PHONETIC REALIZATION AND PERCEPTION OF PROMINENCE AMONG LEXICAL TONES IN MANDARIN CHINESE

By

MINGZHEN BAO

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

UNIVERSITY OF FLORIDA

2008

To my parents and my husband, for their unconditional love

ACKNOWLEDGMENTS

This has been a long journey. There are many individuals to whom thanks are owed. First and foremost, I would like to express my heartfelt thanks to my wonderful mentor and chair, Dr. Ratree Wayland, for all of her wisdom and guidance throughout my studies at the University of Florida. Her dedication to her students is beyond the highest expectations. This work would not have been possible without her extensive knowledge, constant encouragement and support. I hope I will learn from her persistence in academic pursuit in my career.

I would like to acknowledge the rest of my committee members (Professors Caroline Wiltshire, Masangu Matondo, and Jimmy Harnsberger) for their helpful suggestions. I owe Dr. Wiltshire much gratitude, not only for her insights into this study but ever since I applied for the Linguistics Program. My wholehearted thanks also go to Dr. Matondo and Dr. Harnsberger for sharing their valuable reading materials and discussing the design of experiments with me from the beginning to the final stage.

I thank Professors Edith Kaan and Takako Egi for the opportunities to assist them in research projects of language processing and second language acquisition. Professors Andrea Pham and Elinore Fresh also deserve a word of thanks for their supervision in teaching linguistics and language courses.

I also thank Professors Diana Boxer, Eric Potsdam, Fiona McLaughlin, Gary Miller, Roger Thompson, Virginia LoCastro, and Wind Cowles for introducing a great variety of linguistic branches to me. The knowledge gained through their courses is truly beneficial to widen my scope and deepen my understanding of linguistics.

In addition, I am obliged to fellow linguistic students Priyankoo Sarmah, Andrea Dallas, Bin Li, Lili Gai, Rania Habib, Rui Cao, Ye Han, Yunjuan He for their help, friendship and moral support. The administrative staffs of the Linguistics Program also deserve a word of thanks for their dedicated assistance.

Special thanks also go to the University of Florida for the Alumni Fellowship, the College of Liberal Arts and Sciences, and the Graduate Student Council for research-related travel grants which funded my study and completion of this work.

Lastly, I express my gratitude to my family for their love and concern. My parents Shisun and Qiyi, and my husband Tao deserve a special thanks for their caring, patience, understanding and encouragement. I dedicated this dissertation to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

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

PHONETIC REALIZATION AND PERCEPTION OF PROMINENCE
AMONG LEXICAL TONES IN MANDARIN CHINESE

By

Mingzhen Bao

August 2008

Chair: Ratree Wayland
Major: Linguistics

Linguistic prominence is defined as words or syllables perceived auditorily as standing out from their environment. It is explored through changes in pitch, duration and loudness. In this study, phonetic realization and perception of prominence among lexical tones in Mandarin Chinese was investigated in two experiments. Experiment 1 explored phonetic realization of prominence. The primary aim of this experiment was to compare and contrast acoustic characteristics of a target word produced under four conditions: (a) unaccented and unfocused; (b) accented but unfocused; (c) unaccented but focused; (d) accented and focused, among four tones. Ten native speakers of Chinese were recorded reading materials in a natural fashion with the target word appeared in the above four positions. The recorded data were segmented and acoustically measured for acoustic parameters: vowel duration; mean and maximum of intensity; mean, maximum, minimum and slope of F_0. The results showed that vowel duration lengthening was the main acoustic parameter associated with accent while an increase in vowel duration, mean and maximum of intensity and F_0, and slope of F_0 was associated with focus realization. It was also found that acoustic parameters used to realize focus were varied from tone to tone: an increase in duration, F_0, and intensity was presented in focus realization for Tone 1(high level tone) and Tone 4 (high falling tone); duration and F_0 were used to implement focus for Tone 2

(mid-high rising tone); while duration and intensity were used in Tone 3 (low falling-rising tone). Acoustic cues used to perceive prominence were investigated in Experiment 2. In this experiment, acoustic parameters found to have been used to realize focus in Experiment 1 were compared in pairs to test native speakers' preference in focus perception. Twenty native speakers of Chinese participated in the 'preference' judgment. The results showed that duration, mean and maximum of intensity cues were selected more often than pitch cues in focus perception. These results suggested that phonetic realization of prominence in Mandarin Chinese was affected by category of prominence (i.e., focus or accent) and tonal contexts. Moreover, acoustic parameters used by native Mandarin Chinese to produce focus were different from those used in their perception of focus.

CHAPTER 1
INTRODUCTION

All languages use vowels and consonants to distinguish meaning of one word from the other, so ‘pick’ is different from ‘sick’ or ‘pick’ is different from ‘pack’ because their first consonants, [p] versus [s]; or their vowels, [i] versus [æ], are different respectively. Such minimal pairs of words can be found in all of the world’s languages. However, the number of vowels and consonants used to contrast lexical meaning varies from language to language. Besides vowels and consonants, a difference in voice pitch is also employed to change word meaning in the so called ‘tone’ languages such as Mandarin Chinese, Vietnamese and Thai. In these languages, words change their meanings depending on the voice pitch or ‘lexical tones’ in which they are pronounced. These ‘tones’ are defined both by their pitch height or ‘registers’ (e.g., high, mid, and low) as well as their pitch contours (e.g., level, falling or rising) (Wang, 1967; Woo, 1969; Bao, 1990; Hyman, 1993; Odden, 1995; Snider, 1999; Yip, 2002). Mandarin Chinese, for example, include four lexical tones in its phonological system: Tone 1 (high level), Tone 2 (mid-high rising), Tone 3 (low-falling-rising) and Tone 4 (high falling). In Mandarin, the word ‘ma’ spoken with the first tone means ‘mother,’ with the second tone means ‘hemp,’ with the third tone means ‘horse,’ and with the fourth tone means ‘a scold or a reproach.’.

This is in contrast to stress languages such as English in which pitch is used to convey emphasis, contrast, emotion and other paralinguistic information at a larger linguistic unit of phrases and sentences. For example, falling and rising intonation contours over an utterance in English are used to distinguish a statement from a question, as well as displaying doubt, anger, fear and other emotions. Besides, pitch is also used to indicate relative degrees of prominence among syllables in multi syllabic words of English. For example, the first syllable in ‘national’ is perceptually more salient or more prominence than the last two. The relatively higher degree of perceptual salience of this ‘stressed’ syllable is due to its longer in duration, louder in volume or intensity and higher in pitch than its neighboring ‘unstressed’ syllables. A difference in ‘stressed’ location can be used to contrast meanings of such noun and verb pairs as in ‘an export’ and ‘to export’, or ‘an address’ and ‘to address’. Stress patterns in English can also be used to differentiate a compounded word, ‘a blackboard’ from an adjective-noun phrase, ‘a black board’. At the sentence level, timing and intervals between stressed and unstressed syllables affects the rhythm with which the utterance is spoken.

Similar to stress language like English, different intonation contours or pitch movements over an utterance (a phrase or a sentence) is also used in lexical ‘tone’ languages to convey emphasis, contrast and prosodic boundaries. When tone and intonation are concurrently realized in an utterance, voice pitch serves more functions than contrasting lexical meanings. It may signal an intonation pattern as statements or questions; convey doubt, anger and many other emotions. In other words, pitch heights and/or pitch contours of each lexical tone will be modified to additionally represent intonational expressions. Modifications may also be observed in other acoustic dimensions such as duration and intensity when intonation is superimposed on tones (Leben, Inkelas, & Cobler, 1989; Luksaneeyanawin, 1993; Ladd, 1996; Gussenhoven, 2004; Beckman, 2006). As discussed above, these three acoustic parameters: pitch, duration and intensity are most used to give some syllables prominence when compared with other syllables (as in English). Such linguistic prominence is important in informing a rhythmical framework of speech by connecting sequences of prominent and non prominent syllables; they may also convey new or contrastive information at the pragmatic level. In other words, these phonetic features are used to convey sentence-level information, encompassing syntactic and semantic information as well as pragmatic information. In a tonal language, such as Mandarin Chinese, acoustic parameters such as pitch, duration and intensity are expected to be modified to implement prominence while retaining tonal features.

As already mentioned, Mandarin Chinese is a tonal language. The intonational prominence shown on the sentence level can be identified in terms of its source: default sentence accent in a sentence final position marks a rhythmical prominence, and contrastive focus placed in any part of a sentence signals an informative prominence. In a sentence ‘John jiao le xuefei’ (John paid the tuition fee), the last word ‘xuefei’ is prominent as it receives the ‘default’ or ‘grammatical’ accent and marks the prosodic boundary of the sentence. The sentence final position for accent can be justified from several perspectives: syntactically, a non-head component (such as the object in a verb phrase) is more accented (Duanmu, 2000); semantically, rhyme is more prominent than theme in a sentence (McKie, 1996), and direct arguments (such as agent and patient) are more accented than the predicate (Gussenhoven, 1983); phonetically, the word in the sentence final position are accented (Chao, 1968; Yip, 1980). When the sentence is extended to ‘John jiao le xuefei, danshi Mary meiyou jiao’ (John paid the tuition fee, but Mary didn’t.), the sentence-middle word ‘Mary’ receives intonational prominence (or focus), because the utterance contrasts ‘Mary’ with ‘John’ and focuses on the contrast regarding the information delivered.

Many studies have been conducted to investigate prominence in Mandarin Chinese (Yip, 1982; Shen, 1985; Shih, 1988; Tseng, 1981; Liao, 1994; Jin, 1996; Xu, 1999, 2004; Chen, 2004; Liu & Xu, 2005). However, after years of research, some questions regarding the production and perception of prominence remain unsolved. For instances, is focus and accent phonetically realized in a same fashion? Are different tones modified differently to implement prominence? What cues are used in prominence perception? Thus, in this study, we explored the interaction among tone, accent and focus to look for answers to these questions.

Purpose and Significance of the Study

The overall purpose of this study was to investigate the phonetic realization and the perception of prominence caused by accent and focus in the environment of longer utterances to allow for an examination of the interactions among tone, accent and focus in Mandarin Chinese.

The study filled the gaps of previous studies on prominence in Mandarin Chinese in the many important respects. First, unlike previous studies, in this study the sources of prominence were separated to sentence accent and contrastive focus. Second, the study domain was expanded to longer utterances, which provide a more natural context for accent and focus realization. Third, the phonetic realization of prominence among tones were compared and contrasted. Fourth, perception and production experiments were conducted and results were compared with the same set of data. Finally, quantitative analyses were applied to the study of prominence (shown in Figure 1-1).

Previous studies	*This study*
• Examine prominence in general	o Separate prominence categories (i.e., accent and focus)
• Study domain limited to short utterances (e.g., words, phrases, simple sentences)	o Study domain extended to longer utterances (e.g., sentence groups)
• Investigate tone in general	o Exploit tonal differences in the realization and the perception of prominence
• Address either realization or perception of prominence	o Include both realization and perception of prominence, and compare acoustic parameters used for realization with those in perception
• Analyze in a descriptive way	o Analyze in a quantitative way (e.g., repeated-measure ANOVA and follow-up pair-wise comparison)

Figure 1-1. Improvement made in this study

Research Questions

This study was guided by three research questions:

- Research question 1: What are the acoustic parameters used to realize focus and accent among lexical tones of Mandarin Chinese?
- Research question 2: What are the interactions among tone, accent and focus in the realization of focus and accent?
- Research question 3: Among acoustic parameters used to produce focus and accent, which ones are used in the perception of prominence?

Research Design

To answer the three research questions, two experiments were designed: a production experiment aimed at exploring phonetic realizations of prominence and a perception experiment devised to investigate perceptual cues used in prominence perception.

In the production experiment, native speakers of Mandarin Chinese (N=10) were recorded producing utterances where the bi-syllabic target words produced with all possible combination of the four tones were set in prominent and non-prominent conditions. Multiple acoustic parameters including duration, mean and maximum of intensity and F_0, minimum F_0 and F_0 slope of the target words were measured and compared across conditions to determine (a) the frequency with which an acoustic parameter was used (i.e., the percentage of data showing modifications in a particular acoustic parameter) to produce prominence, and (b) the extent of the modification (i.e., the ratio between non- prominent and prominent conditions) of that acoustic parameter.

In the perception experiment, native speakers of Mandarin Chinese (N=20) perceived two digitally modified prominent tokens (of the target word) in each trial and chose the one that sounded more natural to signal prominence. The tokens were modified by adopting one acoustic parameter exclusively at a time to signal prominence. In other words, original target words produced in prominent conditions in the production experiment were replaced by its own modified version with only one 'prominent acoustic parameter' fully realized and played to native Mandarin Chinese listeners for 'preference' judgment. Therefore, listeners' selection of a token indicated the acoustic cue they preferred or adopted in prominence perception.

Main Results

The results found in this study were consistent with previous studies regarding general realizations of prominence in Mandarin Chinese. That is, similar to previous studies, the results obtained from this study indicated that:

- Duration and F_0 were the primary acoustic parameters to implement prominence, while intensity was secondary.
- Modifications in F_0 were observed in Tone 1, Tone 2 and Tone 4, but not Tone 3.
- Focus was more fully realized without the presence of accent.

However, this study also yielded findings that have not yet been reported in previous studies. Specifically, the results obtained from this current study revealed that:

- Focus realization made use of more acoustic parameters than accent.
- Lexical tones differed in terms of acoustic parameters implementing prominence.
- For an acoustic parameter adopted by more than one lexical tone, tones differed in terms of the percentage of data to which the parameter applied and the extent of modifications on that parameter.
- Acoustic parameters used in the realization of accent in an unfocused position were modified to a larger extent than in a focused position.
- The ranking of acoustic cues used to perceive focus was different from the ones used to produce focus.

Outline

The remaining of this dissertation will be organized as followed. In chapter two, background of the study will be introduced. General information and previous literature on phonetic studies of prominence in Mandarin Chinese will be presented in Chapter Three. In Chapter Four, the production experiment designed to investigate Research question 1 ‘acoustic parameters used in focus and accent realization’ will be described, and the data will be presented and analyzed to provide answers to this research question. In the following chapter, Chapter Five focuses on Research question 2 ‘interaction among tone, accent and focus in realization’. The perception experiment will be described in Chapter Six to answer Research question 3 ‘the ranking of acoustic cues in prominence perception’. In the last chapter, Chapter Seven, general discussions based on the analyses of production and perception experiments are provided. Results will be discussed with previous studies and the whole dissertation will be concluded with potential areas for future exploration.

CHAPTER 2
PHONETICS AND PHONOLOGY OF LEXICAL TONE, ACCENT AND FOCUS

In this chapter, general concepts of tone, accent and focus will be firstly elaborated. Models and approaches to describe tone, accent and focus will also be discussed in this section. Next, the phonological interactions of tone, accent and focus will be explained. Then, acoustic parameters used to signal phonological interactions will be introduced. Finally, interactions among acoustic cues used in tone, accent and focus perception will be discussed.

