the height of the pitch level that preceded it or involved a "slight drift upwards." One additional juncture, which is transcribed as /+/, is used in this analysis. The principal function of this internal juncture "is to indicate phonetic cues that separate words or parts of certain compound words during fluent speech." Four "phonemic" stress levels also are used. The levels are marked with the following symbols: //.
Primary stress, designated by /'/, is the loudest stress. (Stress is defined psychcacoustically in terms of loudness.) The secondary stress, /'/, occurs only in compound words or phrases in which the internal juncture," /+/, occurs. Secondary stress is thus one of the phonetic cues of the "internal juncture" phenonema. For example:

How + do + they + study / now + we've + got + their + books#?

phonation are the tension of the laryngeal muscles and the neutral position of the vocal cords during phonation. If the tension of the laryngeal muscles remains constant, then the fundamental frequency of phonation will fall at the end of expiration.

Note that the effect of opening the larynx for breathing should not be ignored. The physiological basis of the gradual fall of the fundamental frequency may be a condition of least articulatory control. This baseform delimits a sentence through amplitude and the terminal fundamental frequency. The terminal fall in fundamental frequency is accompanied by the fall of the amplitude of the waveform and the spectrum of the glottal source with increase of duration. The simplest form of the breath–group is a straight line followed by a terminal fall. This prosodic pattern, according to Lieberman (1967 : 27) is characteristic of the ones that are used to delimit the boundaries of unemphatic, declarative sentences in normal speech. However, it is important to note that although Lieberman proposed the above–mentioned "archetypal normal breath–group", he did not specify the exact prosodic pattern before the terminal fall. As a result, the breath–group theory is less strong in its predictive power, and yet more flexible in its characterization of possible intonation contours for simple declarative sentences.

5.1.2 The declination theory of intonation

Several more recent studies on intonation by Maeda (1976), Pierrehumbert (nee Breckenridge), and Liberman (1977), and Pierrehumbert (1979) have proposed the declination effect on simple declarative sentences in American English. Furthermore, these three versions of the declination theory claim that declination results is a near–universal tendency of intonation to drift downward in pitch.

According to Maeda (1976), these "schematized fundamental frequency patterns" are specified by a set of symbols (attributes). The attributes rise and lowering characterize upward and downward rapid movements of the fundamental frequency. A fundamental frequency peak often occurs with rise, and the baseline represents a gradual fall of the fundamental frequency along a sentence. The Pierrehumbert and Liberman version (1977) of the declination theory proposes the topline as contrasted to the baseline.

That is, the topline is derived by connecting all the fundamental frequency peaks of a sentence, whereas the baseline is the connected line of the fundamental frequency valleys. They claim that the topline always declines at a more rapid rate than the baseline, while both characterize the downdrift effect. Sorensen and Cooper (1980) proposed an even more specified declination pattern for declarative sentences. In their version, they specified the topline and bottomline, and introduced a "key point," which predicts the fundamental frequency peak that occurs within a sentence. On the graph of fundamental frequency versus time, the "key point" occurs halfway between the first and last peaks on the frequency axis and one–fourth distance from the first peak to the last peak on the time axis. They do not attach any special significance to the location of the key point in terms of the speaker's production of the sentence. The key point is conceptually useful because it indicates that half the fall between the first and last fundamental frequency peak occurs during the first fourth of the sentence (in Cole, 1980 : 410).

5.1.3 Declination as a discourse effect

Studies of the declination effect by Umeda (1981b, 1982) concluded that the declination effect probably related more to discourse rather than to the sentence level (By discourse we mean conversation centered around one common topic. The change of topic results in the change of the discourse.) In their study, they imposed the declination intonation on synthetic nonsense syllables and asked the subjects to mark sentence breaks. Their results showed that consecutive series of declination intonation caused the subjects severe fatigue, and apparently proved not to be an effective cue for delimiting sentences. They suggested that it might be the case that whenever a speaker introduced a new topic for a discourse, he would be more likely to use the declination intonation, that is, to start a sentence at a rather high fundamental frequency and gradually drifted downward. During the course of this discourse, a speaker would probably prefer to use a flatter type of intonation, corresponding to the archetypal breath–group. Umeda (1982) also conducted production studies addressing the same issue, in which the production data showed that this was indeed the case. In other words, the declination intonation may exist as a device more likely to be employed in speech production for signaling a discourse, since it might be perceptually more prominent in communication.

5.1.4 Chinese tones and intonation

Traditional Chinese philological studies go back to the T'ang period (618–906 A.D.). Numerous works have been directed toward the area of rhyming, which ultimately lead to the research of prosody. Nonetheless, more recent research on the nature of intonation of Mandarin Chinese can best be represented by Chao (1968). Chao (1968 : 40) attempted to characterize the basic intonation patterns in Mandarin. He described the intonation patterns under the assumption that:

"... The main types of intonation described below are found not only in Mandarin, but in almost all dialects. In some respect, intonation varies less even from language to language. It is, however, never safe to assume that any given type of intonation in one language will have the same function in another language."

He classified Mandarin sentences into 13 intonation types. Only one type is directly related to the data here, i.e., the normal intonation, which he defines as "ordinary statements". He maintains that in short sentences of three or four syllables, there is no special intonational modification. The word tones are maintained following the stress rules involved. In longer sentence he states (p.40), "There is a slight tendency for the pitch to trail off to a lower key toward the end." His observations may correspond to the simplest form described by the breath–group theory.

In this section, Mandarin Chinese intonation is studied and reported. The study of Mandarin intonation and tone address a number of interesting questions. First, it should be of interest to investigate how lexical tones interact with intonation. Previous studies of lexical tones by Denes (1959), Denes and Milton–Williams (1962) and Abramson (1970) have demonstrated that the fundamental frequency contours are dominant cues to the perception of tones, although Denes and Milton–Williams (1962) suggested the existence of a complex interaction among the fundamental frequency contour, segmental duration and intensity. In Chapter 2 it was reported that the fundamental frequency patterns are the primary cue to the perception of tones in citation forms. In Chapter 3 the results from the production data of spontaneous running speech showed that only 36.13% of the data was produced with sufficient acoustic information regarding tones. In other words, only 36.13% of the data possessed identifiable tonal contours.

We suggested a complicated interaction in the processes of the production of spontaneous speech that ultimately derived the phonetic output, using lexical stress, word tones, the phonology, syntax, and moreover the speaker's knowledge of the language as well as the information he intends to communicate. The presence of lexical tones should result in a controlled pattern of fundamental frequency variation in the non–terminal part of a sentence. Spontaneous connected speech should show how these controlled fundamental frequency patterns vary as contrasted to tones produced in citation forms (See Chapter 2).

Second, the presence of level tone in mandarin offers a crucial test of the declination theory. If there are two stressed level tones occurring consecutively by the phonological specification, the declination theory should predict that the second level tone will be lower than the first in fundamental frequency.

Third, it had been noted by Armstrong and Ward (1926), and Jones (1932) that the function of intonation lies mainly in mapping sentences in relation to their meaning. However, both studies pointed out the fact that very often the speakers violate the supposed–to–be–most–normal intonation.

Although Sorensen and Cooper (1980 : 407) do state that "unfortunately, neither Maeda, ourselves, or anyone else has yet arrived at a description of a baseline with such strong predictive power", the declination theory claims that the declination pattern occurs universally. The objectives of this chapter is to explore this claim of "near–universality" for Mandarin Chinese.

Production data oı the intonation of Mandarin Chinese declarative sentences was derived and examined to determine how Mandarin sentence intonation can best be characterized. Moreover, it is also an attempt to test the breath–group theory of intonation as well. (For a more detailed discussion of types of declarative sentences in Mandarin, see 3.2.1.1.)

5.2 Data of speech production – spontaneous speech vs. read speech

5.2.1 Methodology

The production data used in Chapter 3 also served as source of data for the present chapter. For ease of reference, we shall review the data briefly. Initinally, the data body consists of fifty simple declarative Mandarin sentences produced by two female native speakers in spontaneous speech (see Appendix 5.1). (For more detailed discussion, see Chapter 3.2.2.). A detailed analysis of the data from spontaneous speech includes the overall intonation contour of each sentence, the unit energy measurement, using the pitch extracting program on a PDP–11/34 computer (See 3.2.1). Figures 5.1. and 5.2. gives examples of a sample sentences from the data.

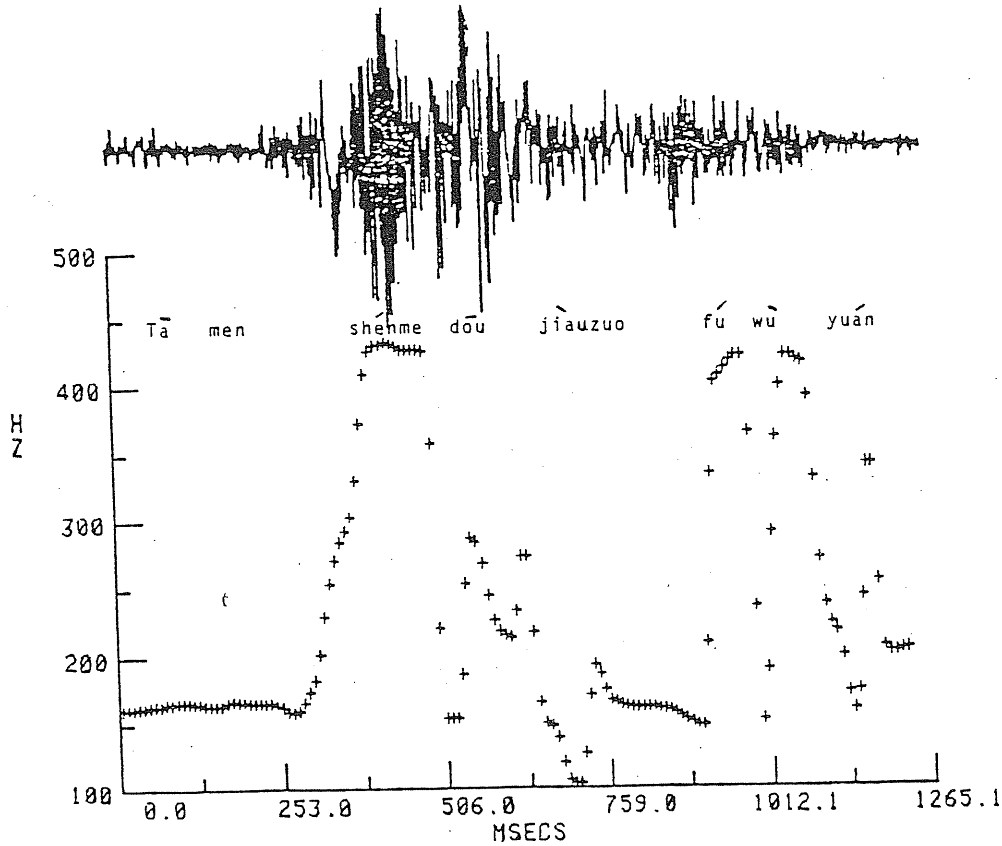


Figure 5.1. Fundamental frequency pattern of Mandarin sentence <u>Tāmen shénme dōu jiàuzuo fúwùyuán</u> 'They everyone all call service staff--They call everyone service staff.

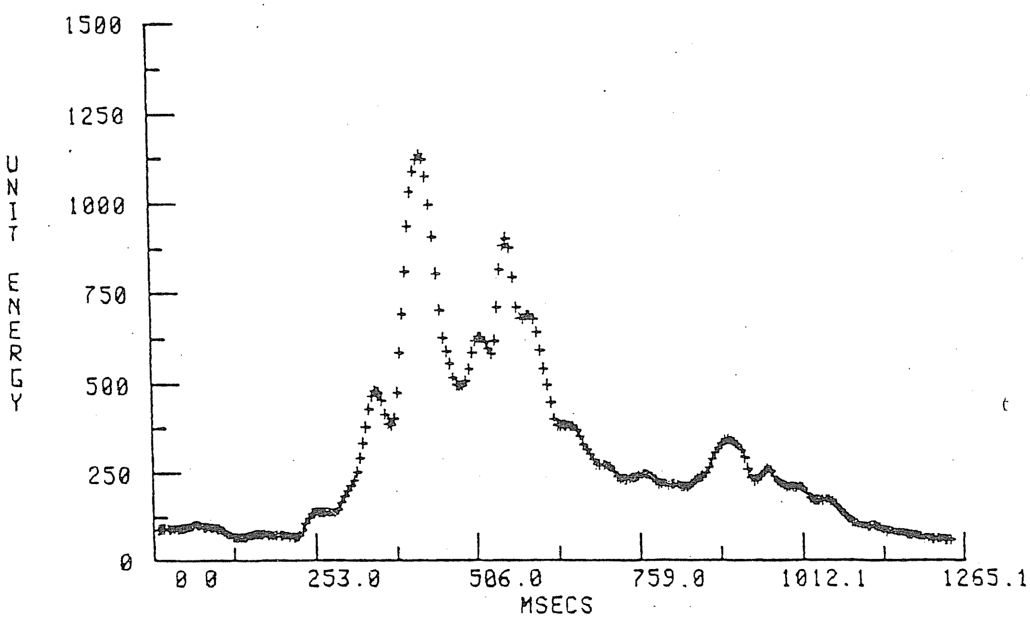


Figure 5.2. Unit energy measurement of Mandarin sentence Tāmen shéme dōu jiàuzuo fúwùyuán They everyone all call service staff--They call everyone service staff.'

However, in order to examine the possible discrepancy between spontaneous speech and read speech, the latter having a tendency to be more enunciated and more regular in its intonation contour, the same 50 sentences were read by the same speakers for later comparison.

5.2.2 Results

5.2.2.1 Measurements of production data of spontaneous speech

The following measurements were made from the graphic display of the fundamental frequency pattern of each sentence. The measurements include the following variables within every sentence: (1) the beginning F0; (2) the highest peak; (3) the lowest valley; (4) the terminal F0; (5) the range of F0 drop between (2) and (3); and (6) the duration of the sentence. Table 5.1. illustrates the range and mean of all the measured variables for the production data included here. Table 5.2 illustrates the same measurements broken down by speakers.