Lexical Tone

In all languages, vowel height and consonantal place of articulation are central to conveying the meanings of words. Among them, a subset of languages also makes use of the pitch (height and/or contour) to distinguish the lexical meaning of one word from another. These languages are called 'tone' languages. In Cantonese, for example, the syllable [yau], can be said with one of six different pitches, and has six different meanings: with a high level tone, it means 'worry'; with a high rising tone, it means 'paint (noun)'; with a mid level tone, it means 'thin'; a low level tone means 'again'; a very low level tone means 'oil'; and a low rising tone means 'have' (Yip, 2002). These 'tones' are defined both by their pitch height or 'registers' (e.g., high, mid, and low) as well as their pitch contours (e.g., level, falling or rising). In Vietnamese, a word can be pronounced with one of the six tones and the meaning of the word changes (Thompson, 1987).

Table 2-1. Words [ma] in Vietnamese

Tone	Pitch height	Pitch contour	Gloss
Ngang	high	level	'ghost'
Huyền	low	falling	'but, nevertheless'
Ngã	high	creaky rising	'horse'
Hỏi	low	falling-rising	'grave, tomb'
Sắc	high	rising	'cheek'
Nặng	low	creaky falling	'rice seedling'

In longer words, it matters where the tones go. For example, in Lingala, a Bantu language spoken along the Congo River between Lisala and Kinshasa, a multisyllabic word can be low-toned among all syllables, or have a high tone somewhere in that word, and the meaning changes completely (Guthrie & Carrington, 1988). The acute accents indicate a high tone in Table 2-2.

Table 2-2. Words [moto] and [kokoma] in Lingala

Word	Pitch height	Gloss
mo.to	low low	'human being'
mo.tó	low high	'head'
ko.ko.ma	low low low	'to write'
ko.kó.ma	low high low	'to arrive'

This is in contrast to stressed languages where pitch is used to indicate relative degrees of prominence among syllables in multisyllabic words. In English, for example, the first syllable in 'national' is perceptually more salient than the last two. The relatively higher degree of perceptual salience of this 'stressed' syllable is represented as being longer in duration, louder in volume and higher in pitch than its neighboring 'unstressed' syllables. A difference in 'stressed' location can be used to differentiate a compounded word, 'a blackboard' from an adjective plus noun phrase, 'a black board'. Stress patterns in English can also be used to contrast meanings of such noun and verb pairs as in 'an export' and 'to export', or 'an address' and 'to address'. In normal statement intonation, 'address (noun)' starts high falling pitch on its first syllable, but 'address (verb)' has the fall on the last syllable. Should we then conclude that these words have high falling tones on different syllables in the lexicon? The answer is no, because the actual pitch of these syllables depends entirely on the intonation pattern of the utterance where they are placed. If the speaker is skeptical when saying the two words, she can use a quite different pitch pattern. For example, 'address (noun)' will have a very low pitch on the first syllable, rising into the second syllable, and 'address (verb)' will have a very low then rising pitch on the last syllable. There is no high pitch in either word in this context. What is constant is that in each word one of the two syllables is more prominent than the other, and attracts the intonation pitch, whether it is the statement's high fall, or the skeptical response's extra low-rise.

Besides, tones are different from pitch used 'to convey "postlexical" or sentence-level pragmatic meanings in a linguistically structured way (Ladd, 1996). Intonation contours or pitch movements over an utterance (a phrase or a sentence) occur in all languages, whether or not they have lexical tone. In English, for example, pitch is used to convey emphasis, contrast, emotion and other paralinguistic information at a larger linguistic unit of phrases and sentences. Falling and rising intonation contours over an utterance are used to distinguish a statement from a question, as well as displaying doubt, anger, fear and other emotions. In other words, when I say 'Tom bought himself a guitar', 'guitar' means 'guitar' whether it has a falling or a rising tone. The pitch used to deliver sentence-level information is not enough to earn a language membership into the class of tone languages.

A significant boost to the study of tonal phenomena was given by Pike (1948), who set out a typology of tone languages and provided means to distinguish tones. According to his definition, only languages in which every syllable has a separate tone can be regarded as tonal languages. Hyman (2006) recently defines tonal languages in a broader sense by including accentual languages (e.g., Japanese) as a sub-type of tonal languages, in which each tone is associated with a particular syllable, but not every syllable requires a tone.

How is Tone Produced?

In the discussion of tone, there are three terms need to be explained first: fundamental frequency (F_0), pitch and tone. Among them, F_0 is a purely phonetic or acoustic term referring to the number of pulses or complete repetitions (cycle) of variations in air pressure per second the signal contains (Ladefoged, 2000; Yip, 2002). In the case of the speech signal, each pulse is produced by a single vibration of the vocal folds and measured in Hertz (Hz) where one Hertz is one cycle per second. Pitch is used as a perceptual term, relating to listeners' judgment as to whether a sound is 'high' or 'low' whether one sound is 'higher' or 'lower' than another and by how much, and whether the voice is going 'up' or 'down'. The relation between the auditory pitch and the acoustic F_0 is not linear. For listeners to judge that one sound is twice as high as another, the frequency difference between the two sounds is much larger at higher absolute frequency, e.g., 1000Hz is judged to be double 400Hz, and 4000Hz is judged to be double 1000Hz. But F_0 values in speech are all relatively low (i.e., usually less than 500 Hz), so pitch can be equated with F_0 (Cruttenden, 1986). Tone, on the other hand, is a linguistic term. It refers to a phonological category that distinguishes two words or utterances, and is thus applied only to languages in which pitch plays some sort of linguistic role. In this study, 'F_0' and 'pitch' are used to describe tone production and perception respectively.

The production of tone is dependent on fundamental frequency or F_0. For distinct tones to be perceptible, the signal must contain F_0 fluctuations large enough to be considered as pitch differences. The F_0 fluctuations or differences are determined by adjusting the mass and stiffness of the vocal folds inside the larynx so that the frequency of vibrations changes (Hirose, 1997). When the crico-thyroid muscle contracts, it elongates the vocal folds, decreasing their effective mass and increasing their stiffness. This action increases the frequency of vibration, and thus raises F_0 in tone languages. On the other hand, when the activity of the crico-thyroid muscle is reduced, while the thyro-arytenoid muscle contracts, thickening the vocal folds and increase their effective mass, the pitch is lowered (Yip, 2002). Besides internal changes to the larynx, some other articulatory mechanisms may also contribute to F_0 control. The main one is larynx lowering. According to Ohala (1978), lowering the larynx may play an important role in lowering pitch, because it stretches and thins the vocal folds.

Tone Languages in the World

There are three main linguistic areas of tone languages in the world: (a) certain clusters of American Indian languages (e.g., Otomanguean, Mixtec, Mazatec); (b) the vast majority of African languages (e.g., Sukuma, Yoruba and Xhosa); and (c) almost all of the languages of the Sino-Tibetan family together with many neighboring languages of Southeast Asian (e.g., Mandarin Chinese, Thai, Vietnamese) (Woo, 1969; Yip, 2002).

Linguists working in different geographical areas have developed different traditions in tonal notation. One of the commonalities is that tone is nearly always transcribed on the syllable nucleus, which is usually a vowel. Starting from area (c) where the majority of Sino-Tibetan family languages are tonal languages, tones are shown numerically in a system known as the 'Chao tone letters', based on work by Chao (1930). These are numbers that divide the natural F_0 range of the normal speaking voice into five levels, with 1 as the lowest and 5 as the highest. Each syllable is given digits, written after the segmental transcription. Most syllables are given two digits, one for the starting F_0 and one for the ending F_0. This is true even for level tones. Three digits are used for tones which change direction in the middle of the syllable. For example, [ta] with a high level tone is noted as ta55, with a high rising tone is ta35, and with a low falling-rising tone as ta214. The central Americanists in area (a) also use numbers to describe tones, but the digits are reversed, so that 5 shows low tone and 1 shows high tone. For level tones, only one digit is used. For example, [si] with a high level tone is shown as si1, and with a high rising tone is si32. Africanists in area (b) convey tones by a set of accent marks. Acute accent (á) is used for high tone, grave accent (à) for low tone and level accent (ā) for mid tone. If a tone is unmarked in the language, no accent will be superimposed.

Besides the difference in tonal notation, tone systems of area (c) differ from those of area (a) and (b) in terms of the number of tones in the system and the mobility of tones when interacting with other aspects of the language. For example, Thai of area (c) has five phonemic tones including both level and contour tones (i.e., high, mid, low, rising and falling tones) while Xhosa of area (b) has only two level tones (i.e., high and low tones). Moreover, tones in Thai are almost exclusively used lexically (There is no interaction between tonal distribution and the syntactic or morphological aspects of the language.), while the high tone position in Xhosa is determined by the verb stem domain and the stress system of that language (Downing, 2003), as shown in Table 2-3 and 2-4.

Table 2-3. Words [k^ha:] in Thai tones (Wayland & Guion, 2003)

Tone	Pitch contour	Pitch height	Gloss
Mid	Level	Medium	‘to be stuck or lodged in’
Low	Level	Low	‘a kind of aromatic root often used in Thai cooking’
Falling	Contour	High to low	‘I, servant’
High	Level	High	‘to engage in trade’
Rising	Contour	Low to High	‘leg’

Table 2-4. Tone shifts in Xhosa (Downing, 2003)

Tone shifts	Examples
High tone of the object prefix shifts to the low verb stem	Stem: ndi-ya-[xoleela ‘I forgive’ ([indicates the verb stem edge) Object prefix: <u>kú</u>- ‘you (object)’ ndi-ya-<u>ku</u>- [xóleela ‘I forgive you’ (the high tone sponsor is underlined)
High tone avoid stressed position	Low-toned verbs in the present, short form preceded by High-toned subject prefix *ba*- ‘they’ ba-[qonondísa ‘they emphasize...(clause)…’ When the penult of a word is lengthened under stress-accent, high tones shift to the antepenultimate syllable instead of shifting further right (to the penult) to avoid the syllable which is prominent for stress-accent ba-ya-[qononóndiisa ‘they emphasize.’

Tone Features

For the past five decades, a number of phonologists have proposed phonological features to account for the patterning and distribution of tones. Among these models, I will first introduce the feature models. The following sections deal with the markedness models and the perceptual models. These sections will be followed by a section on the geometric relation between binary features, Register and Pitch (comparing the approaches of Bao, Clements, Hyman, Shi, and Yip). These models differ in their perspectives from which tones are viewed. Feature models, which serve as the basis of other models, deal with tonal differences in production. Markedness models explain why certain tone features are preferred than others. Perceptual models include articulatory and perceptual considerations in the description of tone systems, and explain why certain tones are preferred to others when both are unmarked or marked. Finally, the tone geometry models focus on the relationship among tone features and discuss the internal structure of tones.

Feature models

It has been known for years that the smallest units of phonological structures are not phonemes, but the properties or distinctive features that make up those sounds. The syllable [bu], for example, is represented as two sounds [b] and [u]. [b] is a symbol for a voiced bilabial stop consonant, and [u] is a symbol for a high, back, rounded vowel. When converted to a binary feature descriptions, [b] is [+anterior, -coronal, -cont, +voice], and [u] is [+high, +back, +round]. If the contrast implicit in the description of the sound is a two way contrast, such as voiced and voiceless, rounded and unrounded, then a single binary feature[+/-voice] or [+/-round] will do the job. If the contrast is multi-valued, such as vowel height, which need to distinguish high, mid and low levels, two features [+/-high] and [+/-low] will be needed (high vowels are [+high,-low], mid vowels are [-high, -low], and low vowels are [-high, +low].

Tones are also properties of sounds, and need the appropriate features to explain their behavior. Feature models consider prosodic features, such as F_0, duration and intensity, as the basis to distinguish tones. Tones are mostly analyzed in terms of F_0 level, F_0 contour and intensity to describe the tonal alternations in the language and to provide the abstract basis from which physical phonetic interpretations can be made. For example: features of F_0 level were described as [+/-high] [+/-low] [+/-central] in Sampson's work (Fox, 2000), or as [+/-high] [+/-low] [+/-modify] in Woo's work (Fox, 2000) as shown in Table 2-5; features of contour were depicted as [+/-rising] [+/-falling] (Gruber, 1964) as shown in Table 2-6 , and features of intensity were analyzed as [+/-maximal] [+/-medial] [+/-minimal] (Trager, 1941).

Table 2-5. Woo's feature system to describe level tones

tone / samples / features	55	44	33	22	11
[high]	+	+	-	-	-
[low]	-	-	-	+	+
[modify]	-	+	-	+	-

Table 2-6. Gruber's feature system to describe contour tones

tone / samples / features	55	35	214	51
[rising]	-	+	+	-
[falling]	-	-	+	+

There exists some weakness in feature models. First of all, internal redundancy is inevitable. For example, linguists need seven binary features (i.e., [contour] [high] [central] [mid] [rising] [falling] [convex]) to describe a total of thirteen tones available in the world's languages (Wang, 1967), but the seven features involved can technically specify up to 128 distinct tones, which indicates a considerable amount of redundancy among the features. Secondly, feature models allow us to deduce what tones are permitted in a language, but do not indicate which tone is favored among them. Therefore, the models could not explain why certain features (e.g., [high]) are exploited more than others (e.g., [contour] and [convex]). Neither could they explain why a four-tone paradigm always has some contour tones, even though many languages do distinguish among four non-contour tones. The second weakness is remedied by Wang's (1967) markedness model and Hombert et. al's (1979) perceptual model.

Markedness model

To describe tone preference, the markedness model (Wang, 1967) applies the 'marking conventions' to tone systems. Each feature can be labeled as 'unmarked' or 'marked' in addition to the binary values. For example, [-contour] or [-central] is unmarked while [+contour] or [+central] is marked. The more marked a tonal system, the more complex the system and the more tones it contains (assuming that the presence of a marked token presupposes the presence of its unmarked counterpart). This knowledge derives primarily from observations of three sorts: the frequency of distribution of the sounds in the languages of the world, the patterns of historical change in sound systems, and the acquisition of sounds in children and the dissolution of sounds in linguistic pathology. Therefore the complexity assigned to tones based on markedness may reflect an integrated effect of perception, production, and learnability (Ke, Ogura, & Wang, 2003).

Perceptual models

Hombert et al. (1979) add perceptual consideration to their model, which aims at maximizing perceptual distance to search for phonetically optimal tonal systems. Contour tones covering a small F_0 range are more difficult to perceive than tones ending at an extremity of the F_0 range. Average F_0, F_0 onset, offset and slope are included in perceptual judgment to keep two closest tones of a system maximally apart. This is a first attempt to predict tone shapes if the number of tones are known in a system from a perceptual perspective. However, this model considers a contour tone as a combination of two level tones, which, as a result, excludes tones involving three levels (e.g., the dipping tone in Mandarin Chinese). Moreover, only pitch cues are considered perceptually and no consideration was given to other possible cues in tone perception, such as duration.

Tone geometry models

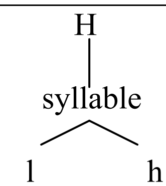
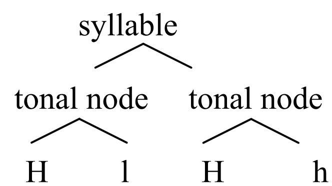
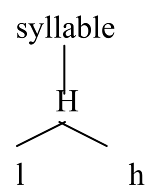
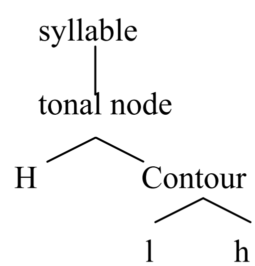
In the early 1980s, it was suggested that distinctive features were not just a list, but the terminal nodes in a structured tree. For example, the features relating to voice of articulation formed a constituent called Voice, and this constituent was a phonological entity which could spread or delete. Since Yip's (1980) feature proposal, phonologists have explored that tonal features could also be organized into a multi-tier structure and provide explanations for tonal changes. Tone geometry models represent a significant theoretical departure from early generative phonology in the number of features postulated and their relationship. The models view tones as independent entities, a multi-tiered representation with intricate internal structure, identifying the similarity and the difference among tones in a system and explaining how changes take place inside a tone.