Table 5.1. Measurement of production data from spontaneous speech

	beginning Fo	highest peak	lowest valley	terminal Fo	range of Fo drop	duration of sentences
range	137.4– 379.4 Hz	200.9– 478.6 Hz	102.0– 300.8 Hz	100.0 – 300.8 Hz	69.3 – 315.4 Hz	722.1– 3171.9 msec
mean	224.98 Hz	326.79 Hz	154.85Hz	190.67 Hz	173.2 Hz	1706.76msec

Table 5.2. Measurement of production data from spontaneous speech by speakers

speaker		beginning Fo	highest peak	lowest valley	terminal Fo	range of Fo drop	duration of sentences
COL	range	137.4– 363.55 Hz	200.9– 461.5 Hz	102 – 175 Hz	100 – 250 Hz	69.32– 315.4 Hz	1155.5 – 3071.9 msec
LH	mean	194.14 Hz	219.11 Hz	132.67 Hz	175.49 Hz	187.74 Hz	1832.49 msec
COL	range	166.36 – 379.4 Hz	241.3 – 478.6 Hz	123.04 300.8 Hz	160.54 300.8 Hz	85.05– 279.7 Hz	722.1– 2629.2 msec
LH	mean	255.83 Hz	334.48 Hz	177.03 Hz	205.84 Hz	158.75 Hz	1481.04 msec

5.2.3 Analysis of Results

The intonation contours of the selected fifty declarative sentences were examined in order to see how they can best be represented. The fundamental frequency pattern of each sentence was fit to different models of intonation and the results were then calculated. They were tested to see whether they fitted (1) the declination theory, (2) the breath–group theory, or (3) neither theory. Due to the fact the physical correlates of the declination theory vary in the models of different formulations, the following criteria were derived. If the intonation pattern of a sentence fitted Maeda's (1976) version of declination, i.e., a number of characteristic fundamental frequency movements superimposed on a straight falling baseline, it was accepted as having the declination effect (DM). If the intonation pattern of a sentence fitted the Breckenridge (nee Pierrehumbert) and Liberman (1977) version, i.e., a topline that connected the fundamental frequency peaks in a sentence declines at a faster rate than a line connecting the fundamental frequency valley, it was accepted as having the declination effect (DBL). The Sorensen and Cooper(1980) version of the declination theory not only specifies the topline and bottomline, but also further specifies the keypoint on the graph of fundamental frequency versus time (DSC). The topline rule specifies that this keypoint occurs halfway between the first and last peaks on the frequency axis and one–fourth distance from the first peak to the last peak on the time axis (in Cole ed., 410). The Topline Rule claims that the line drawn through this "key point" and the last fundamental frequency peak of the sentence constitutes an excellent model of the speaker's Fo topline. Moreover, half of the fall between the first and last Fo peak occurs during the first fourth of the sentence.

On the other hand, if the fundamental frequency contour of a sentence falls only at its end, it was accepted as having the intonation contour that fit the breath–group theory (B–G). Of course the breath–group theory is less powerful in terms of its predictiveness than the declination theory, since it does not specify the fundamental frequency pattern of the non–terminal portion of a sentence. On the other hand, the declination effect can be accounted for as a special case of the intonation encompassed by the breath–group theory, because of the specification of the terminal fall of the coutour.

Finally, sentences whose fundamental frequency patterns fitted none of these theories constituted a third group.

5.2.4 Analysis and Discussion

Table 5.3. illustrates the results of the analysis. 4% of the sentences fit Maeda's version (1976) of the declination theory (DM), having a straight falling baseline. 14% of the sentences fit the Breckenridge and Liberman (1977) version of the declination theory, adding the topline as one more criteria. 16% of the sentences fit the Sorensen and Cooper (1980) version of the declination theory, which met the Topline Rule and the keypoint. Overall, if we consider the sentences having the declination effect as one category, only 20% of the data can be characterized by the theory. (Notice that the figure 20% did not derive from the summation of the sentences fitting any version of the declination theory.) On the other hand, 54% of the data cannot be characterized by the declination theory while it can be characterized by the breath–group (B–G) theory. Moreover, because we consider the declination effect a special case of the breath–group intonation, the overall percentage of the production data that can be characterized by the breath–group theory is thus 74%. Finally, 26% of the production data possessed intonation contours that cannot be accounted for by any of the proposed theories. Examples will be presented in the following discussion to illustrate each of the above–mentioned cases.

Table 5.3. Intonation analysis of production data

proposed intonation theory	percentage
DM	4%
DBL	10%
DSC	16%
B–G	74%
neither	26%

5.2.4.1 Intonation of sentences that can be characterized by the declination theory

Among the selected sentences, only two sentences, that is, 4% of the data, can be characterized as having the declination pattern specified by the baseline (Maeda, 1976). Figure 5.3. presents the intonation contour of one of these two sentences.

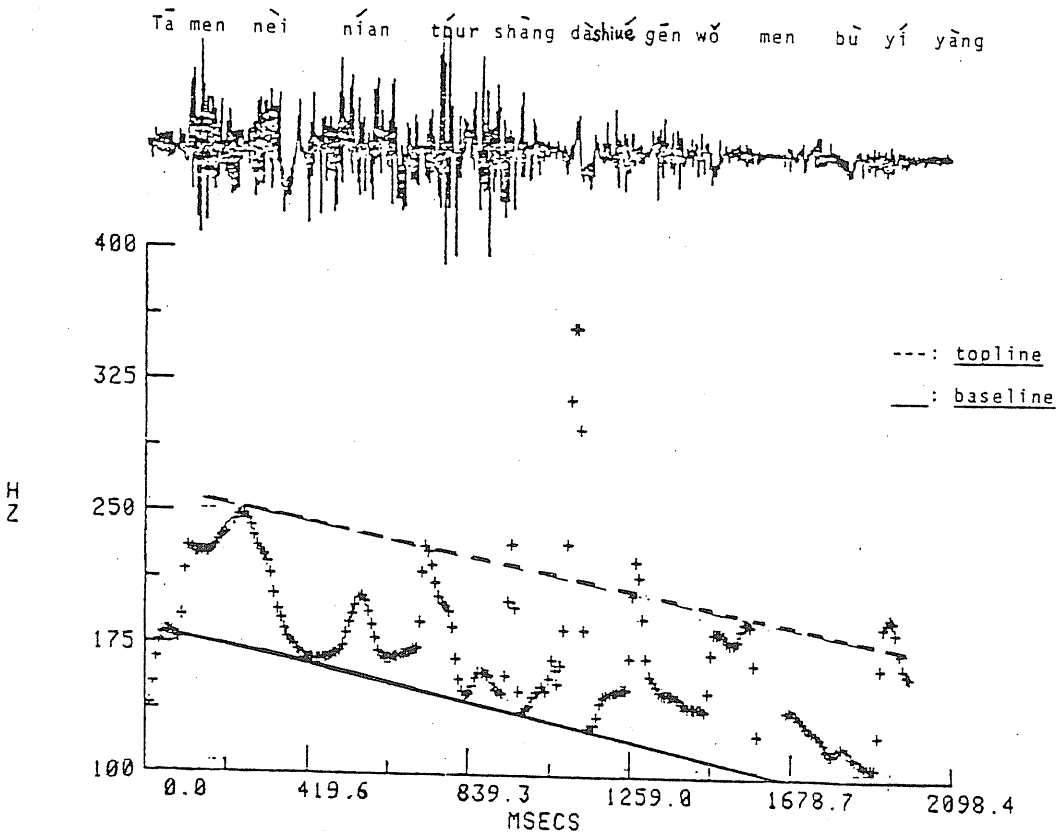


Figure 5.3. Intonation contour of Mandarin sentence Tāmen nèiniántóur shàng dàshiué gēn wǒmen bù yíyàng 'Things were different from the way it was for us when they went to college during those days.' The intonation contour fits both the baseline (Maeda, 1976) and the topline (Breckenridge and Liberman, 1977)

Figure 5.3. demonstrates the intonation contour of a Mandarin sentence which can be characterized by the baseline. The sentence was tāmen nèiniántóur shàngdàshiué gēn wǒmen bù yíyàng 'They in those days going to college with us not the same —— Things were different from the way it was for us when they went to college during those days.' By connecting the valleys of the fundamental frequency contour of the sentence, a baseline can be derived to characterize the overall intonation contour of the sentence. However, the same intonation contour can not

be characterized by the topline proposed by Breckenridge and Liberman (1977), since it does not decline at a faster rate than the baseline.

On the other hand, 10% of the data can be characterized as having the declination pattern specified by Breckenridge and Liberman (1977). This type of sentence can be characterized both by the topline connecting the fundamental frequency peaks and the bottomline connecting the fundamental frequency valleys, with the topline declining at a faster rate than the bottomline. Figure 5.4. shows the intonation contour of one of these sentences.

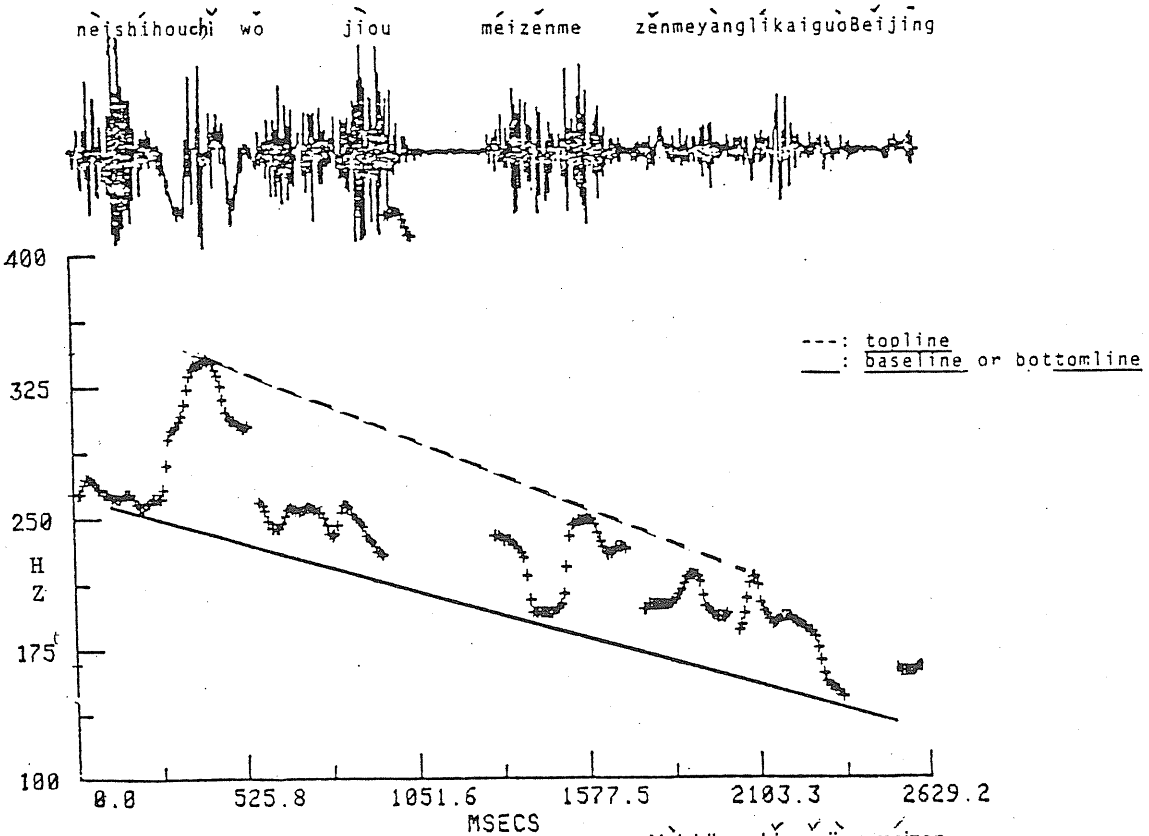


Figure 5.4. Intonation contour of Mandarin sentence Neishihouchi wo jiou meizen zenmeyang likaiguo Beijing 'From that time on I didn't quite ever leave Beijing.' The intonation contour here fits every version of the declination theory, i.e., the baseline (Maeda, 1976), the topline and bottomline (Breckenridge and Liberman,1977) and Topline rule (Sorensen and Copper, 1980).

The sentence in Figure 5.4. is neishrhou chi wo meitzenmeyang likaigue Beijing 'That time on I just not–so–much left Peking —— From that time on I didn't ever quite leave Peking.' The intonation contour of this sentence can be characterized by both the topline and the bottomline and the topline indeed declines at a faster rate than the bottomline.

Moreover, 16% of the sample sentence can be characterized by the Sorensen and Cooper (1980) version of the declination theory. That is, the Topline Rule discussed above represents the overall model of the intonation contour of a sentence, and that half the fall between the first and last fundamental frequency peak occurs during the first fourth of the sentence. Figure 5.5. is an illustration of this type of intonation pattern.

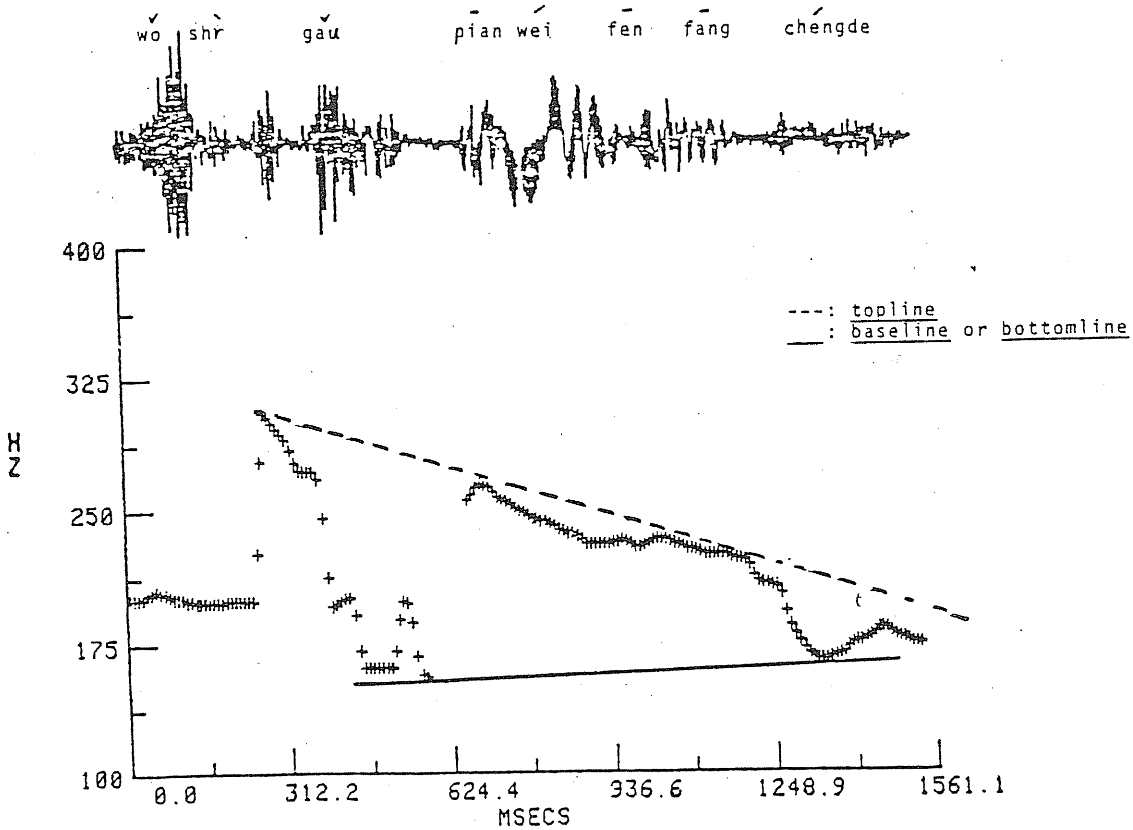


Figure 5.5. Intonation of Mandarin sentence Wo shr gau pianweifenfangcheng de 'My research is on partial differential equations.' This sentence's intonation fits the Sorensen and Cooper (1980) version of declination, i.e., the Topline rule, and the F0 peak occurs at the first quarter into the sentence. However, note that the baseline (or bottomline) results in an almost horizontal line.

The sentence in Figure 5.5. is wo shr gau pianweifenfangchengde 'I copula work partial differential equation. —— My research is on partial differential equations.' This sentence not only exhibits an overall declination, but also a fall of the fundamental frequency of 152.87 Hz. This fall of the fundamental frequency occurred, as the Sorensen and Cooper model predicted, within the first fourth of the sentence, and further it is also more than half of the fall between the first and last fundamental frequency peak.

From the above reported results, it appears that different versions of the declination theory vary in their specification of the proposed declining effect for simple declarative sentence, as well as in their application to spontaneous speech. Maeda's baseline version (1976) seems to least capture Mandarin sentence intonation (4%), the Breckenridge and Liberman (1977) version captured a mere 10% of the data, and the Sorensen and Cooper version at a low 16% as well.

We have previously mentioned that sometimes a sentence can be characterized by only one version of the declination theory, such as the sentence presented in Figure 5.3. This sentence can be characterized only by the baseline proposed by Maeda. The sentence presented in Figure 5.5. on the contrary can only be characterized by the Sorensen and Cooper version. The attempt to draw the base or bottomline by connecting the fundamental frequency valleys of the sentence would result in a near horizontal line, which in turn excluded the baseline version (Maeda, 1976) and the topline–bottomline version (Breckenridge and Liberman, 1977) of the declination theory. However, there existed sentences in the data whose intonation contour, declining to begin with, can be characterized by any version of the declination theory. The sentence presented in Figure 5.4. is an example of this kind. It not only fit the baseline version, topline–bottomline version but also the

Topline Rule (Sorensen and Cooper, 1980) that half of the fall of the fundamental frequency occurred during the first one–fourth of the sentence.

As discussed in 5.1.4., the presence of lexical tones should result in controlled patterns of fundamental frequency variation in the non–terminal part of a sentence, as well as modulation of the individual tones involved (See Chapter 3). If the intonation contour is superimposed on a sentence, we might expect that the intonation can sometimes override lexical tones. The interaction between tones, in this case, the stressed words of a sentence as well as the stressed morphemes of a word and the sentence intonation, a steady declining slope through the entire sentence could roughly be described as follows: (1) The level tone may fall. Nonetheless, it should be less abrupt a fall than the falling tone. The crucial test, as discussed in 5.1.4., would be two stressed level tones occurring consecutively. By the production of the declination theory, the second level tone should be lower in fundamental frequency than the first one. (2) The rising tone, although not falling in its fundamental frequency pattern, however would not rise as high. (3) The falling–rising tone would, as expected, result in more variability because the rising portion of it would somehow not rise and possibly become low level in fundamental frequency (See 3.1.1.4). (4) The falling tone would fall in fundamental frequency, as it normally does.

Among the 20% of the production data which possessed the slope–like declination pattern, the modulation of lexical tones was more or less the case as the above description. A close examination of Figure 5.5. illustrates the point. The sentence was wo shr gau pianweifenfangchengde 'My research is on partial defferential equations'. The stressed words in the sentence were shr 'copula' and pianweifenfangcheng 'partial differential equations.' If the polysyllabic word pianweifenfangcheng 'partial differential equation' were to be produced in isolation, the only morpheme in this word whose tone might be reduced is wei in weifen. However, the tone neutralization for wei is optional in this case. In other words, this word can be pronounced either with every lexical tone present, i.e., pianweifenfangcheng or pianweifenfangcheng. Theoretically, none of the level tone was reduced. This resulted in a series of level tones occurring consecutively, followed by a rising tone morpheme cheng. The graphic display of the intonation of this sentence, presented in Figure 5.5, shows the overall downdrifting. In this instance, the fundamental frequency pattern of the last syllable de, a neutral tone suffix, was a slight fall. In other words, the consecutive level tones in the word pianweifenfangcheng 'partial differential equation' followed the declination slope of the superimposed intonation contour. Thus although both morphemes fen and fang had level tones, the fundamental frequency pattern of fang, which occurred after fen by approximately 12.5 Hz, was lower than that of fen, as the declination theory predicted, although it is not certain whether the Fo fall of 12.5 Hz is really significant. Moreover, the other morphemes of the sentence also followed the downward declination slope. The sentence began with the word wo 'I', which happened to be unstressed in the sentence. Thus, instead of its original lexical tone which is falling–rising, it was produced with a low level (179.32 Hz) and somewhat flat fundamental frequency pattern. The word wo 'I' was followed by copula shr,which happened to be the most stressed morpheme of the sentence. This morpheme shr was produced at a sudden shift from low fundamental frequency of wo 'I' at 179.32 Hz to 307.72 Hz, following by a rapid fall of the fundamental frequency to 268.8 Hz. The word following shr 'copula' was gau 'do' whose lexical tone was falling–rising. However, because it occurred at an unstressed position in the sentence, its tone was not realized, and it was produced at even lower fundamental frequencies than the sentence–initial wo 'I'. That is, the word gau 'do' was produced at about 178 Hz to a low 154.85 Hz, back to 178 Hz, and suddenly going down to 150 Hz. The fundamental frequency pattern of this word became difficult to describe in terms of its shape, as we have previously discussed.

In contrast, examples of sentences occurred when level tone did not seem to be influenced by intonation. Figure 5.6 illustrates an example of this kind.

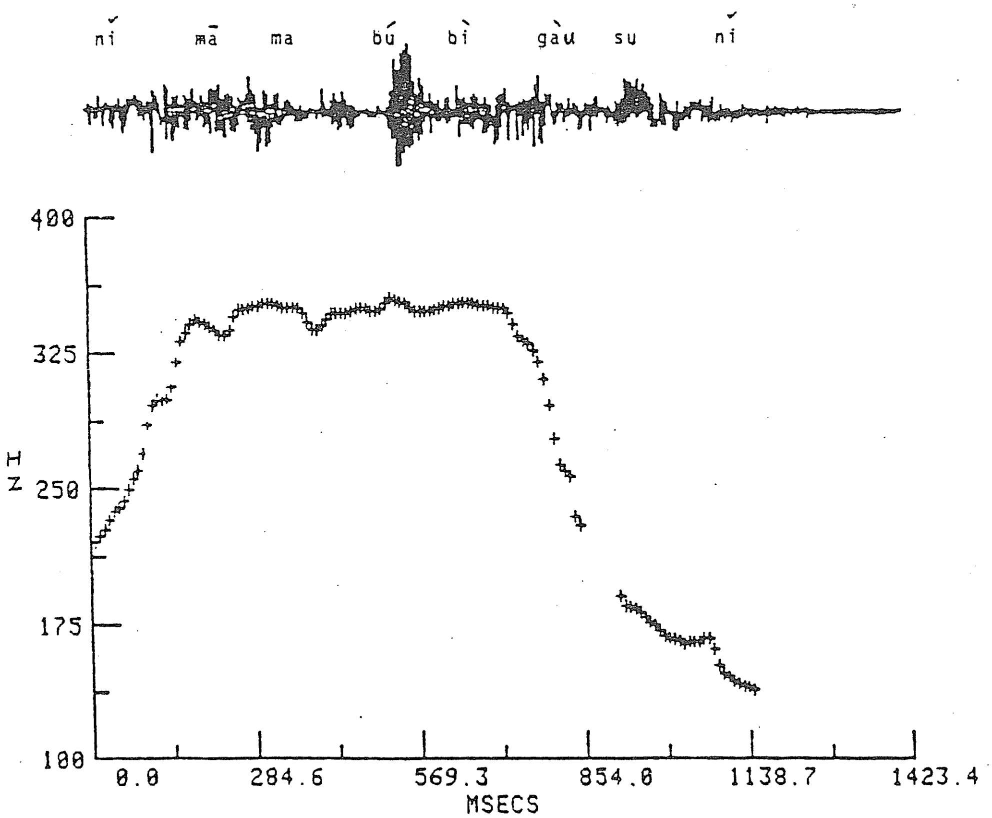


Figure 5.6. Intonation of Mandarin sentence <u>Nǐ māma bú bì gàusu nǐ</u> 'Your monter doesn't have to tell you.', showing that only the breath-group theory of intonation spcifying the terminal fall can characterize this sentence.

The sentence in Figure 5.6. is nǐ māma búbì gàusu nǐ 'Your mother not necessary tell you. —— Your mother doesn't have to tell you.' Notice that the word māma 'mother' should be pronounced as one level tone followed by a neutral tone, as the word formation rule (See 3.1.1.3) predicts that the stressed syllable is the first one in the bi–syllabic word. Nevertheless, in casual speech, it is often pronounced as two stressed level tones, i.e., māmā 'mother'. From Figure 5.6., we see that the intonation contour did not begin to fall until the word gàusu 'tell'. Within the morpheme gàu in gàusu 'tell', the fundamental frequency dropped from 357.8 Hz to approximately 210 Hz. Not only the word māma 'mother' was produced as a high–level plateau at 357.8 Hz, but also the two syllables afterwards. That is, the morphemes bú 'neg. prefix' and bì 'necessary' remained on the same plateau. We thus see no clear rising of fundamental frequency for the morpheme bù, whose lexical tone is rising. Similarly a flat contour is produced for the morpheme bì 'nessary' despite the fact that the lexical tone should be falling. If this sentence followed the same declination contour as the sentence presented in Figure 5.5., the overall intonation contour should definitely begin to fall no later than the beginning of the morpheme bú 'Neg. prefix'. Furthermore, the lexical tone of bì 'necessary', in this case the falling tone, should definitely be present physically. Thus declination theory could not account for cases as such, which is a simple declarative structure without particular emphasis.

We have also noticed among the data that 88% of the lowest fundamental frequency valleys of the sentences were actually lower on the frequency scale than the terminal fundamental frequency of the same sentences (88% of Fo valley terminal Fo fall). This would lead the baseline to go below the terminal fundamental frequency fall of a sentence. That is, if the baseline (bottomline) was to connect all the fundamental frequency valleys, which must go through the actual lowest fundamental frequency valley of a sentence, then such lines of these sentences would go below the terminal fall of the sentences. If the <u>baseline</u> or (<u>bottomline</u>) was to connect the first fundamental frequency valley and the terminal fundamental frequency fall, then these lines derived from such sentences would not capture the actual lowest fundamental frequency of the sentences. This example is further evidence that the declination theory is unable to predict the intonation contour of a simple declarative sentence.

Recall that only 20% of the production data can be described by the declination theory. This suggests that the intonation contour proposed by the declination theory may not be the most dominant intonation pattern in Mandarin Chinese. The evidence presented here i.e., that only 20% of the sentences can be characterized by the declination theory, further suggests that the claim of the declination effect as a near–universal property be seriously questioned. The strong claim made by the declination theory thus requires some modification, as well as the theory itself. The declination theory failed the test because the proposed declining pattern did not seem to capture the general intonation pattern of simple declarative sentences in Mandarin. The modification of the declination theory apparently has to go beyond the refinement of the proposed pattern only.

Furthermore, we are not quite certain whether it is the case that the intonation contour superimposed on a sentence always overrides the lexical tones, whatever the proposed intonation theory might be. Or the reverse also exists, namely, the lexical tones may sometimes override the intonation contour. More discussion regarding this particular question will be presented in the next two sections (5.2.4.2. and 5.2.4.3.).

5.2.4.2 Intonation of sentences that can be characterized by the breath–group theory

As discussed in 5.1.1., the breath–group theory (Lieberman, 1967) proposed a theory of intonation based on breathing. The simplest form of the breath–group, the archetypal normal breath–group is a straight line followed by a terminal fall. However, no specific prosodic pattern before the terminal fall was presented. Compared with the declination theory, the breath–group theory has less predictive power, offers no claims of universality, but apparently is more flexible because it specifies nothing but a terminal fall. We have already mentioned that the declination effect could be accounted for as a special case of the breath–group intonation. As to how the breath–group theory might be applied to tone languages, Lieberman (1980, personal communication) suggested that in Mandarin Chinese the fundamental frequency variations apparently seem to be interpreted as acoustic manifestations of segmental tone features interacting with the sentence intonation's fundamental frequency pattern, which echoes Chao's description (Chao 1968 : 39) of intonation in Mandarin. The presence of those segmental features may affect the fundamental frequency function of the entire breath–group.

The criteria based on which a sentence from the production data was said to have the breath–group intonation pattern were as follows: First, a terminal fall of the fundamental frequency was required. This would include all the sentences whose intonation can be characterized by the declination theory. Due to the presence of the lexical tones, some interaction between the terminal falling intonation and the lexical tones was expected. Although we have not yet been able to derive an algorithm to predict the exact modification of lexical tones occurring at sentence–final position, we did expect that the falling tone remained more or less unchanged, and that the other three tones, namely, the level, rising and falling–rising tones underwent some modification in their fundamental frequency pattern. That is, the level tone would either remain level and would somehow fall toward the end, although it would be less abrupt of a fall than the falling tone. The rising would not rise as high in its fundamental frequency pattern, but it would not fall. The falling–rising tone, as discussed in 3.1.1.4, would lack its rising portion at non–sentence–final position, thus become a falling followed by a plateau or possibly more complicated variations at sentence–final position, due to the interaction with the terminal fall of the intonation.