In Yip's theory, a tone is not an indivisible entity. Rather, it consists of two parts, Register and Tone. Register features indicate an imagined band of F_0 in which a tone is realized, and Tone features specify the way the tone behaves over the band. The concepts of Register and Tone are later adopted by many other studies (e.g., Clements (1981), Shih (1986), Hyman (1993), Bao (1999), etc.), though Tone is referred to as Contour in some cases[1]. The main difference among all these studies lies in the relation between Register features and Contour features. If we use a high rising tone for example, in Yip's (1980) work (shown in Table 2-7 a), the register features and the contour features are entirely independent of each other, and there is no tonal node dominating them. In Duanmu's (1990, 1994) and Clements' (1981) (shown in Table 2-7 b), the

1 Contour will be used in following discussions to avoid confusion between Tone and tone (in general).

register and the contour features are sisters under a tonal node, and each half of the contour tones is entirely independent, which implies that a contour tone is a concatenation of two level tones. In Yip's (1989) and Hyman's (1993) work (shown in Table 2-7 c), the register feature is the tonal node, dominating the contour features, which implies one register feature for one tone. In Bao's (1990) work (shown in Table 2-7 d), the contour features are dominated by a node of their own, called Contour, which is a sister of the register feature, and both are dominated by a tonal node.

Table 2-7. Types of tone geometry models[2]

Types of tone geometry models	Example of a high rising tone
a	
b	
c	
d.	

[2] Register features are shown in capitalized characters, and contour features are shown in small characters.

In all cases, contour cannot change dynamically in model (b), register cannot change without affecting contour features (and vice versa) in model (c). The whole tone can change as a unit only in (c) and (d), and the contour can change as a whole without affecting register features only in (d).

Accent

In an autosegmental model, sentence accent is defined as nuclear pitch accent, which is consistently realized as a high tone, either on a final syllable or a heavy syllable within the last word of a phrase. For example, in Chickasaw, a Western Muskogean language spoken in south-central Oklahoma, sentence accent is assigned to the final (stressed) syllable [ʃa:] in [katimihtã saha'ʃa:] 'Why am I angry?' and the (non-final) heavy syllable [li:] within the last word in [naʃo'ba:t ma'li:ta] "Does the wolf run' (Gordon, 2005). Accent has many synonymous terms, such as primary accent and tonic accent, which designate one stressed syllable as more prominent than other stressed syllables in a stretch of speech (Cutler & Ladd, 1983; Buring, 1997). Liberman and Prince (1977) name it 'designed terminal element', because accents alternate and contrast with less prominent portions syntactically, creating a series of accentual phrases delimited by accents. The boundary distribution of accents is also perceived by Brown (1980):

> "In pragmatically neutral speech, the last stressed syllable in the phrase will normally be more prominent than preceding stressed syllables."

This statement implies the subtle difference between accent and stress that stress is usually related to word level, while the domain of accent is phrase and sentence levels. Compared with word stress, sentence accent does not refer primarily to the properties of individual segments (or syllables) but rather reflects a hierarchical rhythmic structuring that organizes the morphemes in an utterance into larger prosodic structures (Garde, 1968).

Early descriptive linguists describe sentence accent from the view point of physical properties. The physical properties attributed to accent are stated in Sweet (1906)'s definitions of 'stress' and 'force':

> "physically force is synonymous with the effort by which breath is expelled from the lungs… acoustically it produces the effect known as 'loudness' which is dependent on the size of the vibration-waves which produce the sensation of sound…The comparative force with which the syllables that make up longer group are uttered is called stress."

Jones (1950) who also distinguishes stress as force of utterances agrees with this idea. However, even these two phoneticians cast some doubt on the validity of the phonetic delimitation of the category 'stress', because the linguistic 'stress' does not correspond exactly to physical 'stress' or 'force'. Sweet claims that the discrimination of degree of stress is not an easy matter in any case, because of associations of intonation and vowel-quality, leading listeners to think that high intonations or clear vowels (as the opposite of breathy vowels) possess a stronger degree of stress than they really have.

Starting from Bloomfield's primary and secondary phonemes[3], structuralists describe accent as a phonological category, but is limited merely to its distinctive function (Trager, 1941; Hockett, 1955, 1958). They recognize that the single phonological function of accent is to distinguish meanings and differentiate accent languages from tonal languages. After that, Trubetskoy (1969) first explicitly states that accent has other functions besides the distinctive one, which are to organize prosodic units in an utterance and to mark the syntactic boundaries between prosodic units. However, the distinctive function is still claimed to be the primary function of accent.

Later functionalists propose that the primary function of accentual contrasts is to phonologically unite cohering morphemes and to set up larger groups of words and phrases in an

[3] Primary phonemes are segmental phonemes, while secondary phonemes are supersegmental, not fixed to any particular segments. For example, tone languages use features of pitch as primary phonemes.

utterance (Martinet, 1954; Garde, 1968). They state that prosodic properties are not necessarily to serve a distinctive function. Accent could be an organizational feature extended beyond words to a larger pattern that contrasts words within phrases, smaller phrases within larger phrases, and even larger organizational structures within the level of entire utterances.

Focus

Chomsky (1971) claims that focus is a reflex of phonology, and is determined by the intonation center of the surface structure. Intonational focus is usually divided into broad focus and narrow focus (Frota, 2000). Broad focus is often referred to as (new) information focus (which conveys new, non-presupposed information) (King, 1995; Kiss, 1995) and focuses on whole constituents or whole sentences (Ladd, 1980; Gussenhove, 1983; Schmerling, 1976); narrow focus is usually localized to individual words and referred to contrastive information that distinguishes itself within a set of contextually given alternatives that may occur in the same position in spontaneous speech (Drubig & Schaffar, 2001; Lehiste, 1970). Particularly in the prominence patterns of European languages (shown in Table 2-8), broad focus is commonly equated with neutral intonation, and narrow focus with marked accent[4].

Table 2-8. Two types of focus in English

Types of focus	Examples in English
Broad focus	They [participated in the lexical tone perception experiment] $_{\text{Broad Focus}}$ yesterday. (As an answer to 'What did the students in the Linguistics Department do yesterday?')
Narrow focus	No, it is students in [the Linguistics Department] $_{\text{Narrow Focus}}$ who participated in the lexical tone perception experiment yesterday. (As an answer to 'Is it students in the History Department who participated in the lexical tone perception experiment yesterday?')

[4] For the rest of the description, 'accent' will always refer to normal sentence accent and 'marked accent' to narrow focus.

Focus is usually described in one of two approaches: the highlighting-based approach and the structure-based approach. The highlighting-based approach relates focus to discourse context and speaker intention, and depends on a pragmatic factor called ‘radical Focus-To-Accent (FTA)’, which conceives that focus signals discourse salience and is unpredictable without reference to speaker’s intentions. The approach does not explain why words with neutral intonation pattern can also be accented though they are not pragmatically focused. In the structure-based approach, the speaker’s decision about what to be focused is subject to all kinds of contextual influence (such as syntactic, semantic and/or pragmatic prominence). Once the focused part of the utterance is specified, the marked accent pattern follows more or less automatically by language-specific rules. This approach allows for the existence of a neutral intonation, an ‘unmarked’ or ‘default’ pattern. In such a pattern, the whole sentence is a broad focus and the location of unmarked sentence accent is specified according to semantic rules. For instance, Gussenhoven’s (1984) Sentence Accent Assignment Rule (SAAR) claims that (i) the semantic constituents: Argument and Predicate, when adjacent, merge to form a single focus domain, and (ii) that within this composite domain, accent is carried by the Argument. This implies that broad focus has scope over the entire utterance, larger than the accented word. The accent placement obeys structural principles. It works well in explaining how focus interacts with syntactic and phonological organizations (Shown in Table 2-9.).

Table 2-9. Example of neutral intonation

Example of neutral intonation in English
A: How much did they pay you for participating in the experiment? B: FIVE FRANCS.

In B’s answer, both ‘five’ and ‘francs’ are accented. ‘Francs’ is almost entirely predictable if the conversation takes place in a country where the unit of currency is the franc, while ‘five’ is the new information. According to the structure-based approach, unmarked sentence accent is assigned to ‘francs’ in a boundary position by rules, while the highlighting-based approach can not provide explanation for accented ‘francs’.

To summarize the concepts of tone, accent and focus (Shown in Figure 2-1), tone is a segmental phoneme assigned to syllables to distinguish lexical meaning. Accent is the result of the operation of phonological rules on surface syntactic structures (Newman, 1946), assigned to syntactic boundary positions. Focus is a suprasegmental phenomenon in at least sentence level, to signal new and/or contrastive information. In a tonal language, tones are default while sentence accent and focus are optional. Accent, in most cases, locates itself near syntactic boundaries regardless of whether the sentence gets a neutral or a focused intonation. When focus is added, it can be assigned to any part of the sentence. Both accent and focus are suprasegmental representations, but can ultimately be localized to specific segments, comparable to tone.

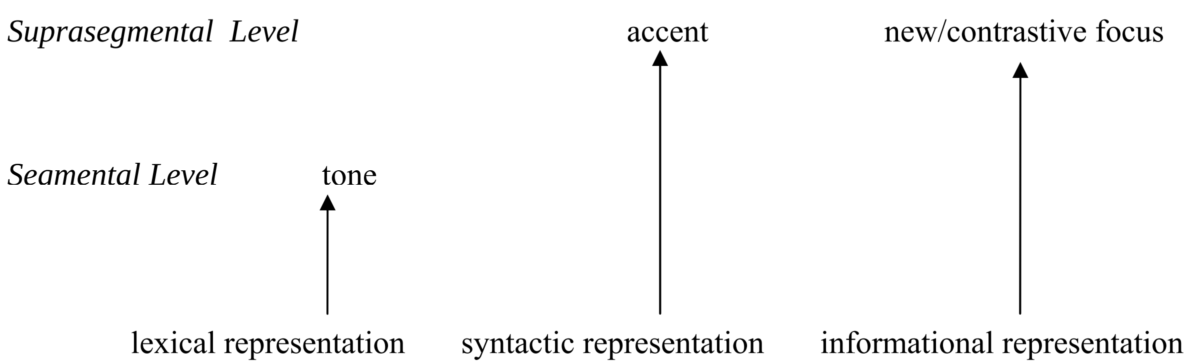


Figure 2-1. Concepts of tone, accent and focus.

Phonological Interactions among Tone, Accent and Focus

In tonal languages, tone bears a close relation with sentence accent and focus. Lexical tone is the most obvious phonological input at the word level, but it is by no means the only input for that word. Drawn from the autosegmental approach for accent and the structure-based approach for focus, once the lexical item is put into a sentence, it may obtain sentence accent as well as focus depending on its position in a syntactic structure and the information it carries. For a tonal language where the default position for sentence accent is sentence final, there are three possible interactions among tone, accent and narrow focus (Shown in Figure 2-2).

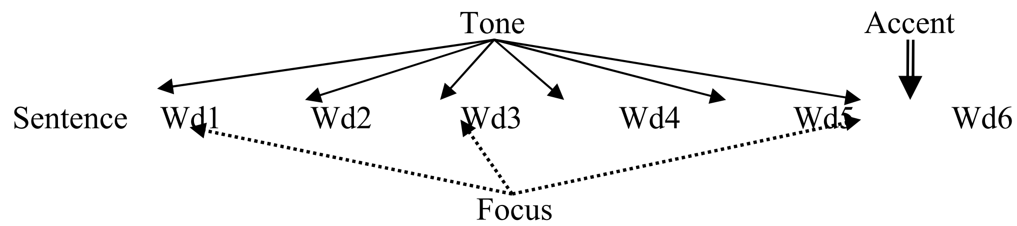


Figure 2-2. Phonological interactions among tone, accent and focus

In Figure 2-2., the sentence consists of six words. Tones are assigned to all words and accent to the sentence final word. Focus is optional, and can be placed on any part of the sentence (i.e., sentence-initial, middle or final word). When focus is not placed, interaction between tone and accent is shown on the sentence final word (i.e., Wd6). When focus is added, there will be interaction among the three (if added to Wd6) or just between tone and focus (if added to other places, e.g., Wd1 or Wd3).

There are usually two ways to deal with the phonological interactions: one is to avoid having tone, accent and focus at the same position. For example, in Chinese and Hausa, focus is realized by emphasis markers to retain tonal intactness; final positions in Otomi are reserved for accent with tones shifting forward. The other solution is to allow tone, accent and focus to be assigned simultaneously, but accent and focus are phonetically implemented in a more restricted fashion in tonal languages than in non-tonal languages. For example, register adjustment is applied to indicate interactions in Mandarin Chinese and Taiwanese (i.e., register is expanded in Mandarin Chinese, while it is being raised overall to a higher level in Taiwanese). However, compared with non-tonal languages, where the entire F_0 register can be moved up and down due to accent and focus, the mechanism cannot be given as free a rein in tonal languages, since lexical tones must remain at least somewhat retrievable to keep tonal features.

Optimal Theory Treatment of Tone, Accent and Focus

Before 1990s, most phonological studies were conducted using the rule-based derivational theory proposed by Chomsky and Halle (1968). Prince and Smolensky, in 1993, proposed a non-derivational approach called Optimality Theory, or OT to analyze differences between the phonological input and the phonetic output (Prince & Smolensky, 1993). The OT theory argues that the output is selected by direct evaluation by various criteria or constraints5. These constraints are universal and violable, but ranked differently in languages. For each language, violations of higher-ranked constraints are fatal, and the winner is the output candidate that survives this winnowing (Archangeli & Langendoen, 1997, Kager, 1999). Recently, tones have been given OT treatments to describe behaviors such as tonal shifting, spreading, alignment (Akinlabi & Liberman, 2000; Cassimjee & Kisseberth, 1998; Myers, 1997; Silverman, 1997; Zhang, 2002; Zoll, 1997). De Lacy (1999) has applied OT to study the interaction between tone and phonological categories such as stress, and posited constraints to deal with the phenomena crosslinguistically that stressed positions prefer high tones and avoid low tones (such as the insertion of a high tone on a stressed syllable in Lithuanian, the movement of a high tone to a stressed syllable in Zulu and Digo, and the tendency for a stressed syllable to avoid low tone in Golin and Mixtec).

OT treatment has also been given to describe right-most accent and discourse new/contrastive focus (Selkirk, 2002; Féry & Samek-Lodovici, 2006; Samek-Lodovici, 2005). In

[5] There are two types of constraints: faithfulness constraints and the markedness constraints. The former encourages underlying tonal forms to resist change (e.g., no insertion of tones, no deletion of tones), and the later encourages more basic and natural forms (e.g., no contour tones, no low tone on heads).