Second, as we have already discussed in 5.1.1., the breath–group theory does not specify the intonation pattern before the terminal fall, thus it has more flexibility in accommodating the intonation of the pre–sentence–final portion.

Third, the unit energy measurement of each sentence was examined as to see whether a normal breath–group was present, because this is an acoustic correlate of the breath–group. At the end of the breath–group both the laryngeal muscle tension and the subglottal air pressure have diminished, which results in low pitch. A permitted modification of the normal breath–group is an increase in the tension of the laryngeal muscles at the end of the breath–group. If the subglottal air pressure were constant, the acoustic correlate of the modified normal breath–group, called the marked breath–group, would always be a rising fundamental frequency contour. We suspect that this type of modification of the breath–group is commonplace in Mandarin. If there was a sudden and abrupt fall of energy toward the end, sometimes followed by a gradual fade–out of the energy, it was said to have a normal breath–group pattern. Figure 5.7. is an example of a normal breath–group in unit energy measurement.

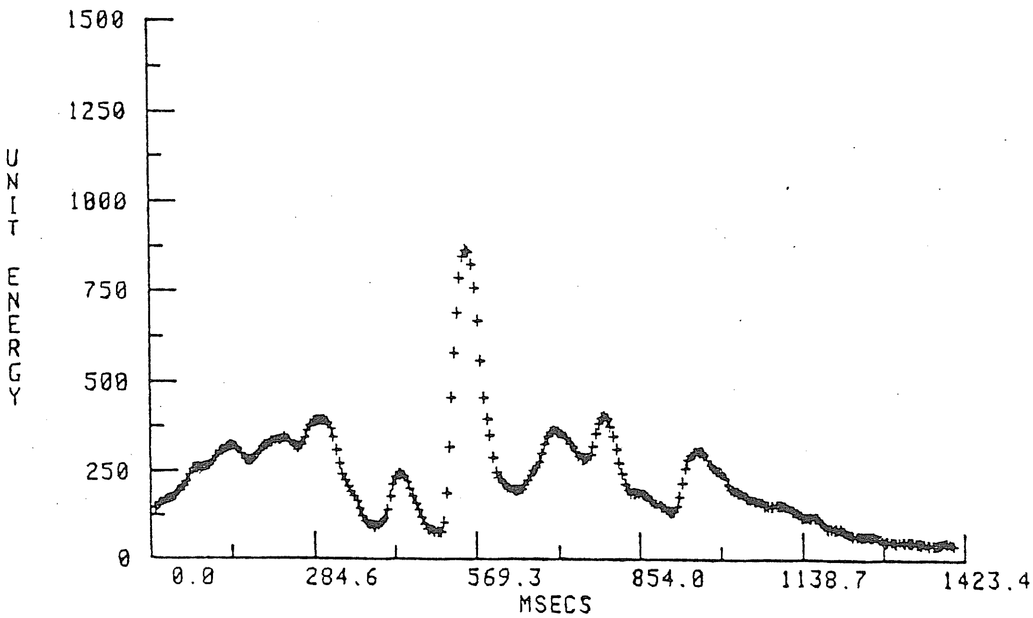


Figure 5.7. Unit energy measurement of mandarin sentence Nǐ māma bù bì ghàosu nǐ 'Your mother doesn't have to tell you.' The intonation contour is shown in figure 5.6. the unit energy measurement shows a normal breath-group.

Examples of sentences in the production data that can be characterized by the breath–group theory will be presented. Except sentences ending with a falling tone which lead to the terminal fall automatically, sentences ending in level, rising and falling–rising tones will be discussed. Figures 5.8. through 5.10. examlify these three different cases of breath–group intonation.

However, another sentence ending theoretically in the level-tone exhibited a different result (Figure 5.8.). In this case, the level tone was modified by the terminal fall of the intonation, showing a falling fundamental frequency pattern.

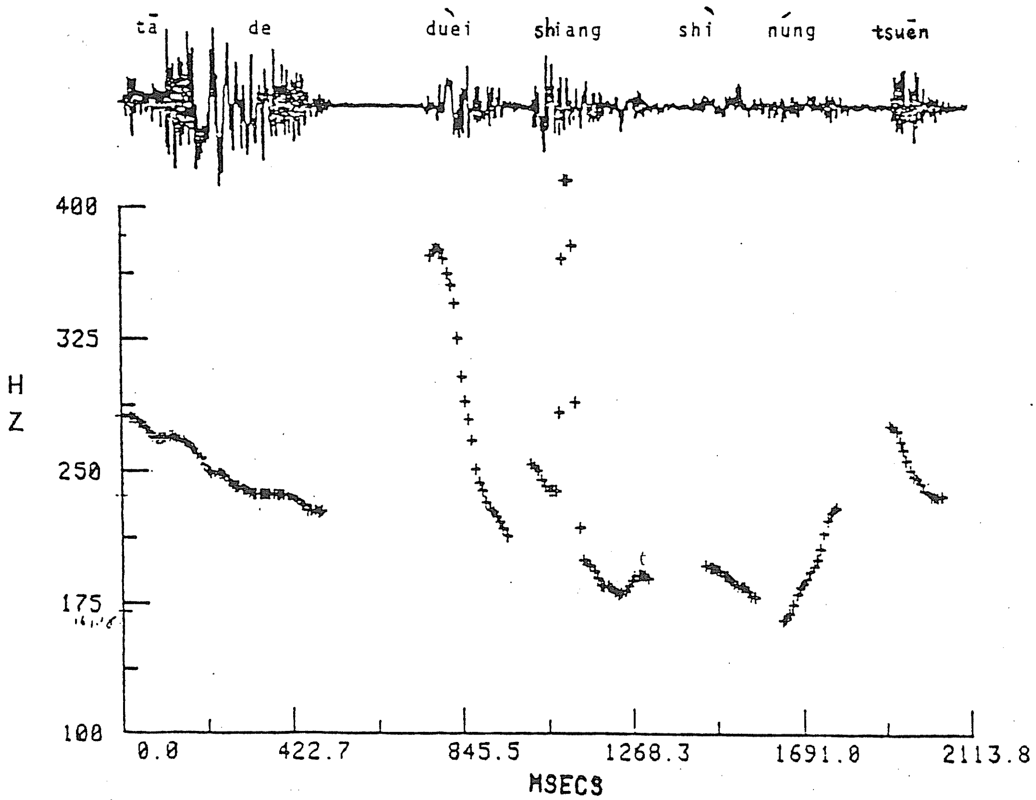


Figure 5.8. Intonation contour of Mandarin sentence Tā de duèishiang shì núngtsuēn 'Its object is village--The main concern of the project was the farming villages'. Note that the sentence ends with a level tone. The terminal fall of the intonation overrides the lexical tone.

The sentence in Figure 5.8. is <u>tāde duèishiang shr̀ núngtsūen</u> 'Its (in this instance, a particular project) object copula village —— The main concern of the project was the farming village'. For the last two morphemes of the sentence, i.e., the last word núngtsūen 'village', the rising tone of the morpheme núng 'agriculture' was clearly produced moving from 169.18 Hz upward to 216.4 Hz. However, the level tone of the morpheme tsūen 'village' was produced in a clear falling fashion, that is, starting at 287.5 Hz with a downward movement, ending at approximately 233 Hz.

Figure 5.9. is an example of a sentence theoretically ending in mid–level tone, as contrasted with the high–level tone called the level tone throughout this study.

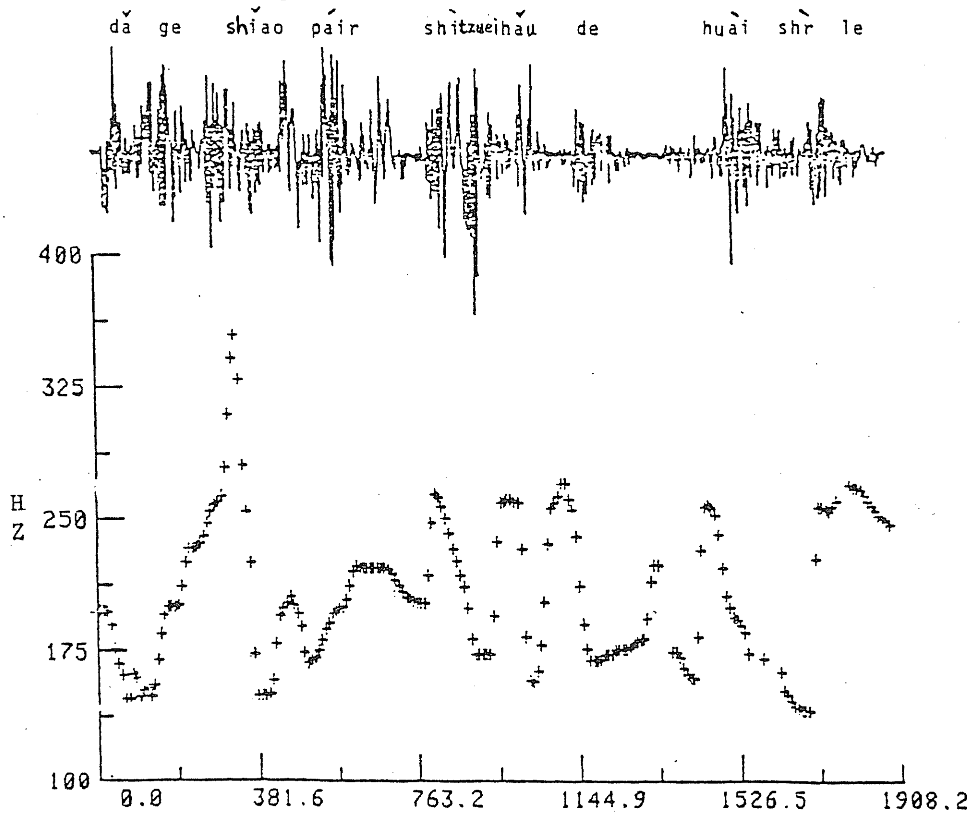


Figure 5.9. Intonation contour of Mandarin sentence Wǒmen dāngshŕ shr̀ dàu jílin 'At that time we were to go to Jilin.', showing a sentence ending with a rising tone. The terminal fall of the intonation overrides the lexical tone.

The sentence in Figure 5.9. is women dangshr shr dau Jilin. 'We then copula go Jilin —— At that time we were to go to Jillin.' The last word Jilin was sometimes pronounced with two stressed rising–tones. However, it should be noted that phonologically the word Jilin was pronounced with a rising tone followed by a neutral tone which was more of a pronunciation from the Peking area. We were interested in the production of the morpheme Ji which was the stressed morpheme of the word. The graphic display showed that the rising did take place, i.e., from 196.6 Hz to 261.5 Hz, then moving downward for the neutral tone morpheme lin. The interesting phenomenon here is that the neutral tone morpheme lin exemplified a falling of the fundamental frequency to 154.84 Hz. In other words, from the rising tone morpheme Ji to neutral tone lin which should be mid–level tone theoretically, the fall of the fundamental frequency was in fact 106.66 Hz (261.5 Hz —— 154 Hz).

Figure 5.10. is an example of a sentence ending theoretically in a falling–rising tone.

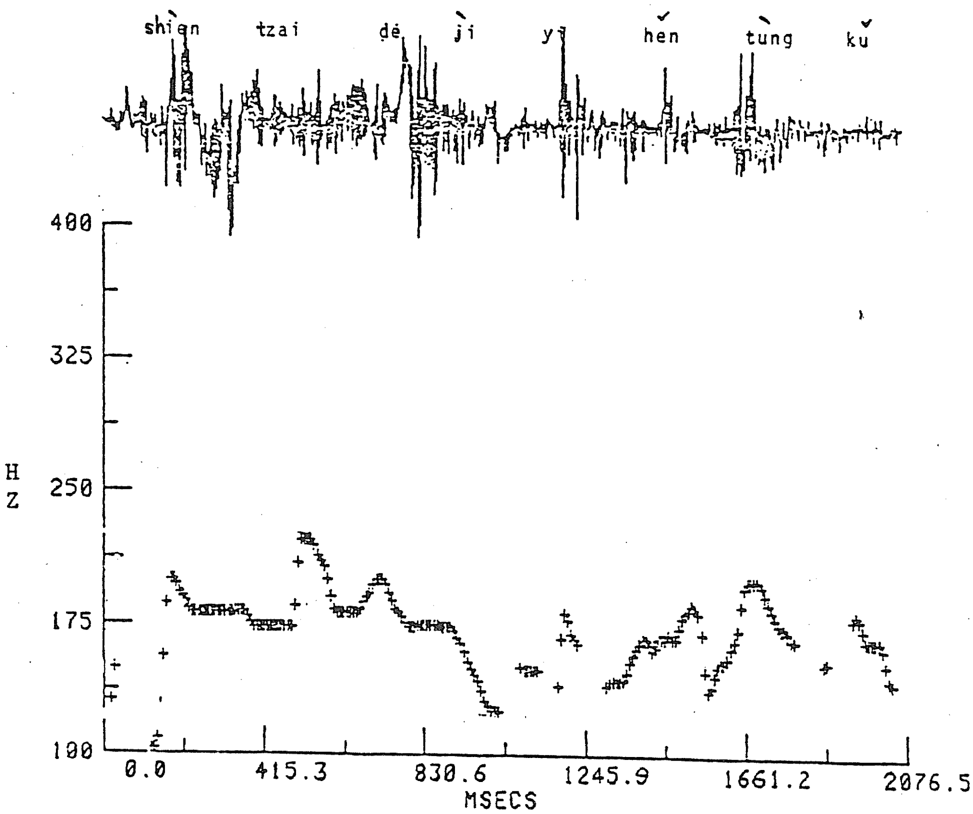


Figure 5.10. Intonation contour of Mandarin sentence <u>Shièntszide jìyi hěn tùngkǔ</u> 'Now (as I recall) the memory is rather painful.', showing a sentence ending in falling-rising tone. The terminal fall of the intonation overrides the lexical tone.