Samek-Lodovici's (2005) study, he examines the prosody-syntax interaction in the expression of focus. He claims that prosodic and syntactic constraints conflict with each other in the expression of focus, where the best position for main sentence accent (rightmost positions) does not necessarily match the best syntactic position for the focused constituent (in situ positions). But focus and stress must match, therefore if stress and focus are not in the same position, either stress or the focused constituent must renounce their best position violating either the syntactic or the prosodic constraints responsible for them. For example, STRESSXP (lexically headed XP must contain a phrasal stress), HP (Align the right boundary of every P-phrase with its heads), HI (Align the right boundary of every I-phrase with its heads) are prosodic constraints; while Stay (No traces) and EPP (Clauses have subjects) are syntactic constraints. In English, when a subject is focused in situ (e.g., 'JOHN$_f$ has laughed.' as an answer to 'Who has laughed?'), the syntactic constraints are ranked higher than prosodic constraints, because the output places a focus at the sentence initial position and violates the rightmost position guaranteed by prosodic constraints. In Italian, syntactic constraints are ranked lower (than prosodic constraints) and violated to correctly express a focused subject, which is moved to the rightmost position of a sentence (e.g., 'Ha riso GIANNI$_{f.}$'[6] as an answer to 'Who has laughed?'). The study argues that human language addresses this tension in optimality theoretic terms and that different focus paradigms across different languages reflect different rankings of a shared invariant set of syntactic and prosodic constraints. In Féry and Samek-Lodovici's (2006) work, they propose discourse constraints to explain how nested-foci in places other than sentence final become most prominent. The discourse constraints, such as SF (A focused phrase has the highest prosodic prominence in its focus domain) and DG (A given phrase is prosodically nonprominent) are

[6] The literal meaning of the sentence is 'Has laughed JOHN' and the English gloss for this sentence is 'John has laughed.'.

ranked higher than prosodic constraints: HP, HI and STRESSXP. As a result, when an utterance does not contain a focus, default accent is assigned rightmost, and when it delivers focus, the focused part is most prominent. Both work provide explanations for focus in situ from syntactic and discourse perspectives.

Besides the work mentioned above which deals with the distribution of focus, investigations on phonetic correlates to focus have also been conducted (Face, 2001; Selkirk, 2002). Face (2001) argues that, in Spanish, early F_0 peak (L+H)*[7] is the result of a focal pitch accent. In addition, it is shown that this is not the only strategy in Spanish for conveying narrow focus through intonation, as increased F_0 peak height may also be used. Selkirk (2002) also claims that contrastive focus gains prominence which is implemented by a L+H* pitch accent. Moreover, a following phonological phrase break, marked by both a L- phrase accent and temporal disjuncture is observed.

Phonetic Representation of Prominence in Tone languages

At the phonetic level, accent and focus are perceived as linguistic prominence, which is defined as words or syllables perceived auditorily by listeners of the given language as standing out from their environment (Terken, 1994). Prominence is usually examined through changes in F_0, duration and intensity at the acoustic level. I have defined F_0 in the section "How is Tone Produced?", and will spend a little time introducing the other two acoustic parameters (i.e., duration and intensity) that are most consistently used for prominence realization, either singly or jointly.

Duration is usually described in msec (millisecond, which is the cycle time for frequency 1 kHz) in speech production. There is little difference whether we view it as the length of time speaker decides to continue to produce a linguistic unit, or the length of time during which a

[7] (L+H)* Indicates the alignment of both tones to the stressed syllable.

listener hears that unit. Hence, we do not differentiate duration in production and length in perception in this study. The word 'duration' is used to for both purposes.

Intensity is proportional to the average size, or amplitude, of the variations in air pressure. It is an acoustic property, usually measured in decibel (abbreviation as dB) relative to the amplitude of some other sounds. Just as duration is the acoustic measurement most directly corresponding to the length of a sound, F_0 is one corresponding to the pitch, so intensity is an appropriate measure corresponding to loudness in perception. The relation between absolute intensity to perceived loudness is not linear, but generally a higher intensity leads to a louder sound, and the lower intensity makes the sound smaller. Therefore, in this study, intensity and loudness are equated with each other. Acoustic 'intensity' is used to describe both physical and auditory properties.

Among the three acoustic parameters, pitch is also the most reliable phonetic cue to perceive sentence accent in English (Fry, 1958). To be more detailed, pitch range, not the absolute pitch height, plays a key role in stress perception (Moore, 1993; Shih, 1988). Besides pitch, Ladd (1996) and Gussenhoven (2004) argue that the phonetic correlates of sentence accent can expand to a longer duration of the stressed syllable. This argument is supported by Beckman (2006) who also agrees that the phonetic properties associated with accent at any level (unmarked or marked) are F_0 and duration, which implied that focus at the sentence level is also related to F_0 movement and tempo changes.

Similar to 'stress' language like English, accent and focus is also used in lexical 'tone' languages to convey emphasis, contrast and prosodic boundaries. When tone, accent and focus are concurrently realized in an utterance, acoustic parameters serve more functions than contrasting lexical meanings and are likely to get modified to realize prominence caused by accent and/or focus. For example, when three intonational patterns (general rising, falling, and a mixed pattern) are assigned to five Thai tones, the behavior of each tone changes when superimposed by intonation, and the systems of tone and intonation interplayed to form the speech melody in spoken Thai (Luksaneeyanawin, 1993). Also in Hausa, a high tone on an individual word is raised to highlight that word (Leben, Inkelas, & Cobler, 1989). An example (shown in Table2-10) with subject focus is taken from their article. High rising is indicated by an upwards directed arrow.

Table 2-10. Example in Hausa where F_0 is raised to highlight a word

Example of raised F_0 to highlight a word
Máalàm ↑**Nú**hù née // yé hánà Láwàn // híirá dà Hàwwá. 'It was Mister Nuhu // who prevented Lawan // from chatting with Hawwa.'

Interactions among Acoustic Parameters in Phonetic Production and Perception

Thinking of phonetic production and perception, no matter it is tone, accent and/or focus , the speech temporal structure integrates all basic acoustic parameters-duration, loudness and fundamental frequency. All speech needs a temporal 'bearer' to carry parameters, such as pitch and intensity to get itself delivered, and the listeners also need these acoustic cues to perceive whether the sound is long or short, high or low, loud or small. It is interesting to ask if there are interactions among the co-existing acoustic dimensions.

Most interactions are shown between pitch and intensity. Regarding speech production, for example, Buekers and Kingma's (1997) study on the impact of phonation intensity upon pitch during speaking claims that pitch appears to rise exponentially with phonation intensity, because the rise results from increased sub-glottal pressure and higher laryngeal muscle effort. The opposite is also tested about the pitch effects on intensity in speaking. It is revealed that with a slight increase in the fundamental frequency, the changes in vocal intensity are considerably greater than at a normal speaking voice (Komiyama et al., 1984). For speech perception, Johnston's (2005) dissertation on the influence of frequency and intensity patterns on the perception of pitch investigates whether exposure to dynamic intensity changes will affect listeners' perception of pitch. In a series of four experiments, listeners hear context sequences of tones that change dynamically in frequency and intensity, and judge whether the pitch of a variable final tone (probe) is the same as or different from the immediately preceding tone. Experiment 1 sequences comprise simple monotonically changing frequency and intensity patterns. In Experiment 2, listeners hear longer sequences that imply periodically changing frequency and intensity patterns. Using the same frequency patterns from Experiment 2, Experiment 3 incorporates regularly recurring intensity accents to investigate whether intensity accent patterns within a periodic frequency pattern can influence pitch judgments and Experiment 4 includes randomly occurring intensity accents to investigate whether temporally irregular accents affect pitch perception. Comparison between Experiments 2 and 3 reveals a significant difference between the pitch perception results, which indicates that pitch perception is affected by the regularly recurring intensity accents.

Tekman's (1995, 1997) studies on interactions of relative timing, intensity, and pitch in the perception of rhythmic structures suggests that rhythmic manipulation of one dimension of sound can create changes in perception of other dimensions of sounds that conform to the same temporal structure. For example, F_0 manipulations are found to change perceived intensity. He explains that the listeners do not discriminate the specific physical variations that created changes in rhythmic structures. In other words the physical manipulation can substitute for each other to get similar impression in auditory properties.

Interactions between duration and pitch, duration and intensity are also observed in perception studies. When three sounds share the same physical length and the pitch level, a rising contour is perceived as being longer than the level pitch, and the level pitch is also longer than the falling contour (Rosen, 1977). Also in a speeded classification experiment, listeners perform faster when one acoustic cue is companied by another cue in a positive fashion. For example, listeners' classification of duration is faster when the sound constantly has louder intensity, or higher F_0. On the other hand, their classification of intensity and pitch is quicker when the sound is longer (Merala & Marks, 1990). So, there are substantial effects of congruity: attributes from one acoustic parameter are classified faster when paired with 'congruent' attributes from another parameter.

CHAPTER 3
MANDARIN CHINESE AND ITS PHONETIC REPRESENTATION OF PROMINENCE

Mandarin Chinese, the official language of the People's Republic of China, is based on the particular Chinese dialect spoken in Beijing (the capital city of China) and across most of northern and southwestern China. According to the 1999 Ethnologue Survey, the language is spoken by 867million native speakers. It is a tone-language where each syllable has a tone exclusively used lexically, with no interaction with the syntactic or morphological aspects of the language (Wang, 1967). In this chapter, first, the tonal system in Mandarin Chinese will be described from production, perception and formal linguistics perspectives. Next, the representations of accent and focus in Mandarin Chinese prosody are discussed. Then, phonetic representation of prominence in Mandarin Chinese from production and perception perspectives will be reviewed. In this section, phonetic models (i.e., contour model, F_0 range model, and register model) for prominence realization in Mandarin are introduced, followed by a literature review of previous studies on the production of prominence and the mismatches between production and perception of prominence. Finally, gaps in previous research on prominence and the research questions investigated in this study will be addressed.

Mandarin Chinese Tones

Production of Mandarin Chinese Tones

As shown in Figure 3-1, there are four lexical tones in Mandarin Chinese, referred to by their Wade-Giles numbers and by the shaping of their pitch contours as Tone 1 - high-level tone; Tone 2 - mid-rising tone, Tone 3 - low-dipping tone and Tone 4 - high-falling tone (Sun, 1997). When produced in isolation, Tone 4 has the widest F_0 range from the onset to the offset; Tone 1 has a very limited F_0 range since it is a level tone; the F_0 range from the onset to the turning point in Tone 3 is also narrow.

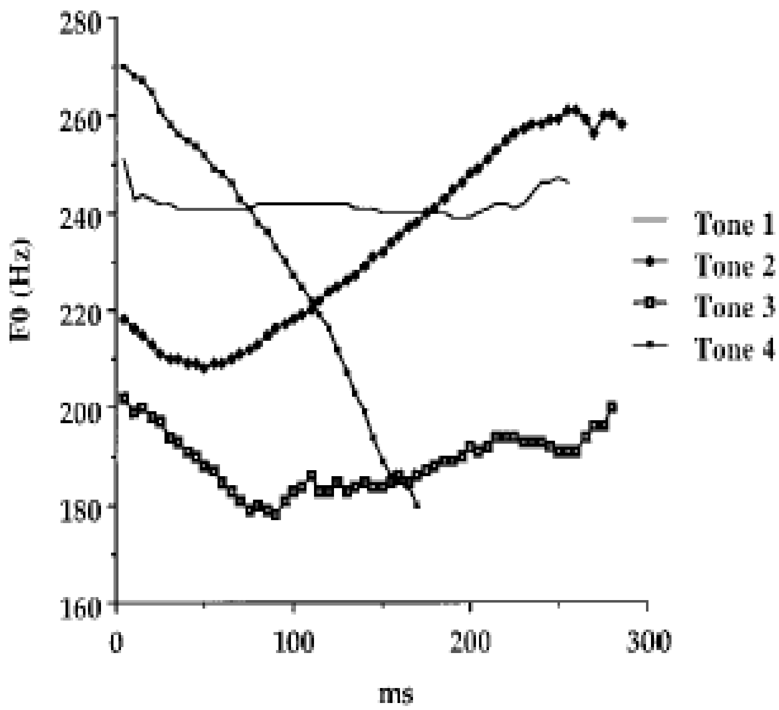


Figure 3-1. Four tones in Mandarin Chinese (Moore & Jongman, 1997)

In connected speech, the F_0 contour of a tone is influenced by the surrounding tones (as shown in Figure 3-2). The most apparent influence is from the preceding tone, whose offset value virtually determines the starting F_0 of the following tone. The influence is assimilatory, that is, a tone with a low offset lowers the F_0 of the following tone, and a tone with a high offset raises the F_0 of the following tone. The magnitude of the assimilatory effects decreases over time: during the initial nasal consonant [m], there are rapid F_0 movements, which are larger when the adjacent values of two neighbouring tones are far apart than when they are more similar to each other; the effects remain sizeable during the vowel, though with reduced magnitude. The high F_0 region seems to be more susceptible to contextual effects, and the lowest F_0 region seems to have strong resistance to the effects.

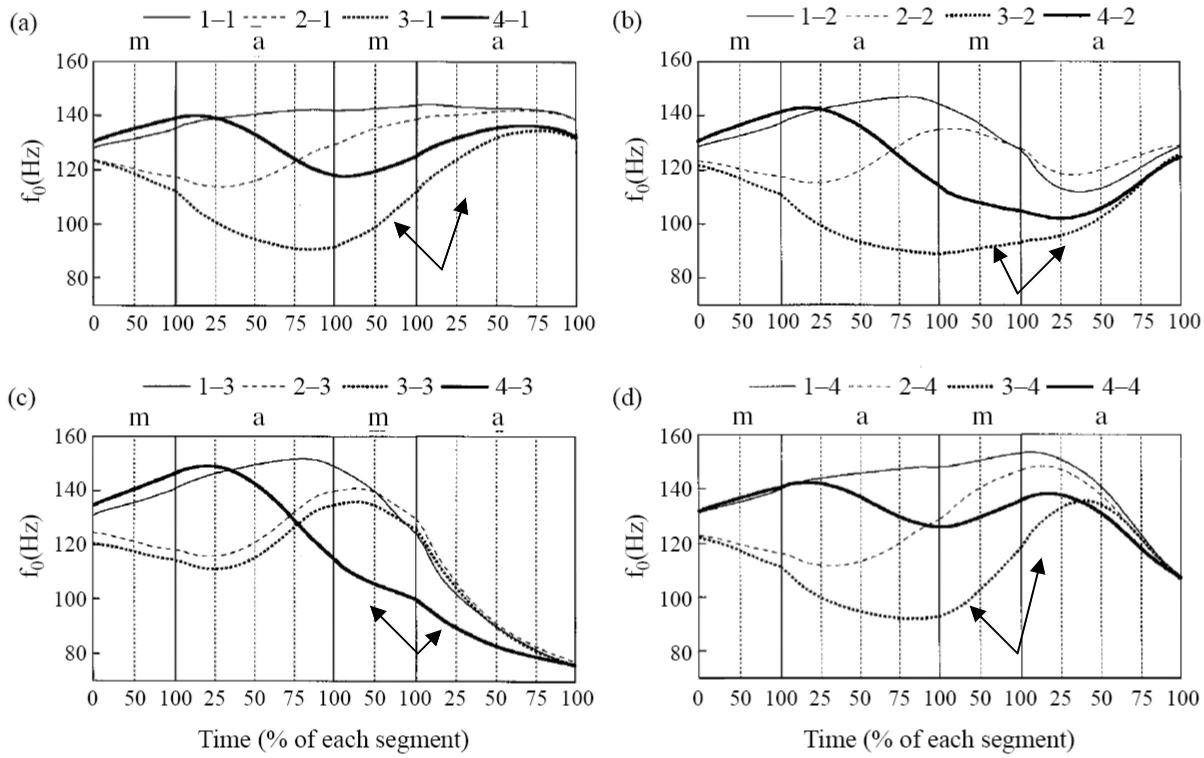


Figure 3-2. Contextual tonal variations influenced by previous tones (Xu, 1997)