The sentence presented in Figure 5.10. is shiàntzai jìyi hěn tùngkǔ 'Now memory very painful —— Now as I recall the memory is rather painful'. It is guite clear that in this case no version of the declination theory can account for its intonation contour, except that the peak of the fundamental frequency occurred during the first one–fourth of the sentence. No approximation of a declination slope could represent this contour. On the other hand, if we were to look for nothing but an effect of a terminal fall of the fundamental frequency, we could then fit this sentence into the model proposed by the breath–group theory. Notice that the last word of the sentence is tùngkǔ 'painful', which consisted of a falling tone followed by a falling–rising tone. The falling tone of tùng was present graphically on the figure as we expected. It descended from 203.8 Hz to 169.24 Hz on the frequency scale. The falling–rising tone of kǔ, being the last syllable of the sentence, only exhibited a falling contour. However, the falling contour of kǔ descended on the frequency scale from 186.54 Hz to 143.2 Hz, which means that the beginning and ending tonal contour of the morpheme kǔ were both lower than the beginning and ending fundamental frequency of the preceding morpheme tùng, as their respective lexical tones would be on the frequency scale. As the breath–group theory predicted, the terminal–fall of the intonation caused the interaction with lexical tones to modify their fundamental frequency patterns. Notice also that the lexical tones in this sentence were somewhat identifiable which suggests that this could be a case where lexical tones override the intonation contour.

The above reported examples provide support for the breath–group theory of intonation. The reader should also have noticed by now that none of the examples presented in Figures 5.8. through 5.10. can be characterized by the declination theory. In short, none of the version of the declination theory can offer an explanation of these intonation contours, because none of these sentences possessed a steady downward slope–like intonation contour. Moreover, the interaction between tones and intonation appears to be quite complicated. In the data, we found that sometimes it was the case that the intonation contour overrode the lexical tones, whereas the opposite also occurred. That is, sometimes it was the case that the lexical tones seemed to override the intonation contour. This suggests that the speaker may freely choose either form of performance in their production of sentences. The choice, at this point, seems to be random.

So far we have discussed sentence intonation patterns that can be characterized either by the declination theory or the breath–group theory. In the data analyzed, there were also sentences whose intonation could be characterized by neither of the proposed intonation theories. The next section will be discussion of this type of sentences.

5.2.4.3 Intonation of sentences that can be characterized by neither the breath–group theory nor the declination theory.

We have mentioned in the discussion of 5.2.3. that in the processes of analyzing the production data for the present intonation study, we came across a group of sentences that neither the declination theory nor the breath–group theory can account for. This group of sentences was represented under the title of neither in Table 5.3., and the resulted showed 26% of the data fell into this category. To group these sentences into one category, however, did not in any sense imply that they share any common intonation features. In other words, we do not suggest that these sentences form any pattern for another potential intonation theory. We are merely stating the fact that the existence of the intonation of these sentences offered evidence for the inadequacies of both of the proposed intonation theories, i.e., the declination theory and the breath–group theory. The following discussion is an attempt to characterize these sentences by roughly dividing them into three different types. The first type contained sentences whose terminal amplitude rose before it came to a fall. In other words, the sentences were not produced with the "archetypal breath–group intonation (See 5.1.1.). The irregularity of the amplitude might very well be one of the main reasons for the failure of the terminal fundamental frequency to fall. Figures 5.11. and 5.11.a are an example of this type.

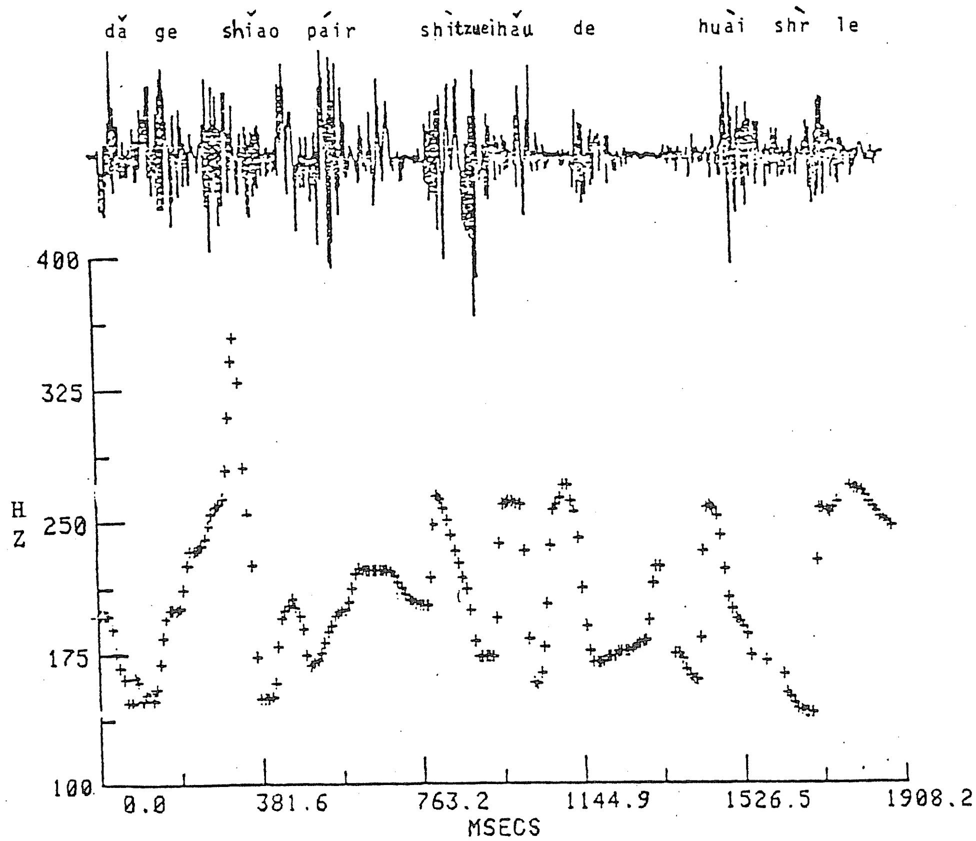


Figure 5.11. Intonation of Mandarin sentence that fits neither the declination theories nor the breath-group theory. the sentence is Dǎge shǐaopáir shì tzuèihǎude huàishr̀le 'Play nahjiang copula the best bad thing--Playing the nahjiang was probably the best among the existing bad habits.

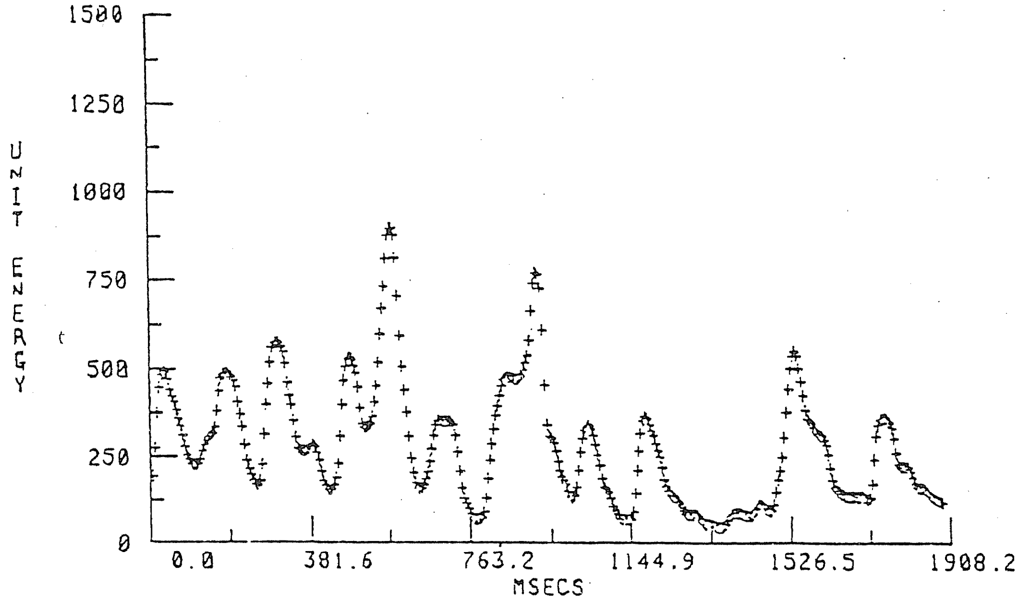


Figure 5.11.a. Unit energy measurment of Mandarin sentence shown in figure 5.12. Here the amplitude of the sentence shows a normal breath-group.

The sentence in Figure 5.11. is dage shiaupair shr tzueihaude huaishrle 'Play mahjiang copula the best bad thing —— Playing the mahjiang was probably the best among the existing bad habits.' The unit energy measurement (Figure 5.11.a.) showed that after two major rises halfway through the sentence, another rise of the amplitude occurred at 1526.5 msec into the sentence, and a fourth rise occurred shortly afterwards. That is, toward the end of the sentence, instead of a gradual fall of the amplitude, it rose and created two energy peaks. The intonation contour of the sentence (Figure 5.11) in turn could not be described by any of the proposed theories. Notice that the last word of the sentence huaishrle 'bad thing aspect–marker' consisted of two consecutive stressed falling tone followed by a neutral tone suffix. The word formation rule (3,1,1,3) would predict that the lexical tone of the first morpheme huài 'bad' would not fall completely before the second morpheme shr̀ 'thing' was produced. Both the declination theory and the breath–group theory would agree that at this position of the sentence, the fundamental frequency should be falling. The last toneless morpheme le 'aspect–marker' should be predicted as low–level too. On the contrary, we can see in Figure 5.11. that the fundamental frequency contour for the morpheme huài 'bad' falls from 164.42 Hz to a low 140.38 Hz, which is the lowest point of the entire sentence. Theoretically, the contour of this syllable should only fall halfway before the following falling–tone was produced. The following morpheme shr̀ 'thing', and le 'aspect marker' instead of picking up the end of the previous morpheme, started at 157 Hz, then moving upward to 275.96 Hz before falling to 250 Hz. That is, although the sentence ends in a terminal fall, it is marked by a high frequency fall (250 Hz) which cannot be explained by either of the previous two theories. However, if we analyze the syntactic structure of the sentence, i.e.,

```
[  [dǎge shiǎupáir]  [shr̀  [tzuèihǎude huàishr̀]  le]  ] ]
S  NP             NP VP    NP                   NP   VP
```

We find that the last NP of the sentence consists of an adjective modifying a noun. That is,

```
       [    [tzuèihǎude]   [huàishr̀]   ]
NP    ADJ            ADJ  NP       NP  NP
```

The noun huaishr 'bad thing' is the portion of the sentence where the predicted fundamental frequency did not fall. The only explanation for the fundamental frequency pattern of this example is probably the syntactic structure that also governs the stress above the word level. In order for the NP to stand out in the sentence, the lexical tones override the intonation, causing both the word formation rule and the intonation pattern to modify according to the speakers intention. If this is indeed the case, the phenomenon might be accounted for as a marked breath–group.

The second puzzling type is causes where a normal breath–group is produced, yet the terminal intonation with falling lexical tone somehow becomes higher in frequency. In other words, the last stressed syllable of a sentence appears to show a rise in fundamental frequency before falling. This can be called terminal shift of register. Figures 5.12. and 5.12.a. are an example of this type.

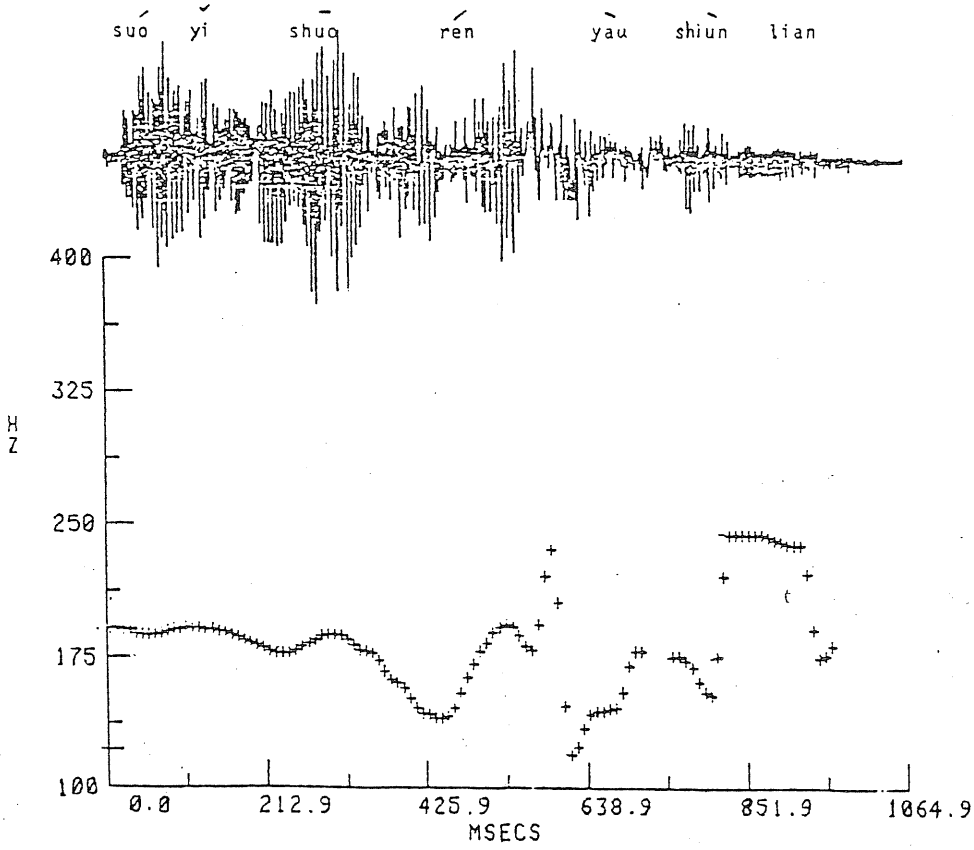


Figure 5.12. Intonation contour of a Mandarin sentence that fits neither the breath-group theory of intonation, not the declination theories. The sentence is Suóyǐ shuō rén yàushiùnlian 'So say people need training--That's why people need to trained.

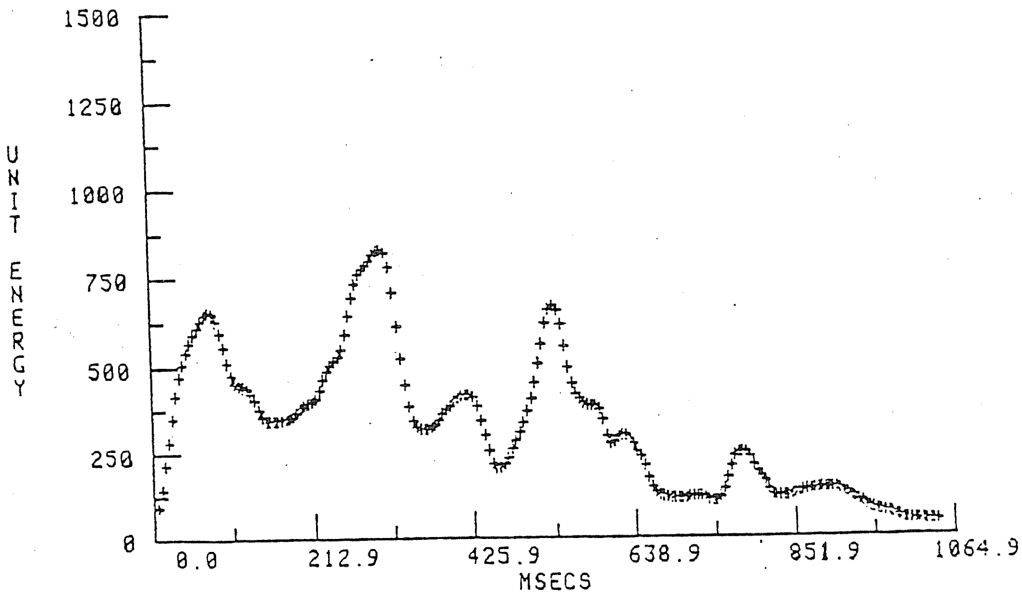


Figure 5.12.a.Unit energy measurment of Madnarin sentence shown in Figure 5.13. The amplitude here shows that the sentence was produced within a normal breath-group.

The sentence in Figure 5.12. is suóyǐ shuō rén yàu shiùnlian 'So say people need training.— That's why people need to be trained.' Figure 5.12.a., the unit energy measurement of the sentence, shows a normal breath–group. That is, the sentence is produced with some amplitude peaks, in this case three in number, followed by a gradual fall of amplitude. However, the intonation contour of the sentence (Figure 5.12.) is marked by a sharp rise of fundamental frequency at the terminal portion. For the last word shiùnlian 'training', it consists of a stressed falling tone morpheme shiùn followed by a neutral tone morpheme lian. Analyzing the syntactic structure may offer some clue for a possible explanation. The sentence can be analyzed as:

```
[ [suóyǐ shuō] [ [rén]  [yàu  [shiùnlian]  ] ] ] ]
S            S  NP    NP VP         NP    NP  VPS  S

      so say      people  need    training
```

notice that the last NP <u>shiùnlian</u> 'training' is an object of the verb <u>yàu</u> 'need', which is also in falling tone. However, the fact that the NP is the object of the verb probably explains the following two phenomena. First, there is a shift of fundamental frequency between the verb <u>yàu</u> 'need' and the object <u>shiùnlian</u> 'training'. The syntactic boundary prevented the two falling tones <u>yau</u> 'need' and <u>shiùn</u> in <u>shiùnlian</u> 'training' to interact. In other words, the rule for two consecutive falling tones (3.1.1.4) does not apply across major syntactic categories. Thus we do not find the first falling tone falling only halfway before the second falling tone is attached. Second, <u>maybe</u> the speaker intended to produce <u>shiùnlian</u> 'training' more saliently than the rest of the sentence. Nevertheless, it is important to notice that the fundamental frequency pattern for the word <u>shiùnlian</u> 'training' no longer possesses its lexical tones, and its corresponding amplitude is very low, but this causes no confusion whatsoever in perception.

The third type stems from causes where the sentence is rather short in duration, about 700–900 msec. The lexical tones are clearly produced, leaving no trace of intonation. In other words, the overall fundamental frequency pattern of the sentence is the lexical tones after modification triggered by word formation rules. That is, after the application of stress rules and if necessary sandhi rules. Nevertheless, the role of an intonation pattern appears to be rather minor in such cases. Figures 5.13. and 5.13.a. are one of the examples of this kind.

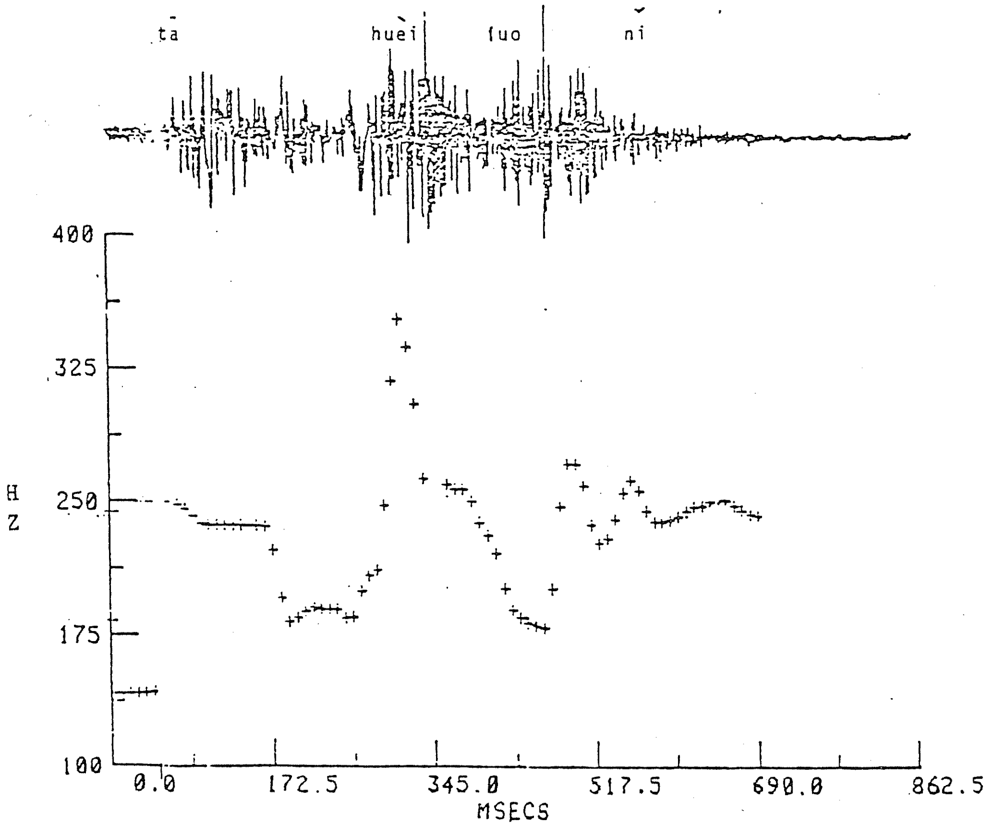


Figure 5.13. Intonation contour of Mandarin sentence <u>Tā huèiluo nǐ</u> 'He bribes you.' Note that the sentence consists of only 4 syllables and that the lexical tones override the intonation contour.

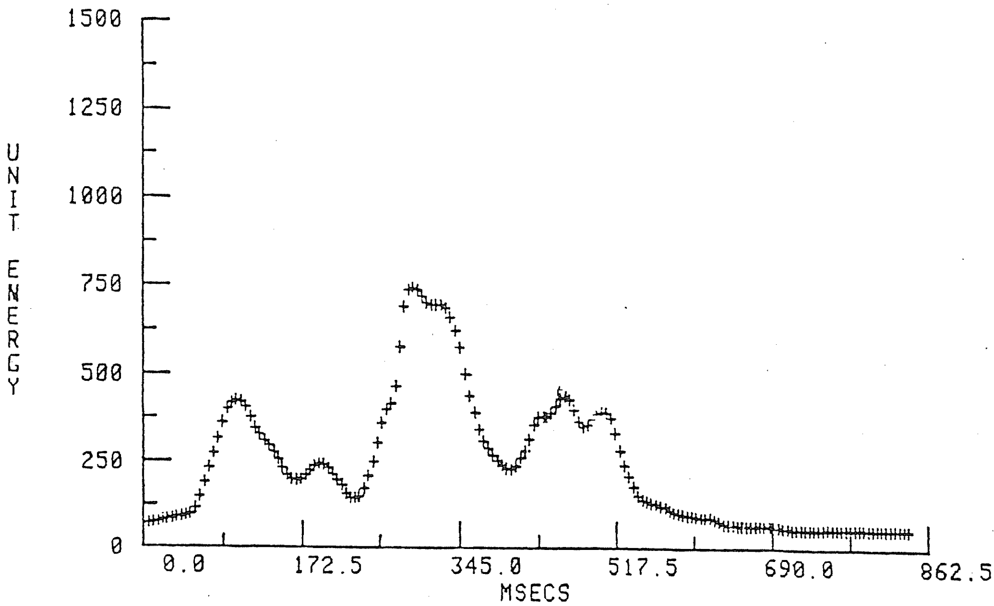


Figure 5.13.a. Energy measurement of Mandarin sentence Tā huèiluo nǐ 'He bribes you.' as shown in Figure 5. 14.

The sentence in Figure 5.13. tā huèiluo nǐ 'He bribes you' consists of four syllables only. The sentence was produced with a normal breath–group, as Figure 5.13. shows. In this case, each lexical tone is clearly identifiable, except the unstressed morpheme –luo in huèiluo 'bribe'. (For a detailed discussion of this sentence, see also 3.2.1.). No clear intonation exists except the lexical tones. Such cases had previously been observed by Chao (1968 : 40–44), and the present acoustic data provide instrumental evidence for this observation. Chao had also noted that such phenomena only occur with sentences not more than four syllables in structure. Whether the same situation occurs in other tone language involves further study, especially with respect to word structure.

Another point that is worth some attention is the fact that speakers of the same language may vary in their production of intonation. The results demonstrate that 9% of speaker COL's sentences has the declination effect. whereas 30% of speaker LH's sentence demonstrates this effect, 72% of speaker COL's production and 26% of speaker LH's fit neither theory. The results give evidence of individual variation in production, which also weakens any intonation theory's predictive power.

5.3. Comparison of spontaneous speech vs. read speech

In order to obtain the production data in different forms, the same subjects were later asked to read the same sentences they had previously produced in natural spontaneous speech in isolated forms using the same tape recording equipment. The analysis of both sets of sentences were facilitated by a Digital PDP–11/34 minicomputer using a pitch extraction software.

The result is summarized in Table 5.4.:

Table 5.4. Comparison of intonation patterns of Mandarin simple declarative sentences produced in natural spontaneous conversation vs. in read isolated form.

intonation type	spontaneous conversation	read form
declination	20%	50%
terminal fall only	45%	27.5%
other	35%	22.5%

In the form of spontaneous natural conversation, 20% of the data can be characterized by the declination theory, that is, having an overt visual declination in their F0 contour. 45% of the data can be characterized as having a terminal fall rather than an overall downdrift, thus giving support to the breath–group theory. However, 35% of the data cannot be characterized by either theory.

Table 5.4 also summarizes result of the same sentences read in unrelated isolated form by the same speakers: 50% of the sentences can be characterized by the declination theory; 27.5% of the data can be characterized by the terminal fall theory; whereas 22.5% of the data cannot be characterized by the either theory. Figure 5.14. shows a read sentence having a distinct downdrift pattern as its intonation.

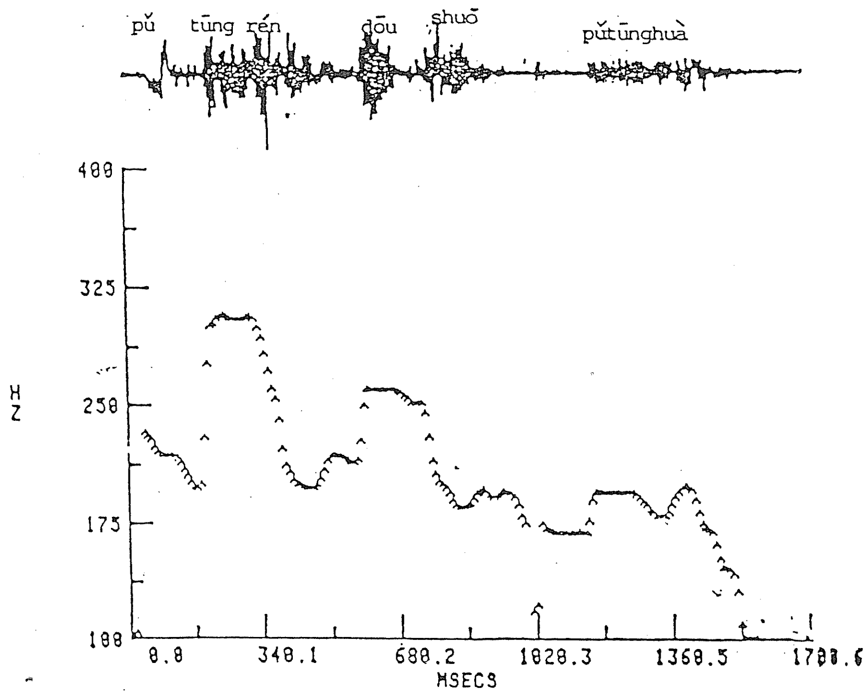


Figure 5.14. Intonation contour of Mandarin sentence <u>Pǔtūngrén dōu shuō pǔtūnghuà</u> 'Oridnary folks all is putunghua.' produced in read isolated form. The intonation contour can be characterized by the downdrift pattern of the declination theory.

Figure 5.15. is a read sentence whose intonation pattern shows a terminal fall only.