Perception of Mandarin Chinese Tones

Work by Gandour (1981, 1984) includes perceptual dimensions to describe tones. Gandour (1981) extracts three perceptual dimensions labeled 'height', 'direction', and 'contour' that are related to listeners' perception of Cantonese tones. Gandour interprets the 'height' dimension to reflect average F_0 level, the 'direction' dimension to reflect the direction of F_0 change, and finally the 'contour' dimension to reflect the magnitude of F_0 change. He (1984) argues that language background affects relative weighting placed on acoustic dimensions, and perceptual cues work integratively to allow for correct identification of tones. English speakers pay more attention to pitch height (e.g., average pitch, extreme endpoint), while listeners of tonal languages (e.g., Chinese, Cantonese, Taiwanese, Thai) pay more attention to pitch contour. Recent study by Khouw and Ciocca (2007) suggests that among the three pitch cues to distinguish Cantonese tones, the direction of F_0 change is used by listeners to perceptually distinguish contour tones and level tones, and differentiate rising and falling tones; the magnitude of F_0 change is used to distinguish tones with the same contour shape but different pitch levels, such as high rising and low rising tones; the average F_0 level cues the distinction among level tones. Similar to Cantonese tones, Mandarin Chinese tones also differ in 'height', 'direction' and 'contour' in perception. Among these dimensions, the direction of F_0 change is crucial to distinguish contour tones (Tone 2, Tone 3 and Tone 4) and the level tone (Tone 1), as well as to discriminate the rising tone (Tone 2), the falling tone (Tone 4) and the falling-rising tone (Tone 3). The pitch height is used to differentiate high tones (Tone 1 and Tone 4), the mid tone (Tone 2) and the low tone (Tone 3). In a word, listeners from different first language backgrounds use different acoustic cues to perceive tones. For a particular listener, s/he may apply different dimensions of pitch to perceive tonal contrasts (depending on the tones in that tonal system).

Formal Description of Mandarin Chinese

Chao's five-scale model

In Chao's five-scale model (1930), a vertical line, analogous to an ordinary F_0 range, is divided into four equal parts to represent five levels of F_0: low, half-low, medium, half-high and high (level1 stands for the lowest level and level 5 the highest). Each Chinese tone has a numerical label consisting of digits denoting the tone's starting, turning and ending F_0 values. For example, a high falling tone without a turning point may be transcribed as 53 (where the starting F_0 value is of level 5 and the ending of level 3); a low-dipping tone with a turning point as 214 (where level 1 is the turning F_0 value). The model provides a convenient method of phonetically transcribing auditory impressions of tone height. However, too many tones can be generated through the combination of five F_0 levels in tonal starting, turning and ending points.

Theoretically, 125 possible tones can be generated. Mandarin Chinese does not contain so many distinctive tones in its tonal inventory. Neither do any other tone languages in the world.

Also, the choice of five levels is not based on phonological principles, but on a balance between phonetic details and phonological distinctions. A distinction between one degree (e.g., 44 and 55, 24 and 35) is usually not significant, so it is common to get two different transcriptions for the same tone. The flexibility causes problems when translating Chao's numerical values into level tone models. For example, Yip (1980)'s model describes tone contours as high (H) and low (L). Level 2 can be an H tone in the lower register, but if it is transcribed as level 3, it may be an L tone in the higher register. Its dubious status between a phonetic system and a phonemic one also allows people to make modification of the phonetic transcription. For example, in Shen (1981)'s work, it is claimed:

> "The real value of Yin Ping is 52. This paper marks it as 53. The real value of Yin Qu is 33 or 24, this paper marks it as 35."

The modification is justified if there is no contrast between 52 and 53; 33, 24 and 35 in the language, the tone can be transcribed in either value. However, if some people modify the phonetic value, and some do not, there are sure to be confusion.

Autosegmental models.

Feature models treat Chinese tones as single-tiered representations with an unstructured bundle of phonological features (Woo, 1969; Wang, 1967). Later studies adopt autosegmental phonology to the Chinese data concerning the internal structure of tones among tonal features (Yip, 1980, 1889, 1993; Clements, 1981; Shih, 1986; Bao, 1999). They use register features [+/-upper] (and [+/-low]) to describe Register; contour features 'H' and 'L' to represent a raised pitch and a lowered pitch. For example, Tone 1 can be transcribed as [+upper, H], Tone 2 as

[+upper, LH][8], Tone 3 as [-upper, HLH] and Tone 4 as [+upper, HL]. The weakness of autosegmental models is the over-generation of tones (though better than the feature models). According to these models, the feature sequence of HLHL is possible under the contour node. However there is no language that contains tones with pitch contours more complex than convexity or concavity. Hence, the models need to have a stipulation that the maximum number of tone feature occurrences in sequence is three, which will allow tones like [-upper, HLH], but rule out such non-occurring tones as [-upper, HLHL].

Besides four lexical tones, Chinese also has a neutral tone, labelled as 0 in Chao's five-scale system. It usually comes at the end of a word or an unstressed position, and is pronounced in a light and short manner. Its pitch depends on the tone carried by the syllable preceding it as shown in Table 3-1.

Table 3-1. Pitch of a neutral tone (Luo & Wang, 1957)

Tone of preceding syllable	Pitch of neutral tone	Example	Gloss
Tone 1(55)	2	tian1qi0	weather
Tone 2(35)	3	fu2qi0	luck
Tone 3 (214)	4	xiao3qi0	stingy
Tone 4 (51)	1	ke4qi0	polite

Prosody in Mandarin Chinese

Prosody of Mandarin Chinese usually contains the following main aspects: rhythm, stress (or accent) and intonation. Perceptually, prosody is referred to the perceived impression of so-called 'the cadence of speech sounds' (Cao, Lu, & Yang, 2000). In natural speech, the three aspects are not completely independent, but integrated with each other, and achieved mainly through the common ground of modulations in pitch duration, and intensity.

[8] For a mid tone, such as Tone 2, its register can be described as either [+upper] or [-upper]. For example a mid level tone is labeled as [+upper, L] or [-upper, H]. Another way to transcribe its register is [-upper, –low] so as to differentiate itself from high tones [+upper, -low] and low tones [-upper, +low] (Bao 1999).

Rhythm is mainly related to the timing behavior of speech, and the rhythmic elements are organized as in hierarchy in terms of particular coherent properties within a unit (Cao, 1999). It consists of three main layers: prosodic word (PW), prosodic phrase (PP) and intonation phrase (IP). Generally, PW is a disyllabic or tri-syllabic word, and it serves as the principal building-block of rhythmic structure. As the intermediate layer, PP is larger than word but smaller than the syntactically defined phrase or clause. IP is a rhythmic group that contains one or more PPs, and is identical to syntactically defined sentence.

Stress is also organized as a hierarchy in terms of the domain investigated, and classified into word stress and sentence accent. The word stress system in Mandarin Chinese is not salient (Wang et. al, 2003). Similar to English, the majority of Chinese words are polysyllabic, especially disyllabic words (Duanmu, 1999). Syllables with one of the four lexical tones are all stressed, compared to those with a neutral tone, which are unstressed (Deng et al, 2004; Duanmu, 1990). As shown in Example (1), the stress contrast at the word level indicates the difference between the neutral tone and the normal lexical tone (Lin et al., 1984; Cao, 1995). Sentence accent in Mandarin Chinese can also be called grammatical or normal accent. In running speech, sentence accent always fall onto certain stressed syllable of a unit that bears semantic or syntactic prominence. More detailed description of accent distribution will be provided in the next section.

Example (1)

Word	Tone combination	Stressed syllable	Gloss
qi1.zi0	Tone 1+ Neutral tone	first syllable	wife
hou2.zi0	Tone 2+ Neutral tone	first syllable	monkey
jiao3.zi0	Tone 3+ Neutral tone	first syllable	dumpling
ku4.zi0	Tone 4+ Neutral tone	first syllable	pants

Intonation, in general, is characterized by pitch movement of the whole course of utterance. Because Mandarin Chinese uses pitch contour (lexical tones) to contrast word meanings, intonation is sometimes expressed not as F_0 variation on lexical words themselves, but as boundary tones that are added after lexical tones as shown in example (2) (Duanmu, 2006).

Example (2)

Tone		Intonation		
LH	+	L	→	LHL
nan				nan
'difficult'		'affirmation'		'Surely difficult!'
HL	+	H	→	HLH
mai				mai
'sell'		'question'		'Sell?'

On the other hand, intonation also interacts with lexical tones, for example, to express contrast or focus. Lexical tones are modified to implement contrastive focus regarding pragmatic or informative needs. The modification will be described in the following section.

Mandarin Chinese Accent

In a disyllabic word, the syllable with a fully realized lexical tone is more stressed than the one with a neutral tone. When words are connected in a larger domain of compound words, phrases or sentences, the degrees of prominence in the stressed syllables are not equal, which generates sentence accents. Chao (1968) argues that in a prosodic unit (a compound word or a phrase) followed by a pause, the final syllable is primarily accented, the initial syllable is secondarily accented and other syllables are weaker than the initial and the final ones. Tseng (1988) draws the same conclusion that Mandarin Chinese has final accent in both word and phrase levels consisting of full-toned syllables. Duanmu (1999, 2004) further argues that the distribution of sentential accent is based on syntactic structures. Accent is assigned to the complement in a head-complement relation. For example, an object is more likely to be accented than its verb head. Though Chao and Duaman use different approaches to study sentence accent in Mandarin Chinese (one from a phonetic perspective, and the other from syntactic perspective), but since Chinese is a left headed structure, the sentence accent is still placed right most. Recent studies on accent in continuous Chinese speech (Chu et. al, 2003, 2004; Wang et. al, 2003; Bao et. al, 2007) differentiate sentence accent in terms of their semantic and syntactic functions. The normal accent near the sentence boundary, showing syntactic prominence is referred to as rhythmic accent, and the accent carries more semantic meaning, showing semantic prominence is labeled 'semantic accent'. The studies guarantee the existence of sentence accent in the sentence final position, and provide explanation for possible accented syllables in other parts of the sentence if heavy semantic weight is placed. These studies suggest, similar to other Asian tone languages such as Thai (Potisuk et. al, 1996), that Mandarin Chinese accent is an independent system and partially serves an organizational function by being located at syntactic boundaries to link syllables in an utterance into larger prosodic structures and create a series of prosodic units.

Mandarin Chinese Focus

Intonation in Mandarin Chinese is comparatively flat. The function of identifying sentence types (e.g., questions and statements) can partly be identified by sentence final markers, such as 'ma' for interrogation and 'le' for declarative (both of which are assigned neutral tones). Even without the sentence makers, intonation can be realized by adding a tone to a syllable without affecting the original lexical tone assigned to that syllable as shown previously in example (2) above. The existence of sentence makers and 'tones' largely prevents the interaction between intonation and tone systems. However, words in Mandarin Chinese, like any other languages, can be focused in an utterance to signal newness or contrast. What speakers decide to focus is not a matter of syntax or semantics, but a matter of what they are trying to say on a specific occasion in a specific context. In other words, focus is adopted for non-lexical purpose; it depends on the needs of speech mood and discourse expressions (Cao, 2004; Gussenhoven, 2004). The location of focus is complex. It can put emphasis on any part of the utterance, signaling contrast in terms of communicative dynamism, closely related to speaker's attitudes, individual and stylistic variations (Halford, 1994).

Phonetic Representation of Prominence in Mandarin Chinese

The interactions among tone, accent and focus in Mandarin Chinese are bi-directional. Accent tends to affect duration and F_0 of tones (e.g., an unaccented tone usually has narrow F_0 range and relatively short length), while tones also affect the assignment of accent (e.g., a neutral tone doesn't obtain sentence accent, even in the sentence final position) (Pike, 1974; Yip, 1995). Among research in Chinese, there are mainly three phonetic models describing the interactions.

Phonetic Models for Realization of Prominence

Contour model

The contour model (Chao, 1968) claims that Mandarin intonation is characterized by contrasting contour shapes. These contour shapes provide a global rise or fall onto which the local tone contours are superimposed. In Chao (1968)'s proposal, the relation between tone and intonation is explained by a model of small ripples (i.e., tones) riding on large waves (i.e., intonation). The output is an algebraic sum of the two kinds of waves (When the two are both high in F_0, the result will be a plus; when only one is high in F_0, the algebraic addition will be an arithmetical subtraction). This 'algebraic sum' notion is called into question when it is used to explain how tones are realized in different intonation patterns such as questions and statements (Shen, 1985). Based on the model, an arithmetical addition is always assigned to questions and subtraction to statements, because questions are high in pitch and statements are low. However, questions and statements are two different registers (i.e., high for questions and low for statements) regarding intonation. Tones need to be realized within the intonation registers, while retain tonal features. An algebraic sum of contour simply puts questions and statements into the same reference frame. The results show contour changes, but the changes are not controlled or adjusted to fit the contours into two separate intonation registers, or to retain contour distinctions among tones.

Pitch range model

The pitch range model (Garding, 1983; Shih, 1988) claims Mandarin intonation to be a combination of different pitch ranges, and tones to be local pitch perturbations within the given ranges. In Garding (1983)'s proposal, a grid has two parallel lines standing for the top and the bottom lines of an intonation contour. When a word is focused, the grid will expand to create the distance between the top and the bottom. Slightly different from Garding's model, the bottom line in Shih (1988)'s model is claimed to be fixed and only the top line is moveable.

Register model

The register model (Shen, 1990; He & Jin, 1992) argues that Mandarin intonation contours are exhibited on different registers according to the grammar and the speaker's attitude. In Shen (1990)'s study, different intonation patterns are not necessarily on the same pitch level. Intonation contours in Mandarin Chinese can be exhibited on two separate registers: an upper one for questions and a lower one for statements; tones are local F_0 variations on these two separate levels. The model is supported by Cao (2004), who agrees that the relationship between tone and intonation is an 'algebraic sum' of pitch register, instead of that of pitch contour. The intonation pattern is mainly related to the pitch register movement of the utterance, which depends on physiological mechanisms and the needs of semantic expression. For example, the pitch register for a statement has a gradually falling top line and an unchanged base line throughout the whole utterance; a question raises its baseline while lowering its top line. Each tone must be modified by intonation through adjusting its relative register on one hand, and keeping its basic tone shape on the other hand. Meanwhile, intonational elements must be manifested through the F_0 movement of each local tone.

Implications from the Three Phonetic Models

First, the contour model depicts the interactions as an algebraic sum of contours, changes in tone contour rather than register are expected in phonetic representation. A dynamic acoustic parameter that indicates changes could be the slope of F_0. Secondly, the pitch range model suggests changes to both tone register and contour. Hence, an increase in average and maximum F_0 values is expected, as well as changes in slope. Finally, the register model explains the interactions as algebraic sum of pitch register, which implies that the tone contour is not affected by interactions. In other words, the F_0 range remains unchanged while the average F_0 value is raised.