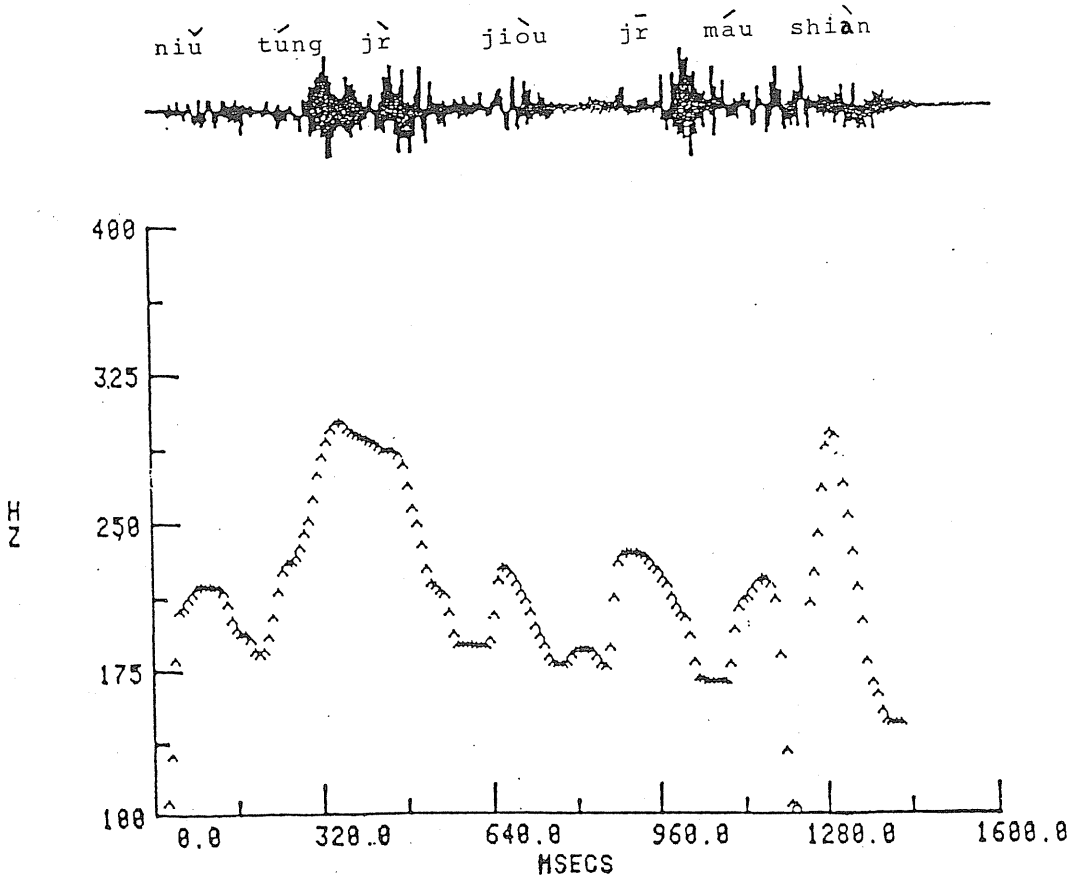


Figure 5.15. Intonation controu of Mandarin sentence Niu tung jr jiou jr maushian 'The female comrades did the knitting.' produced in read isolated form . the intonation contour can be characterized by the breath-group/ terminal-fall theory only.

Figure 5.16. is a sentence whose intonation does not possess either the declination or the terminal fall type of intonation.

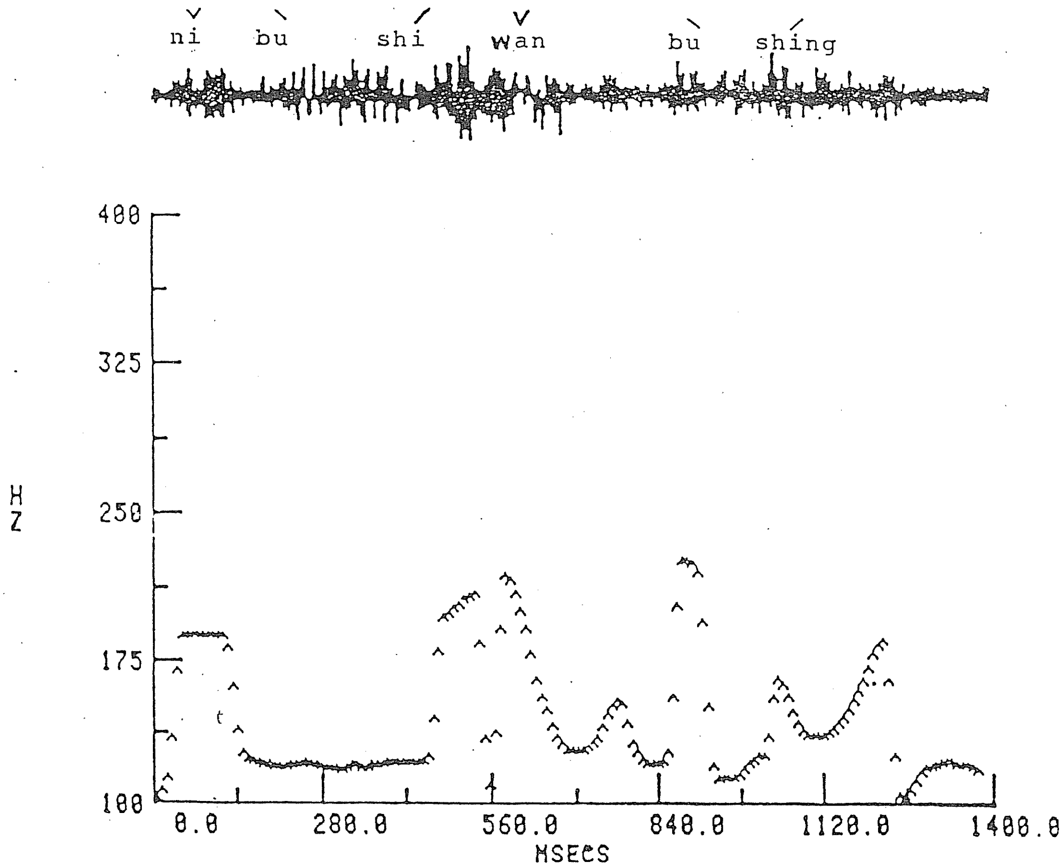


Figure 5.16. Intonation contour of Mandarin sentence <u>Ni bu shi wan bushing</u> 'If you don't do the dishes, it wouldn't be all right.' produced in read isolated form. The intonation contour can not be characterized either by the declination theory or by the breath-group/terminal-fall theory.

A logical question as a consequence of the above result is: Since the declination pattern is found more in read, isolated speech, would it lead to the conclusion that the underlying baseform of sentence intonation is then the downdrift baseline? A further analysis by the speakers showed a different result from this hypothesis. Table 5.5 summarizes the comparison of speakers.

Table 5.5. Comparison of intonation patterns of Mandarin simple declarative sentences by speakers: natural spontaneous conversation vs. isolated read form.

intonation type	Speaker LH		Speaker COL	
	spontaneous conversation	read speech	spontaneous conversation	read speech
declination	30%	75%	10%	25%
terminal fall only	65%	15%	25%	40%
other	5%	10%	65%	35%

The result shows that speaker LH and speaker COL behave rather differently. Speaker LH tended to use the declination pattern more than speaker COL in either conditions. That is, in spontaneous natural speech, 30% of speaker LH's sentences were produced with a declination pattern; 65% of the sentences with a terminal fall; and only 5% cannot be characterized as either. In read form, 75% of speaker LH's sentences were produced with a declination pattern; 15% with a terminal fall; and 10% cannot be characterized as either. Whereas with speaker COL in spontaneous natural fluent speech, only 10% can be characterized by the declination pattern; 25% with a terminal fall; and 65% cannot be characterized as either. In read form, 25% of speaker COL's sentence were produced with a declination pattern; 40% with a terminal fall, and 35% cannot be characterized as either. This analysis shows a distance difference of preferred intonation pattern by speakers. As it is shown here, speaker LH has a tendency to use the declination pattern more than speaker COL in both natural spontaneous fluent speech and read speech. Speaker LH also shows a higher tendency to produce simple declarative sentences with a terminal. It is mostly speaker COL who tends to produce intonation pattern that cannot be characterized by either theory, both in natural spontaneous fluent speech, and in read form. In other words, even in careful reading, speaker COL prefers terminal fall to the declination pattern.

A further look into the transcription of the conversation corpus shows the following: Many sentences were incomplete or interrupted by laugh, or by the other speaker using vocatives such "oh, eh, yah". Both speakers often talk simultaneously. Very long sentences were rarely produced. The two speakers often competed to gain attention, but using different strategies. Speaker LH tended to start a sentence with rather high F0 when she tried to gain attention, which often resulted in a distinctive downdrift intonation pattern. The same strategy was also used by speaker LH to start a now discourse. The remaining sentences within a discourse tended to have flatter contours. In other words, most of her sentences with distinct declination pattern were either discourse initial, or attention attracting type of sentences. On ther other hand, speaker COL's strategy to gain attention was by increasing the volume of her voice and to enunciate the lexical tones, thus resulting in a slower rate of speech; but not necessarily starting a sentence with a higher F0 on the first content words. Thus one gets the impression that speaker COL is louder and clearer. However, there is one strategy both speakers used to gain attention, namely, they tended to repeat the last portion of the other talker's sentences to signal the intention of wanting to be the one to take over the talking.

The repetition was usually very fast and incomplete sentences and in rather low voice.

In read form, the two speakers also used different strategies. Speaker LH is a very efficient reader and tended to use the same intonation pattern across the sentences, since they were all uniform in syntactic structure. Speaker COL, on the other hand, tended to hesitate and halt, even at the second and third reading. she also left long pauses between syntactic boundaries. In so doing, she tended to produce the sentences with emphasis on the tones of the stressed syllables rather than using an overall intonation pattern. The terminal fall intonation pattern is the second preferred intonation by speaker COL. However, the read forms resulted from both speakers sounded perfectly clear and fluent, thus does not give support to the hypothesis that declination pattern is the baseform.

From this study of comparison, a tentative speculation can be drawn from the observations, that is, the complexity of contextual as well as discourse information governs the choice of intonation pattern in natural spontaneous conversation rather than syntactic structures alone. The baseline is not clear for one speaker in most of the cases. Both speakers expressed the degree of importance of each word through F0, resulting the highest peak of an utterance not necessarily falling on the first content words.

Furthermore, it is also observed that in casual speech, speakers prefer a series of short, simple sentences rather than long and syntactically complex ones. As a result, the intonation pattern of a series of connected similar sentences would have to vary to avoid sounding monotonous or boring. Thus the governing principle of selecting sentence intonation goes beyond the sentence itself, giving support to further research in the direction a context/discouse oriented theory of sentence intonation.

5.4. Summary and conclusion

In this chapter, production data of spontaneous Mandarin Chinese sentences as well as the same sentences in read forms were studied for their intonation and the interaction between tones and intonation and some possible governing factors in producing the actual output were also touched upon. Each sample sentence in the data was measured and examined to test two intonation theories, namely, the breath–group theory versus the declination theory.

The data presented suggest the following points: (1) The declination effect does exist in Mandarin Chinese, but only for a very small portion of the sentences studied. (2) The declination theory is not only weak in its predictive power, but also proves to be too strong a claim in terms of its universality. (3) The breath–group theory, offers a more plausible alternative for characterizing intonation. (4) The interaction between syntax and intonation does occur, thus further supporting previous research on intonation. (5) The existence of intonation patterns which can not be explained by either theory should not be ignored. (6) Individual variation in producing spontaneous connected speech as well as in well–formed read sentences poses more difficulty for any proposed intonation theory. (7) The declination theory may prove to be more feasible in characterizing the discourse effect rather than a sentential phenomenon. We are aware of the fact that there might exist a considerable discrepancy between spontaneous fluent speech and read speech, the former being more difficult to characterize not only in intonation alone, but also in segmental information as well. The present attempt to include both forms should thus do more justice to any intonation theory under investigation.

Appendix 5.1.: Production Data of Spontaneous Speech

1. 老魯知道他是一個單身漢.
2. 家裡小孩很多.
3. 你不洗碗不行.
4. 說他是一個年輕的人.
5. 他們什麼都叫服務員.
6. 耿爾點了幾盤牛羊肉.
7. 那種舖子雞鴨可以豬肉不行.
8. 這個麵叫做粉絲.
9. 不能夠隨便吃多少.
10. 那個時候沒有什麼知識.
11. 他們那年頭兒上大學跟我們不一樣.
12. 我們那陣子又跟他們不一樣.
13. 打個小牌兒是最好的壞事兒了.
14. 現在的記憶很痛苦.
15. 所以說人要訓練.
16. 你媽媽不必告訴你.
17. 在第一百三十六頁倒數第三行.

18. 大 陸 上 的 朋 友 都 好 會 講 .

19. 台 灣 不 注 意 這 些 .

20. 後 來 我 才 發 現 這 玩 意 兒 很 重 要 .

21. 那 個 可 是 好 的 .

22. 而 且 過 年 政 府 都 多 配 給 東 西 .

23. 其 實 絲 棉 襖 也 不 怎 麼 頂 事 兒 .

24. 女 學 生 都 有 軍 訓 .

25. 身 上 錢 不 夠 .

26. 普 通 人 都 講 普 通 話 .

27. 家 裡 的 人 也 都 說 普 通 話 .

28. 當 時 我 的 一 個 親 戚 在 那 裡 .

29. 但 是 我 的 祖 先 不 是 北 京 .

30. 祖 先 是 山 東 蓬 萊 .

31. 然 後 我 就 在 北 京 了 .

32. 它 的 對 象 是 農 村 .

33. 當 時 我 們 是 到 吉 林 .

34. 當 時 是 因 爲 錢 不 夠 .

35. 到 處 都 一 樣 .

36. 他 賄 賂 你 .

37. 我 也 是 覺 得 普 通 話 好 聽 .

38. 那 時 候 起 我 就 沒 怎 麼 樣 離 開 過 北 京 .

39. 這 就 是 我 們 可 以 幫 他 們 做 的 事 情 .

40. 那 個 好 吃 極 了 .

41. 我 織 毛 線 的 歷 史 比 較 短 .

42. 就 研 究 工 作 就 基 本 上 停 下 來 了 .

43. 男 同 志 就 抽 煙 啊 .

44. 女 同 志 就 織 毛 線 .

45. 重 新 就 恢 復 科 室 .

46. 就 重 新 恢 復 原 來 的 研 究 工 作 .

47. 我 是 搞 偏 微 分 方 程 的 .

48. 那 我 丈 夫 跟 我 的 工 資 完 全 一 樣 .

49. 這 個 問 題 老 早 就 應 該 採 取 嚴 厲 的 措 施 .

50. 我 再 給 你 說 一 個 故 事 .

CHAPTER 6 SUMMARY AND CONCLUSION

The purpose of this study was to investigate some of the acoustic properties of Mandarin tones in citation form and in spontaneous running speech. The experiments described in Chapters 2–5 are designed to investigate the primary acoustic correlates for tones in Mandarin, as well as the interaction between tones and intonation.

The production data can be summarized as follows. Mandarin syllables produced in citation forms have best demonstrated the fundamental frequency patterns of tones hypothesized in phonological studies. These distinct tonal patterns, however, are not always present in spontaneous running speech. When produced in citation forms, the intrinsic duration of the vowels varies according to different tonal patterns, yet such distinctions in duration were not maintained in spontaneous running speech. This result suggested that duration may be a concomitant acoustic correlate for tones. Our study of the production of spontaneous running speech demonstrated a low correlation between the theoretical prediction of lexical tones and the actual acoustic information. This result suggested that in the production of spontaneous running speech in verbal communication, the major acoustic correlate needed for producing tones, i.e., the fundamental frequency patterns, was only partially present. Speakers' integrated knowledge of the linguistic system enables them to produce only partial acoustic information of lexical tones in a sentence in spontaneous running speech. The perceived phonetic output of sentences is thus the ultimate consequence derived from a complicated interaction among the different levels of linguistic/extra–linguistic knowledge involved. Phonology and phonetics alone are not sufficient in predicting the actual output of a sentence. In establishing the context, there appears to exist some trade–off effect among all the interacting linguistic/extra–linguistic levels of information. Nevertheless, well–formed, read sentences would seem much less complicated since a good deal of the possible interacting factors could be eliminated. In other words, the more the constraints are, the less likely the output appears to vary. Production data were also utilized in testing some different theories of intonation, i.e., the breath–group theory (Lieberman, 1967) and the declination theories (Maeda, 1967; Breckenridge and Liberman, 1977, 1979; Sorensen and Cooper, 1980). The study has shown that the declination theories failed to capture the overall intonation patterns of simple declarative sentences in Mandarin Chinese. The claim that the declination effect is a near–universal property is thus seriously questioned. Declination in intonation might be more of a discourse effect rather than the baseform for declarative sentences (Umeda and Coker, 1980).

The perception data can be summarized as follows. In the perception of Mandarin tones produced in citation forms, native speakers were able to identify these tones without the support of context, due to the fact that the fundamental frequency patterns of these tones were sufficient acoustically. That is, in order for the subjects to perceive such tones, the fundamental frequency contours of the tokens possess the specified tonal patterns. In the perception of syllable tones edited from spontaneous running speech, which often carried insufficient acoustic information specified for tones, subjects' behaved rather poorly. That is, without the support of contextual information, native speakers were unable to identify syllable tones on the basis of insufficient acoustic information alone. This result indicates that in the perception of tones in spontaneous speech, native speakers may very well "reconstruct" or "recover" the lexical tones that are not physically present from the context. Speech perception is thus a complicated decoding process based not on the acoustic information only.

BIBLIOGRAPHY

Abramson, A. S. (1962). The vowels and tones of Standard Thai: Acoustical measurements and experiments. International Journal of American Linguistics, 28 (2), (Part II).

———(1975). The tones of Central Thai: Some perceptual experiments. In Studies in Thai Linguistics, J. G. Harris and J. Chamberlain (eds.)., Central Intitude of English Language, Eangkok.

Anderson, M. D., Pierrehumbert, J. B. and Liberman, M. Y. (1984). "Synthesis by rule of English intonation patterns." Proceedings of the 184 International Conference on Acoustics, Speech and Signal processing. San Diego, CA, March 19–21, 1984.1. 2, 811–44.

Armostrong, L. E. and Ward, I. C. (1926). Handbook of English Intonation. E. G. Teubner, Leipzig and Berlin.

Atkinson, J. R. (1973). Aspects of Intonation in Speech: Implications from an Experimental Study of Fundamental Frequency. PhD dissertation. University of Connecticut, Storrs.

Bolinger, D. L. (1972). Accent is predictable (if you're a mind–reader). Language, 48, 633–644

Blot, R. H., Cooper, F. S., David, E. E. Jr., Denes, P. B., Picket, J. M. and

Stevens, K. N. 91973). Speaker identification by speech spectrograms: some further observations. Journal of Acoustical Society of America, 54, 531–534.

Breckenridge, J. and Liberman, M. Y. (1977). The declination effect in perception. Bell Lab, New Jersey.

Bresnan, J. R. (1971). Sentence stress and syntactic transformations. Language, 47 (2)., 257–281.

Chao, Yuan–Ren. (1930). Tone and intonation in Chinese. Academic Sinica, Bulletin of the Institute of History and Philology, IV. 2, 121–134.

———. (1968). A Grammar of Spoken Chinese. University of California Press, Berkeley, California.

Cheng, Chin–Chiuan. (1973). A Synchronic Phonology of Mandarin Chinese. The Hague, Mouton.

Cheng, B. L. (1966). Mandarin phological structure. Journal of Linguistics, II, 135–158.

Chiba, T. (1935). A Study of Accent, Research into Its Nature and Scope in the Light of Experimental Phonetics. Phonetic Society of Japan, Tokyo.

Cook, Enug–Do. (1971). Vowels and Tones in Sarcee. Language, 47, 164–179.

Cole, R. A. and Jakimik, J. (1980). A model of speech perception. In Production and Perception of Fluent Speech, R. A. Cole, ed., Lawrence Erbaum Assoc., Hillsdale, N. J.

Collier, R. (1975). Physiological correlates of intonation patterns. JASA, 58, 249–255.

Denes, P. and Milton–Williams, J. (1962). Further studies in intonation. Language and Speech, 5, 1–4.

———. (1959). A preliminary investigation of certain aspects of intonation. Language and Speech, 2, 106–122.

Erickson, D. and Abramson, A. S. (1973). Electromyographic study of the tones of Thai, JASA, 53, 346.

———. (1975). A Laryngeal Description of Thai Tone. L. S. A. Handbook of 1975 Meeting, San Francisco.

Fant, G. (1960). Acoustic Theory of Speech Production. The Hague, Mouton.

Formkin, V. A. (1972), Tone features and tone rules. Studies in African Linguistics, 3, 47–76.

———, ed. (1978). Tone: A Linguistic Survey. Academic Press, New York.

Fry, D. B. (1955). Duration and Intensity as Physical Correlates of Linguistic stress. JASA, 35, 765–769.

Goldsmith, J. (1975). Tone melodies and the autosegment. Conference on African Languages, 6, 135–147.

———. (1976a). "Autosegmental Phonology." PhD dissertation. MIT, Cambridge, Massachussetts.

———. (1976b). An overview of autosegmental Phonology. LA, 2, 23–68.

Hollien, H. (1962). Vocal fold thickness and fundamental frequency of phonation. Journal of Speech and Hearing Research, 5, 237–243.

———. and Michel, J. F. (1968). Vocal firing as a phonational register. Journal of Speech and Hearing Research, 11, 600–604.

Hambert, J. M. (1978). Consonant Types, Vowel Quality and Tone. In Tone: A Linguistic Survey, V. Fromkin, ed., Academic Press, New York.

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

Hsieh, Hsin–I. (1976). On the unreality of some phonological rules. Lingua, 38, 1–19.

Hudson, R. A. (1972). Syllables, Moras, and Accent in Beja. Journal of Linguistics, 9, 53–63.

Jacobson, R. C., Fant, G. M., and Halle, M. (1952). Preliminaries to Speech

Analysis MIT Press, Cambridge, Massachusetts.

Jiang, Wei–Sung and Yin, Huan–Xien. (1957). Zidiau han Yudiao. Xinzhishi Chubanshe, Shanghai.

Jones, D. (1932). An Outline of English Phonetics. Dutton, New York.

Kratochvil, P. (1968). The Chinese Language Today. Hutchinson and Company, Ltd., London.

Ladefaged. P. (1975). A Course in Phonetics. Harcourt Race Jovanovich, New York.

Larson, J. (1975). Downstep, downdrift, and diacritics. Conference on African Linguistics, 2, 171–181.

Leben, W. R. (1971). Suprasegmental and segmental representation of tone. Conference on African Languages, 2, 183–200.

———. (1973). The Role of Segmental Phonology. In Consonant Types and Tone, L. M. Hyman, ed., (Southern California occasional Papers in Linguistics, 1). Los Angeles: Linguistics Program, University of Southern California, 115–141.

Lehiste, I. (1961). Some acoustic correlates in accent of Serbo–Croatian. Phonetica, 7, 114–147.

———. (1970). Suprasegmentals. MIT Press, Cambridge, Massachusetts.

Lenneberg, E. H. (1967). Biological Foundations of Language. Wiley, New York.

Li, C. N. (1975). Subject and Topic. Academic Press, N. Y.

Liberman, M. Y. and Pierrehumbert, J. (1984). "Intonational invariance under changes in pitch range and length." In Languagge and Sound Structures., edited by M. Arnoff and R. Oehrle, MIT Press, Cambridge, MA, 157–233.

Lieberman, P. (1960). Some acoustic correlates of word stress in American–English. JASA, 33, 451–454.

———. (1965). On the acoustic basis of the perception of intonation by linguists. Word, 21, 40–54.

———. (1967). Intonation, Perception, and Language. MIT Press, Cambridge, Massachusetts.

———. (1968). Direction comparison of subglottal and esophageal pressure during speech. JASA, 43, 1157–1164.

Lieberman, P., Knudsen, R. and Mead, J. (1969). Determination of the rate of change of fundamental frequency with respect to subglottal s'. pressure during sustained phonation. JASA, 45, 1537–1543.

———. (1970). The articulatory implementation of breath–group and prominence:

Crico–throid muscular activity in intonation. Language, 46, 312–327.

———. (1970). Towards a unified phonetic theory. Linguistic Inquiry, 1, 307–332.

———. (1974). A study of prosodic features. In Current Trends in Linguistics, Vol. 12, T. Sebeok, ed., The Hague, Mouton.

———. (1976). Phonetic features and physiology: a reappraisal. Journal of Phonetics, 4, 91–112.

———. (1977). Speech Physiology and Acoustic Phonetics. Macmillan, N. Y.

———. (1979). Phonetics and physiology: Some Current Issues. In Current Trends in Linguistic Theory, Perspectives in Experimental Linguistics, G.D. Prideaux., ed., John Benjamins, Amsterdam.

Massaro, D.W., C. Tseng and M. M. Cohen. (1983). "Vowel and lexical tone perception in Mandarin Chinese: psycholinguistic and psychoacoustic contributions." In Quantitative Linguistics, 19, 76–102.

Massaro, D.W., M. M. Cohen and C. Tseng. (1985). "The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese." In Journal of Chinese Linguistics, 13.2, 267–290.

McCawley, J.D. (1973). Global rules and Bangubangu tone. In Issues in Phonological Theory. M. J. Kenstowicz and C. W. Kisseberth, eds. The Hangue, Mouton.

Maeda, S. (1976). "A Characterization of American English Intonation," unpublished PhD dissertation. MIT, Cambridge, Massachusetts.

Mertus, J. (1980). WAVE: Waveform File Editor Version LA–10. 16, Brown University, Providence, R. I.

Miller, J. (1961). Word tone recognition in Vietnamese whispered speech. Word, 17, 11–15.

Morton, J. and Jassem, W. (1965). Acoustic correlates of stress. Language and Speech, 8, 159–181.

Peterson, G. E., Wang, W. S–Y. and Sieverson, E. (1958). Segmentation techniques in Speech synthesis. JASA, 30, 739–742.

Peterson, G. E. and Lehiste, I. (1960). Duration of syllable nuclei in English. JASA, 32, 693–703.

Pierrehumbert (nee Breckenridge)., J. (1979). The perception of fundamental frequency declination. JASA, 66, 363–369.

———. (1980). The phonlogy and phonetics of English intonation. Ph.D. dissertation, MIT.

Pike, K. L. (1945). The Intonation of American English. University of Michigan, Ann Arbor.

——. (1948). Tone Languages. The University of Michigan Press, Ann Arbor.

Schachter, P. (1969). Natural assdmilation rules in Akan. IJAL, 35, 342–355.

Sorensen, J. M. and Cooper, W. E. (1980). Synthetic Coding of Fundamental Frequency in Speech Production. In Perception and Production of Fluent Speech, R. A. Cole, ed., Lawrence Erlbaum Assoc., Hillsdale, N. J.

Stockwell, R. P. (1972). The Role of Intonation. In Intonation, Bolinger, ed., Penguin, New York.

Studdert–Kennedy, M. (1974). The Perception of Speech. In Current Trends in Linguistics, Volume 12, T. Sebeak, ed., The Hague, Mouton.

Thorsen, N. (1978). "An acoustic analysis of Danish intonation." in Journal of Phonetic, 6, 151–175.

——. (1984). "Fo timing in Danish word perception." In Phonetica, 41(1), 17–30.

——. (1984). "Variability and invariance in Danish stress group." In Phonetica, 41(2), 88–102.

Tranger, G. L. and Smith, H. I. (1951). An Outline of English Structure. Washington American Council of Learned Societies, Seventh Printing.

Tseng, Chiu–yu. (1979). "The production and perception of Mandarin vowel i: In all four tones regarding duration." In Speech Communication Papers, J. J. Wolf and D. H. Klatt, eds., Acoustical Society of America, N. Y.

——, D.W. Massaro and M. M. Cohen. (1986). "Lexical tone perception in Mandarin Chinese: Evaluation and integration of acoustic features." In Linguistics, Psychology, and the Chinese Language, Kao, H.S.R. and Hoosain, R. Ed.s. University of Hong Kong Press, 91–104.

Umeda, N. (1981b). F0 rule for discourse, JASA, 69, S82.

——. 1982). "F0 declination" is situation dependent. Journal of Phonetics, 10, 279–290.

Wang, Li. (1936). Zhongguo Yinyunxue. Shanguyinshuguan, Shanghai.

——. (1959). Chungkuo Hsien Tai Yu Fa (Zhongguo Xiendai Yufa). Zhonghuashuju, Beijing.

——. (1963). Han Yu Yin Yun. Zhonghuashuju, Beijing.

Wang, W. S–Y. (1967). Phonological features of tone. IJAL, 33, 93–105.

Hsie, Yun–Fei. (1971). Yin Xue Shi Lun. Wufengchubanshe, Wufeng.

Yip, M. (1980). "The Tonal Phonology of Chinese." Unpublished PhD dissertation. MIT, Cambridge, Massachusetts.

Chou, Fa–Gao. (1975). Junggguo Yuyenshiue Lunwenji. Lienjing Chuban Shiye Gungsz, Taipei.

Zhou, You–Guang. (1955). Pinyin wenzi zhongde shengdiau wenti. Zhongguoyuwen, 37, 4–6. Beijing.

中央研究院歷史語言研究所專刊之九十四

國語字調的聲學語音研究

定價　平裝　新臺幣250元　美　金　10元　精裝　新臺幣300元　美　金　12元

（如匯率大幅變動，本所得以新匯率換算售價）

著　者　鄭　秋　豫

發行者　中央研究院歷史語言研究所
臺北市南港區

印刷者　文匯印刷資訊處理有限公司
臺北市環河南路二段211號

代售處　臺灣商務印書館
臺北市重慶南路一段37號
臺灣學生書局
臺北市和平東路一段198號
三民書局
臺北市重慶南路一段61號

中華民國七十九年二月出版