Both the contour and the pitch range models support the idea that tone contour is, to some extent, independent from tone register. It is consistent with the tonal geometry shown in Table 2-1 (b), where tone register and contour are defined as "sisters" rather than "mother-daughter" relationship.

Previous Literature on Phonetic Production of Tone, Accent and Focus and their Interaction in Mandarin Chinese

By exploring the interactions among tone, accent and focus, some researchers investigate tone and prominence in general. Their studies suggest three acoustic parameters implementing prominence. F_0 variation is known to be an important acoustic manifestation of prominence in Mandarin Chinese. Shen (1985) claims that the Chinese tonal ranges could be expanded both upward and downward, but only the expansion of the top-line is relevant to the expression of sentential prominence. Besides F_0 raising, duration and intensity also play important roles in the realization of prominence. Shih (1988) reports that, in addition to the F_0 range expansion, duration and intensity are both involved in stress production: ' it is apparent that prominence is reflected by expanding F_0 range: high targets become much higher, while low target remain at the same level or are slightly lower. Aside from the increased F_0 range, more prominent forms also have longer duration and higher intensity.' Tseng (1988) examines the disyllabic stress pattern in Mandarin and finds that 'the main difference between emphatic and non-emphatic forms appears to be in the domain of syllable duration rather than a wider F_0 range or more energy information'. Jin (1996) investigates the sentence stress in Mandarin Chinese. In this study, four native speakers of Mandarin Chinese are asked to read four simple six-syllable sentences using the intonation that they feel will answer the question posed to them. Acoustic parameters such as F_0, duration and intensity are measured. The results show that when a syllable is stressed, its F_0 range expands dramatically, its duration is lengthened, and its intensity's effect on stress is related to the position of the stressed word in the sentence. At sentence-initial or sentence-medial positions, intensity is not much related to stress. Only at the sentence-final position does he find high correlation between intensity and stress. From these results, Jin (1996) concludes that F_0 and duration play primary roles in sentence stress production and intensity plays a secondary role.

More recently, the examination of tone and prominence is carried out in a more detailed fashion, focusing on each individual tone in Mandarin Chinese separately. Yip (1993) argues that when a tone is prominent, Tone 1 is raised throughout; the end of Tone 2 is higher with the start unchanged; the start of Tone 4 is higher with the end unchanged; Tone 3 is lowered throughout. In Chen (2004)'s study, results show that the four Chinese lexical tones behave quite regularly yet distinctively under prominence. Tone 1 continuously raises its F_0 level; Tone 2 constantly raises its high end with its low start moderately rising only under strong prominence; Tone 3 generally keeps unchanged, with its prominent level indicated by the F_0 level of the following tone; Tone 4 constantly raises its high start, with its low end moderately lowering only under strong prominence. As a summary, both works show that the realization of prominence is more dependent on the raising of the high points of the lexical tones, while different opinions remain in the low targets. However, the studies do not separate prominence due to accent or to focus.

Research conducted by Jin (1996) and Xu (1999, 2004) investigates how lexical tones and focus in Mandarin are realized concurrently in an utterance. Results show that the domain of focus is much wider than that of tone (i.e., tone identities are implemented as local F_0 contours, while focus patterns are implemented as pitch range variations imposed on different regions of an utterance). For instance, figure 3-3 shows a sentence consisting of three words (the first and the last are disyllabic words with H tones, the one in the middle is a monosyllabic word with H tone). Focus is assigned to three different words, one at a time. These three utterances are compared with the same sentence read in a neutral intonation. The focused utterances all show (i) the pitch range of tonal contours directly under focus is substantially expanded; (ii) the pitch range after focus is severely suppressed (which is consistent with Garding et al's (1983) finding of a compression of the pitch range after the focused part); and (iii) the pitch range before focus does not deviate much from the neutral-intonation condition.

Studies on the interaction between focus and accent suggest that there exists a competition between accent and focus if they coincide in the sentence final position (Liao, 1994). The results of recent study by Liu and Xu (2005) are consistent with the previous conclusion drawn by Liao and Tseng. Focus is acoustically manifested much less effectively with the presence of accent than in the sentence middle position. It is also worth noting that the results do not exclude the possibility that the combined manifestation of accent and focus together is more effective than focus alone in the sentence middle position.

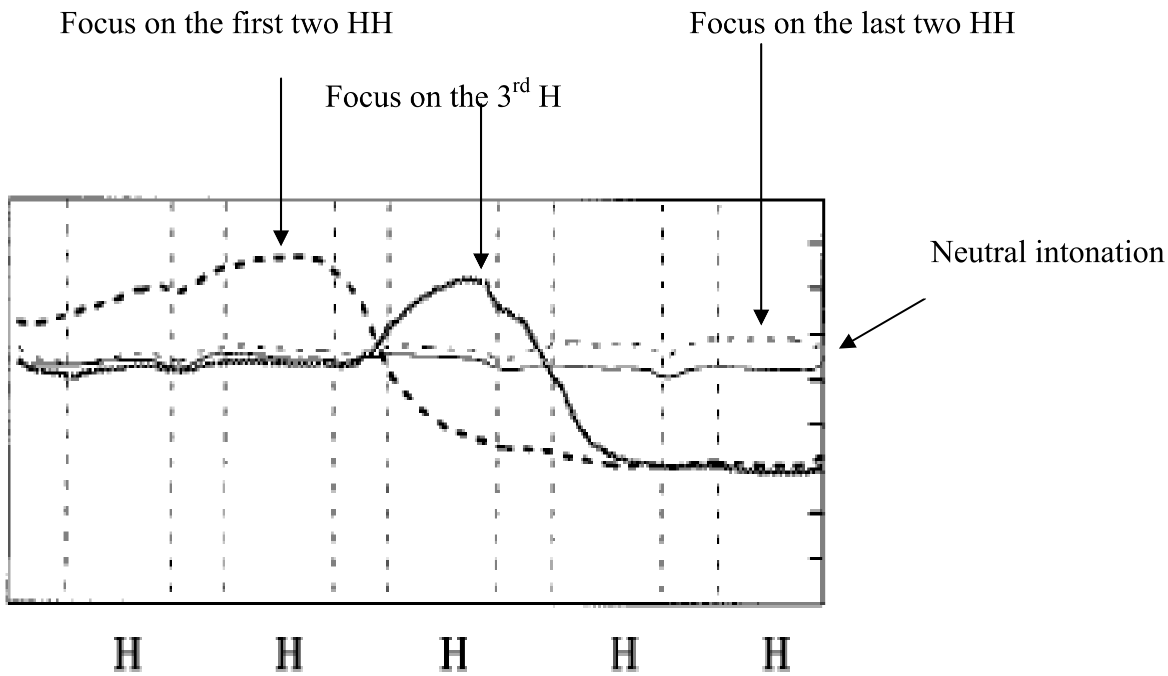


Figure 3-3. Effects of focus on F_0 curves. (The original figure was from Xu, 1999)

Previous Literature on Phonetic Perception of Tone, Accent and Focus and their Interaction in Mandarin Chinese

The perceptual level of prominence analysis concerns listeners' perception of sensory information. The sensory system, different from acoustic analyzer, is subject to psychophysical ranges and limits of sensitivity. Three phonetic parameters are responsible for the coding of prominence: duration, intensity and fundamental frequency (F_0), which are perceived as length, loudness and pitch (Dogil, 1999).Generally, words are more prominent to listeners when they display higher pitch, greater loudness and longer duration than other words in the neighborhood. Among all acoustic cues adopted, duration is a more important cue than intensity and pitch (Shen, 1993). In Shen's study, she examined if pitch was necessary to perceive stress in Mandarin, and if not, which cue, duration or intensity was necessary in stress perception. Four native speakers of Mandarin Chinese were asked to perceive stress in five-syllable sentences. The recorded sentences were manipulated in three ways: in the first set, utterances were low-pass filtered with a cutoff frequency of 400 Hz through a linear phase filter so that the segmental information was removed to prevent listeners from using semantics in their judgment of stress. In the second set, F_0 was held constant at 135 Hz in the filtered utterances. In the third set, the intensity was fixed at a constant 60 dB, in addition to the elimination of F_0 variation. The temporal patterns of the stimuli in all three sets remained intact. Thus, subjects had only duration information available in Set 3, duration and intensity in Set 2, and duration, intensity, and F_0 in Set 1. It was postulated that (1) if there were no significant differences among subjects' responses to stimuli of Sets 1 and 2, then F_0 was not crucial to the perception of stress, and (2) if subjects responded similarly to stimuli of Sets 2 and 3, then intensity likewise was not important in cuing stress. The results revealed that more differences were observed between Set 2 and 3, which indicated, listeners were more likely to notice the intensity difference between stressed and unstressed vowels of the same quality (the difference was near 8 dB), but neither the presence of F_0 nor the variation of intensity changed the judgment of stress significantly. From the results, Shen concluded that duration was the most important cue that listeners used in perceiving stress, that intensity cue was also adopted and that the pitch cue was not necessary.

Comparing studies of prominence production and perception, mismatches are observed in terms of the acoustic dimensions involved. Besides studies mentioned above, Waterson's (1976) work in the acquisition of phonology shows that, in the early stage of 17-19 months, high intensity and long duration are important cues for the child's phonological discrimination. The child does not become aware of all acoustic cues simultaneously. By comparing their perception with production, he suggests the child forms perception patterns (i.e., cues that he pays attention to) based on his initial discriminations. Later, he is able to pay more attention to the other cues because use of original patterns becomes almost automatic. His refined perception results in a mismatch between what he perceives and the actual acoustic signal. Moreover, mismatch also occurs between his own perception and production, because he continually refines his own production based on refined auditory discrimination.

Yuan's (2005) study on the production and perception of intonation in Mandarin also implies possible acoustic and perceptual mismatches for native speakers. He finds different speakers choose different strategies to modify lexical tones in intonation and adopt different cues in perception. The differences could be due to mismatches exist between production and perception or subject-dependent difference. If different speakers have different intonation phonology, each speaker should identify intonation of his/her own speech better than intonation of others'. The results show it is not the case that a speaker identifies intonation of his/her own speech better than intonation of others', which does not support the hypothesis that different speakers have different intonational phonology, but provide an evidence for mismatches in production and perception in general. .

Gaps in Previous Literature

Besides the great achievement in previous literature, there are some shortcomings that could still be improved. Firstly, some of previous studies investigate the interactions among tone, accent and focus in relatively short utterances (e.g., a word, a phrase, a simple sentence)[9]. These settings are not perfectly natural for sentence accent or focus to take place, because accent systems are best illuminated by an examination in a more complicated organization of larger utterances (Beckman, 1986).

[9] Duanmu (1999) and Tseng (1988) studied stress within disyllabic words; Shi (2004) studied narrow focus on the third word of a simple sentence; Surendran et.al (2005) studied focus and tone recognition in Mandarin in 3-word phrases.

Secondly, most studies have analyzed the phonetic realizations of prominence in a descriptive fashion without the support of quantitative analyses.

Thirdly, interactions are examined mostly in a general way and only between two phonological categories (e.g., between tone and focus; accent and focus), and the interactions among the three have attracted little attention. Since the data collected are different among experiments, it's difficult to compare and contrast the results thus far obtained from existing research.

Fourthly, most phonetic studies on prominence have examined the realization of prominence (i.e., how tone, accent and focus are produced), and less attention has been given to the perception of prominence. The human perceptual system is separated from the production system, and is subject to psychological ranges and limits of sensitivity, so the question remains as to whether the same information used in prominence production will be the same as that used in its perception.

Objectives of Current Study

The goal for current study is to investigate prominence caused by accent and focus respectively in the environment of longer utterances (e.g., sentence groups) to quantitatively examine the interactions among tone, accent and focus in Mandarin Chinese. It can be reified to four specific aims.

- Enlarge the study domain to sentence groups.
- Analyze the data in a quantitative way.
- Investigate the production and perception of accent and focus respectively with the same set of data, comparing and contrasting among the four tones.
- Study perception as well as realization of prominence to examine if acoustic parameters adopted for realizations could be perceived in a similar fashion.

Research Questions

There are three research questions for the study.

Research question 1: What are the acoustic parameters used to realize focus and accent among lexical tones of Mandarin Chinese?

Research question 2: What are the interactions among tone, accent and focus in the realization of focus and accent?

Research question 3: Among acoustic parameters used to produce focus and accent, which ones are used in the perception of prominence?

The hypotheses being testes: (1) Four lexical tones use different acoustic parameters to realize focus and accent, but focus and accent are realized in a similar fashion for a particular tone; (2) There exist interactions among tone, accent and focus; (3) Perceptual rankings of acoustic cues are different from those found in production, because perception and production are different systems, and different constraints may be applied to this two domains.

CHAPTER 4
ACOUSTIC PARAMETERS FOR FOCUS AND ACCENT REALIZATION

The goal of the production experiment was to investigate the acoustic parameters used for focus and accent realization among lexical tones, and the interaction among tone, accent and focus in these acoustic dimensions to signal prominence. The research questions addressed were 1: What are the acoustic parameters used to realize focus and accent among lexical tones of Mandarin Chinese? and 2: What are the interactions among tone, accent and focus in the realization of focus and accent?

Native speakers of Mandarin Chinese were recorded producing utterances where the target words were set in prominent and non-prominent conditions. Acoustic parameters were measured for the target words, and compared between conditions in terms of how often they were adopted to implement prominence (i.e., the percentage of data showing modifications in a particular acoustic dimension) and the extent of the modification (i.e., the ratio between non- prominent and prominent conditions).

Chapter Four will focus on the acoustic parameters used for focus and accent realization among lexical tones (RQ1), and Chapter Five will discuss the interactions among tone, focus and accent (RQ2). The chapter will be organized as follows. First, the design of the production experiment will be described and justified. Measurement procedures implemented to normalize across-talker differences, and the coding of prominence realizations among the talkers will also be described. Next, the results will be presented and analyzed for each acoustic dimension examined. Acoustic parameters used to implement focus and accent will be discussed among lexical tones in this section.

Methods

Subjects

Ten native speakers of Mandarin Chinese (six female and four male), ages between 27 and 32, participated in this experiment. They were born in Beijing, the capital city of the People's Republic of China, and grew up there. They had lived in the US for less than two years at the time of testing, studying in various Ph. D programs at the University of Florida, including Engineering, Liberal Arts, Education, Public Health and Pharmacy. All reported normal language and speech development and passed a bilateral hearing screen in the range of 250 to 8,000 Hz measuring at 25 dB HL (by DSP Pure Tone Audiometer).

Materials

The stimuli used in this experiment are disyllabic, real words 'LiZhi' [li.tʂi] produced with all possible combination of the four tones of Mandarin Chinese, yielding 16 tonal combinations in all, including same tone (i.e., 1-1, 2-2, 3-3 and 4-4) combinations. All target disyllabic words are embedded in four sentence frames to generate 64 utterances (16 target words x 4 sentence frames =64) where target words are placed under four conditions (Shown in Table 4-1). In Condition (a) and (b), the target disyllables/tones appear in the unfocused position of the utterances. Specifically, target words in (a) are placed in the unaccented, sentence medial position [-A-F], while target words in (b) are in the sentence final position with the default sentence final accent [+A-F] (see Chapter 3 Chinese accent section). Target words in Condition (c) and (d) are the corresponding focused counterparts of (a) and (b) respectively. In other words, target words in (c) are unaccented, but focused [-A+F], and target words in (d) are both accented and focused [+A+F].

Table 4-1. Target words under four conditions.

Conditions	Descriptions	Labels
(a)	Target words are unaccented and unfocused in the sentence middle position	[-A-F]
(b)	Target words are accented but unfocused in the sentence final position.	[+A-F]
(c)	Target words are unaccented but focused in the sentence middle position	[-A+F]
(d)	Target words are accented and focused in the sentence final position	[+A+F]

An example of the target Tone 4 and Tone 2 combination is listed in Table 4-2. A disyllabic word is selected as the target, because disyllabic words are the most common word forms in Mandarin Chinese as far as prosody are concerned (Chu et. al, 2004). Moreover, the word is designed to be a person's name to represent a single morpheme without any internal relationship between the two syllables[10]. The target has a consistent high vowel [i] in CV structures to minimize the influence of vowel quality on acoustic realizations of prominence.

To further control for the environment of the targets, sentence frames under four conditions where the same target word is inserted are similar, except for minor differences to make the context natural. For example, the syllable immediately after the target word in Condition (a) and (c) is the possessive marker '的' [tə] with neutral tone to minimize the tonal effects after the target word, and to warrant the comparison between the target words in these two conditions and the ones in the sentence final positions under Conditions (b) and (d). In addition, the syllable immediately before the target word is the same verb '提' [t^hi] with Tone 2 across conditions. Note that the verb is always preceded by '不' [pu] with Tone 4 expressing negation to form

[10] The internal construction of a compound word affects the prosodic distribution among syllables involved. (Bao et. al, 2007).

disyllabic words for prosodic purposes (Disyllabic words are the most common word forms in Mandarin Chinese).

Table 4-2. Example of target Tone 4 and Tone 2 under four conditions*

Conditions	Examples	Literal Meanings	Labels
(a)	信息传播已经进入无限量的时代。博客的出现让炒作成了大众能够尝试的玩意儿。作者在文中故意不提**丽植**的名字。问其原因，他说不想有炒作之嫌。	Here comes the information age. The emergence of blogs makes it possible for ordinary people to publicize other people's privacy. The author didn't mention **Lizhi**'s name. Being asked why, he answered he still kept his professional ethics.	[-A-F]
(b)	信息传播已经进入无限量的时代。博客的出现让炒作成了大众能够尝试的玩意儿。作者在文中故意不提**丽植**。问其原因，他说不想有炒作之嫌。	Here comes the information age. The emergence of blogs makes it possible for ordinary people to publicize other people's privacy. The author didn't mention **Lizhi**. Being asked why, he answered he still kept his professional ethics.	[+A-F]
(c)	信息传播已经进入无限量的时代。博客的出现让炒作成了大众能够尝试的玩意儿。作者在文中提到了一植和二植，就是故意不提**丽植**的名字。问其原因，他说不想有炒作之嫌。	Here comes the information age. The emergence of blogs makes it possible for ordinary people to publicize other people's privacy. The author mentioned Yizhi and Erzhi in his article, but didn't mention **Lizhi**'s name. Being asked why, he answered he still kept his professional ethics.	[-A+F]
(d)	信息传播已经进入无限量的时代。博客的出现让炒作成了大众能够尝试的玩意儿。作者在文中提到了一植和二植，就是故意不提**丽植**。问其原因，他说不想有炒作之嫌。	Here comes the information age. The emergence of blogs makes it possible for ordinary people to publicize other people's privacy. The author mentioned Yizhi and Erzhi in his article, but didn't mention **Lizhi**. Being asked why, he answered he still kept his professional ethics.	[+A+F]

* 'lizhi'is the target word, bolded and underlined under each condition.

Procedures

The production experiment was carried out in a sound booth in the phonetics lab at the Program in Linguistics, University of Florida. In order to ensure a consistent level of recording volume, all readings were recorded at a fixed 4-inch distance and a 15-30º angle between the head-mounted microphone (Shure SM 10A) and the participants' lips so that the input level can be made relatively stable. Care was also taken to set a sampling rate of 44.1 kHz and 16-bit PCM[11] on the Marantz PMD660 Professional Solid State Recorder and saved for all speakers. The 64 sentences were presented to participants in a random order for recording. They were recorded reading the sentences in a fluent and natural fashion after they were familiar with the context and had practiced reading to themselves once or twice. The resulting 640 utterances (64 utterances x 10 speakers) were transferred to a PC and saved as WAV files for subsequent acoustic measurements.

Acoustic Measurements

Acoustic measurements of target words were taken from the vowel portions only. Using both waveforms and spectrograms, vowels were segmented in Praat (Boersma & Weenink, 2004). F2 onset and offset were taken to be the onset and offset of the vowel respectively (shown as the first vowel segmentation in Figure 4-1). When the onset or the offset of F2 was hard to identify, the onset of periodicity and the point at which the amplitude is minimum were used to define vowel onset and offset respectively (shown as the second vowel segmentation in Figure 4-1).

---

[11] PMC stands for pulse code modulation. In the context of audio coding PCM encodes an audio waveform in the time domain as a series of amplitudes. This parameter specifies the amount of data used to represent each discrete amplitude sample. 16 bits gives a range of 65536 amplitude steps.

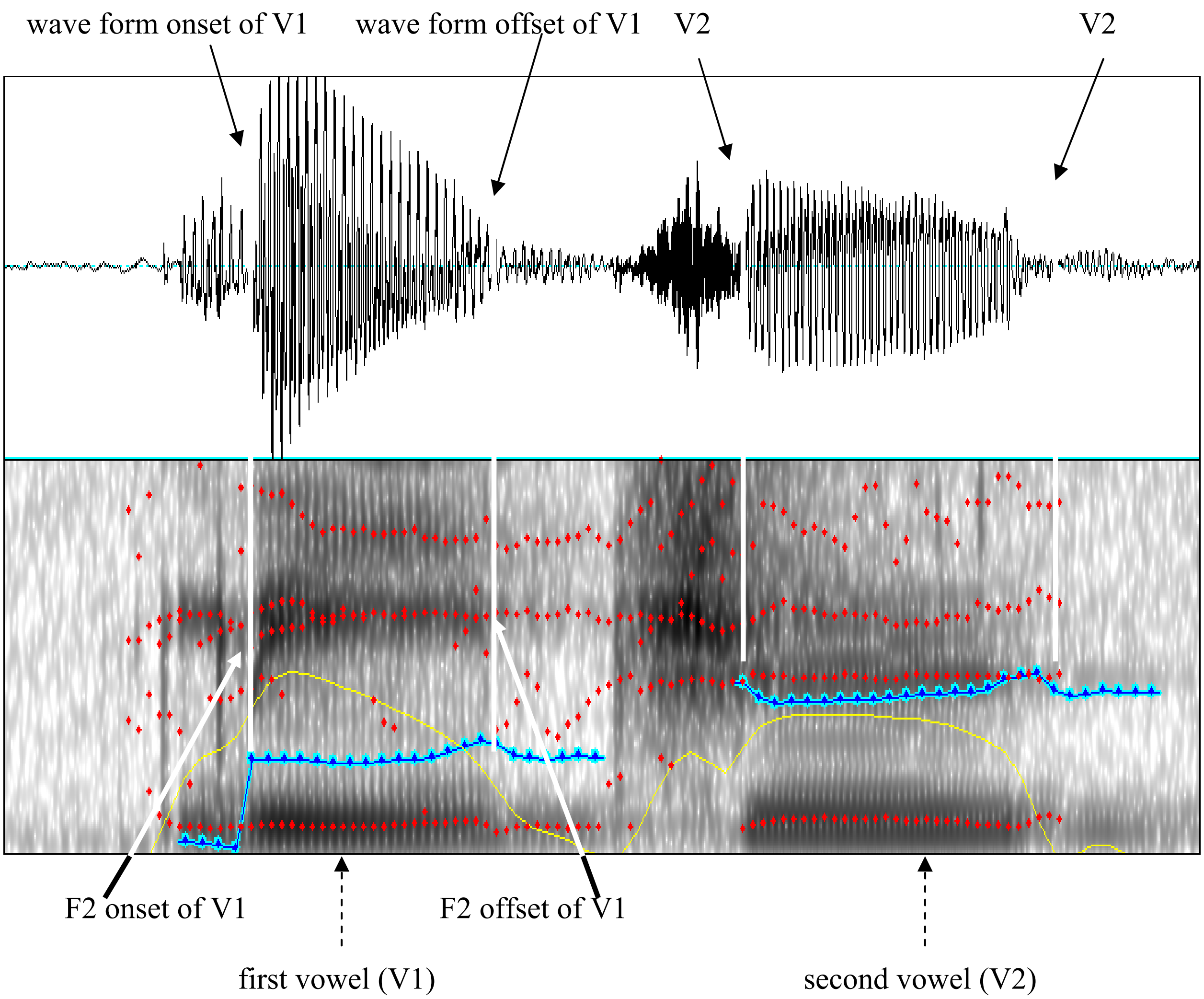


Figure 4-1. Vowel segmentation

Altogether seven acoustic parameters were measured for each of the target tone. These include vowel duration; average and maximum values of intensity; average, maximum and minimum values of F_0; and F_0 slope. The same measurements were also taken for the whole sentence for the purpose of across-talker normalization. Table 4-3 illustrates acoustic parameters measured for each target words. For Tone 1, maximum, minimum F_0 and F_0 slope (alpha F_0) were not measured, because it was a level tone and the numerical F_0 differences in the contour were not considered as tonal features. F_0 slope (alpha F_0) was not measured for the dipping Tone 3 either, because its contour in natural speech was often partially realized.

Table 4-3. Acoustic parameters measured for four lexical tones*

Tone	T1	T2	T3	T4
Duration	√	√	√	√
Mean of intensity	√	√	√	√
Maximum of intensity	√	√	√	√
Mean F_0	√	√	√	√
Maximum F_0	-	√	√	√
Minimum F_0	-	√	√	√
F_0 Slope	-	√	-	√

* "√" indicates acoustic parameters measured.

Acoustic Normalization among Speakers

Controls on the variation of speakers' speaking rates and vocal F_0 ranges were taken by normalization acoustic measurements obtained across speakers. For the acoustic parameters described above, ratio values between targets and sentence contexts were derived.

Duration was normalized by adjusting the duration value of target tone with the speaker's speaking rate. The formula used for this ratio was:

$$(4.1)\ \text{Normalized Duration} = \frac{\text{Msec (Target Tone)}}{\text{Speaker's Speaking Rate}}$$

where Msec (Target Tone) is the measured duration of the vowel of target tone. Speaker's Speaking Rate is defined as the average duration of a syllable in the sentence shown in equation (4.2)

$$(4.2)\ \text{Speaker's Speaking Rate} = \frac{\text{Msec (Sentence)}}{\text{Num. of Syllables}}$$

The comparable logarithmic ratios for the amplitude measurements could be computed by subtracting the sentence average from the value of the target, because the intensity measurements were already in decibels (shown in formula 4.3).

(4.3) Normalized Intensity = dB (Target Tone) - dB (Sentence Average)

F_0 normalization was performed using four different frequency scales: hertz (Hz), semitone, ERB-rate and Mel scale[12]. No significant difference was found among these scales. Since most researchers used the Mel scale for segment measurements (consonants and vowels), this scale was selected in this study. Mel scale is a logarithmic frequency scale defined as in formula (4.4), where f is the fundamental frequency. Comparable ratios were calculated by subtracting the average Mel value of the sentence from the target as shown in equation (4.5).

(4.4) Mel scale = $1127.01048 \log_e (1+f/700)$

(4.5) Normalized F_0 = Mel scale (Target Tone) - Mel scale (Sentence Average)

Coding of Prominence Realizations

Realizations of prominence were coded by comparing target words in prominent conditions with those in non-prominent conditions. Among the four conditions (i.e., [-A-F], [+A-F], [-A+F], and [+A+F]) (shown in Table 4-1), target words were compared in Figure 4-2.

[12] Hertz (Hz) is a linear frequency scale. It is defined as the number of cycles per second (Ladefoged, 1996). Semintone is a musical scale used to express the relative distance between two tones in a musical interval. Equivalent rectangular bandwidth rate (ERB-rate) is a psychoacoustic scale. It represents the perceived excursion size of prominence-lending pitch movements in different pitch registers (Hermes and van Gestel, 1991). The Mel scale is a perceptual scale of pitches judged by listeners to be equal in distance from one another (Stevens, Volkman and Newman, 1937).

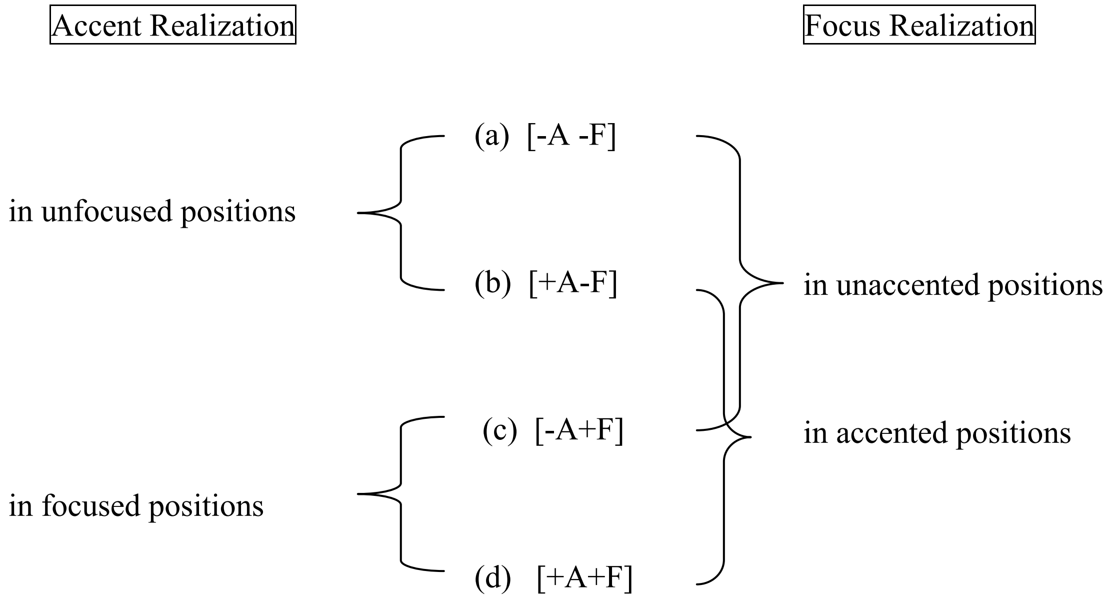


Figure 4-2. Realizations of prominence

Comparisons were made between (a) and (b) for accent realization in unfocused positions, (c) and (d) for accent realization in focused positions, (a) and (c) for focus realization in unaccented positions, (b) and (d) for focus realization in accented positions. All conditions (e.g., [-A+F]) have two variables: A(ccent) and F(ocus). The two conditions within each comparison shared a value in one variable, and differed in the other (i.e., [-A-F] versus[-A+F] or [-A-F]. versus[+A-F]).The variable acquiring the same value in the two conditions was the environment for prominence realization, the other variable with differing value in the two conditions indicated the source of prominence realization. For example, for the [-A-F] versus [-A+F] comparison, the environment of prominence comparison is 'unaccented' and the source of 'prominence' was whether the target words were 'focused' [+F] or 'unfocused' [-F]. In other words, in this comparison, acoustic realizations of the target 'focused' versus 'unfocused' (both being produced in an 'unaccented' position in the utterance) were being compared.

For all acoustic parameters studied here, comparisons were operationalized as ratios of normalized acoustic values obtained from the prominent conditions to those obtained from the non-prominent conditions. For example, the ratio for duration was defined as the normalized duration value in Condition P (prominent) divided by the one in Condition NP (non-prominent). The formula used for this ratio was shown in (4.6):

$$\text{(4.6) Duration Ratio} = \frac{\text{Normalized Duration in Condition P}}{\text{Normalized Duration in Condition NP}}$$

Where if the ratio value is larger than 1, duration in Condition P is longer than in Condition NP, indicating a numerical lengthening in the prominent condition.

The comparable ratios for the normalized intensity were computed by a subtraction between normalized logarithmic intensity values in two difference conditions. The same was true for comparisons of normalized F_0 between conditions in Mel scale, as shown in formulae (4.7) and (4.8).

(4.7) Intensity Ratio

= dB (Normalized Intensity in Condition P) - dB (Normalized Intensity in Condition NP)

(4.8) F_0 Ratio

= Mel scale (Normalized F_0 in Condition P) - Mel scale (Normalized F_0 in Condition NP)

Where ratio values > 0 indicates an increase from non-prominent condition to prominent condition.

Statistical Analyses

Along with reporting the descriptive statistics for the acoustic measures mentioned above, the processed data were compared for significant differences using appropriate Analysis of Variance (ANOVA) methods: repeated measures and follow-up pair wise comparisons with Bonferroni adjustment. The significant level was set as $\alpha = .05$.

Results and Analyses

The production experiment investigated the two research questions proposed. In this chapter, I will discuss the first question regarding the acoustic parameters used for focus and accent realization among lexical tones. The interaction among tone, accent and focus in realization will be analyzed in the next chapter (chapter Five).

Research Question 1: What are the Acoustic Parameters Used to Realize Focus and Accent among Lexical Tones of Mandarin Chinese?

Prominence could be conceivably implemented by either a decrease or an increase in acoustic values, but prominence realization in Mandarin Chinese was in general implemented by F_0 rising and expansion, duration lengthening and intensity increasing as mentioned in Chapter 3. Therefore, to answer Research Question 1, I took an increase, as opposed to a decrease in acoustic values as an indication of prominence (i.e., focus and accent). In other words, I assumed that the focus and accent were implemented by an increase rather than a decrease in the value of all acoustic dimensions measured. The increase was indicated by a greater than 1 (>1) duration ratio in formula (4.6), greater than 0 (>0) intensity and F_0 ratios in (4.7) and (4.8). It was found that not all acoustic parameters measured were simultaneously realized to implement prominence and some parameters were presented more frequently than others. For example, in some cases, 'focus', was realized by lengthening of the duration, raising the mean value of F_0 , and increasing the maximum value of intensity; while in other cases, the intensity was not used, and 'focus' was realized only by means of an increase in duration and F_0. Therefore, the frequency at which each acoustic parameter was used to implement 'prominence' (focus or accent) differed across the data set. In other words, the percentage of data in the prominent conditions that actually showed an increase in a particular acoustic dimension to implement prominence varied. An example of how the percentage was calculated was shown in Figure 4-3.

In Figure 4-3. , the duration data of Tone 1 were first normalized under each condition in the upper table. They were then compared between conditions to obtain the duration ratio: the comparison between [-A+F] and [-A-F] signaled focus realization alone without the effect of accent; the comparison between [+A-F] and [-A-F] signaled accent realization without the effect of focus. From the duration ratios displayed, not all ratios showed an increase (ratio value >1). Results confirmed that 80% of the focused data and 90% of the accented data indicated an increase in duration. Since an increase in duration was considered as evidence for prominence in Mandarin, I concluded that 80% of Tone1 used the duration parameter to implement focus and 90% use the duration parameter to implement accent.

For acoustic parameters measured in Table 4-3[13], the percentage was calculated for both focus and accent realizations (24 measurements * 2 prominent realizations= 48 calculations in total). The results showed the percentage of data making use of a particular acoustic parameter to implement prominence. In other words, the results indicated the proportion of prominent data in which a particular acoustic parameter was used in its realization: the higher the presence (or percentage), the more frequently a parameter was used to implement prominence. The total 48 percentage values were listed and grouped in Figure 4-4.

In Figure 4-4., there were 12 instances where acoustic parameters were used in the realization of more than 76% of the prominent data. Among them, 'focus' had more parameters listed in this range than 'accent', which implied that 'focus' was implemented by a greater variety of acoustic parameters than 'accent'. Tones were observed using more than one parameter to implement prominence. For instance, both duration and mean F_0 were frequently adopted by Tone 1 to implement focus; moreover, four parameters were used by Tone 4 to

[13] In Table4-3, four acoustic parameters were measured for Tone 1, seven for Tone 2 and Tone4 respectively, and six for Tone 3, which were added up to twenty-four measurements.

realize focus. No intensity parameters were listed in this range. Ranked lower, 7 cases were listed between 61%-75%, most of which made use of intensity parameters to implement focus. There were no acoustic parameters that were used to implement prominence in 45% -60% of the data. This gap segregated parameters used in less than 45% of the data from those appeared in 60% or more.

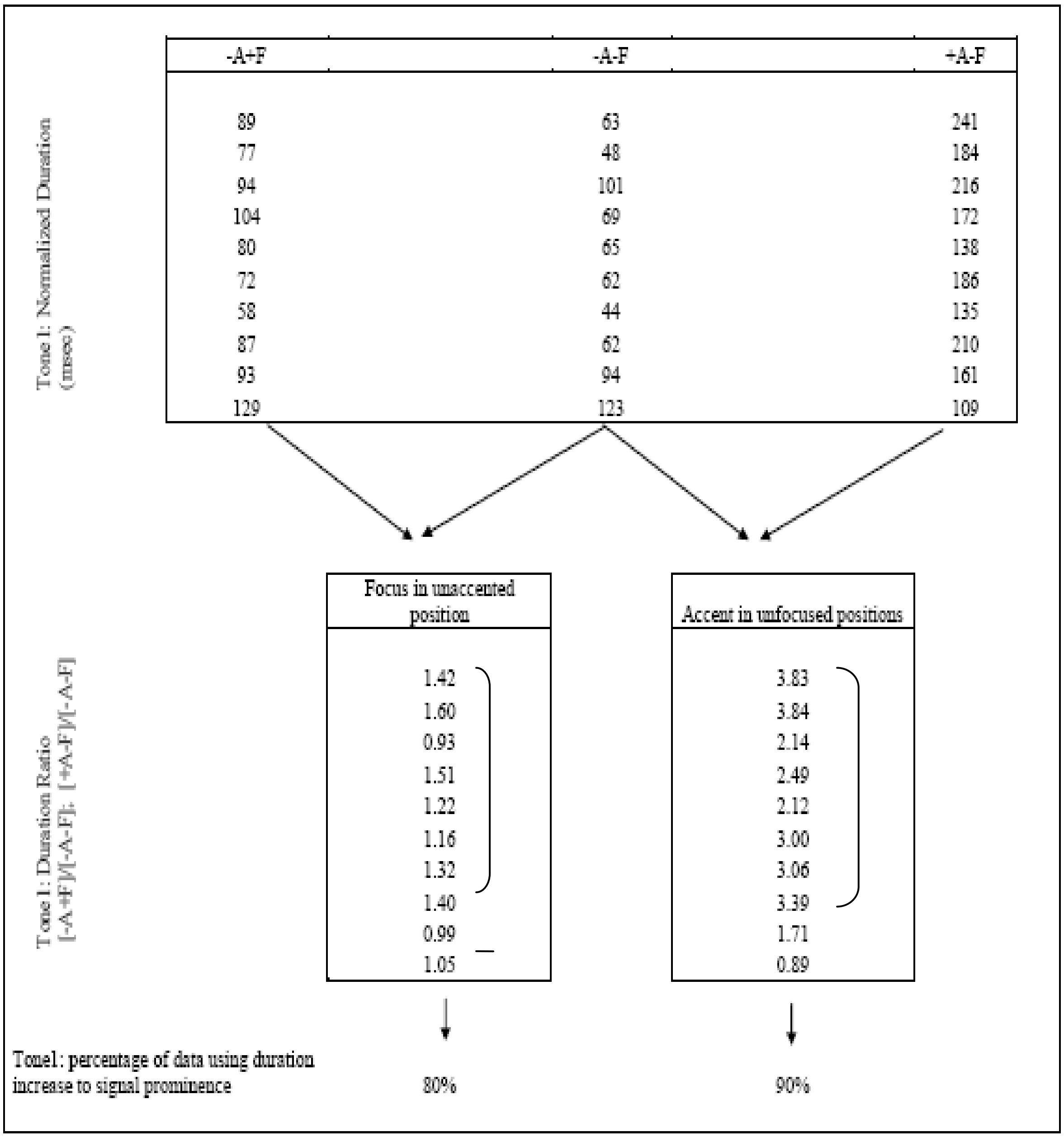


Figure 4-3. Calculation of duration increase to implement prominence in Tone 1

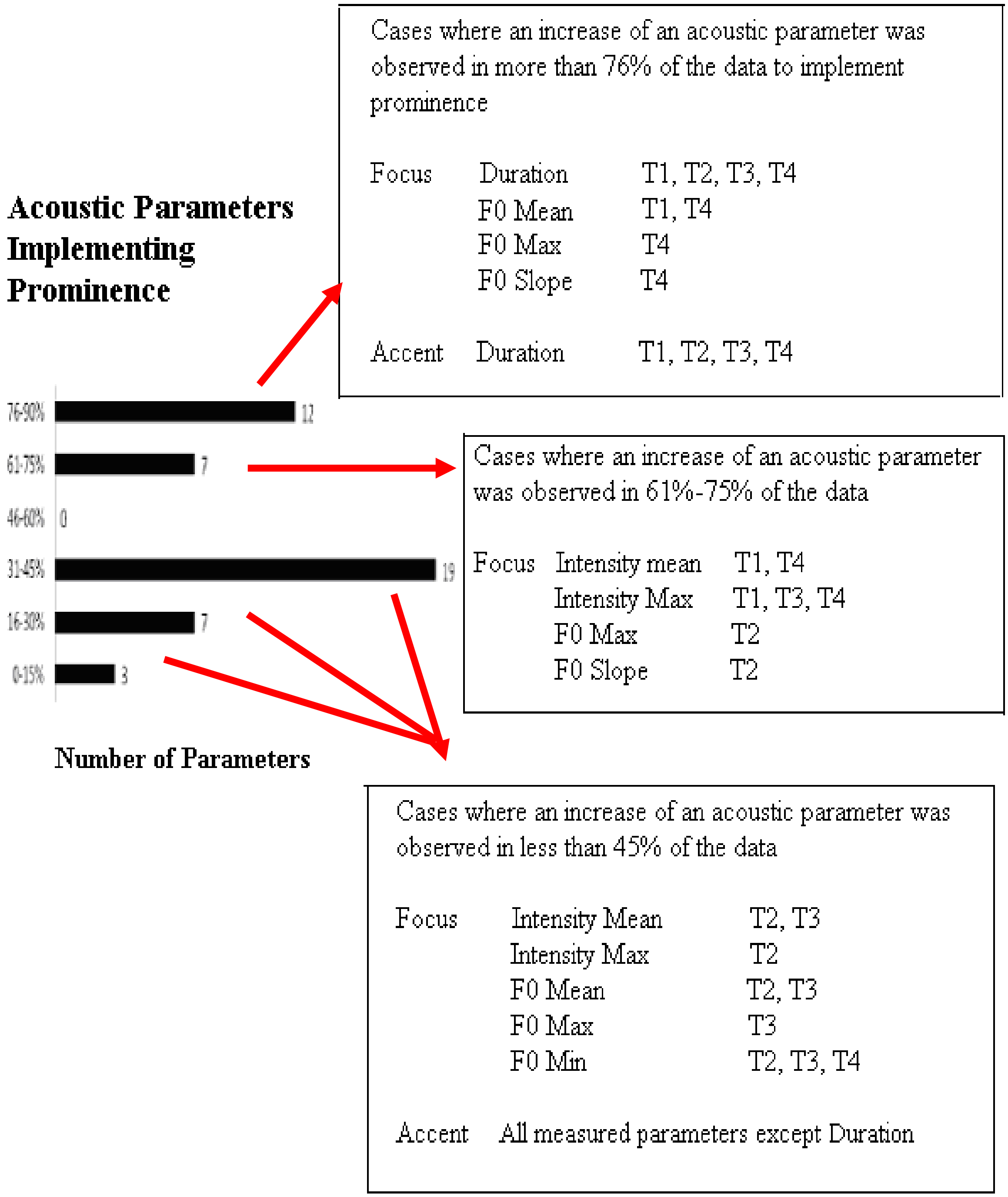


Figure 4-4. Distribution of acoustic parameters in terms of their frequencies

Table 4-4 showed the percentage of data in which each acoustic parameter was used in the implementation of focus and Table 4-5 showed the percentage of data in which each acoustic parameter was used in the implementation of accent. (The shaded cells in the two tables were acoustic parameters not measured, and the bolded cells were acoustic parameters used in more than 60% of the data to implement focus).

Table 4-4. Acoustic parameters for focus realization in unaccented positions

Tones	DUR	INTEN-MEAN	INTEN-MAX	F_0-MEAN	F_0-MAX	F_0-MIN	SLOPE
T1	**82%**	**63%**	**62%**	**80%**			
T2	**84%**	39%	44%	21%	**66%**	36%	**74%**
T3	**81%**	38%	**64%**	36%	34%	30%	
T4	**86%**	**65%**	**70%**	**85%**	**88%**	43%	**90%**

Table 4-5. Acoustic parameters for accent realization in unfocused positions

Tones	DUR	INTEN-MEAN	INTEN-MAX	F_0-MEAN	F_0-MAX	F_0-MIN	SLOPE
T1	**81%**	37%	45%	12%			
T2	**81%**	31%	39%	21%	28%	15%	35%
T3	**83%**	22%	39%	39%	39%	26%	
T4	**85%**	34%	41%	41%	41%	7%	41%

Acoustic parameters for focus realization

In Table 4-4, seven acoustic parameters were measured to examine focus implementation. Acoustic parameters differed in their frequencies to implement focus among tones.

Tone 1 Four acoustic dimensions were used in the implementation of 'focus' in unaccented positions for Tone 1: lengthening the duration, increasing the mean and the maximum values of intensity, and raising the mean value of F_0 (shown in Table 4-6).

Table 4-6. Descriptive analysis of parameters used for focus realization in Tone 1

Parameters	Mean (%)	Std. Deviation
Duration (Dur)	82.14	17.09
Mean Intensity (Inten-mean)	62.57	18.62
Max Intensity (Inten-max)	61.82	20.23
MeanF_0 (F_0-mean)	80.08	8.28

Numerically, the duration was the most frequent dimension adopted (present in 82.14% of the data), followed by F_0 (present in 80.08% of the data). The two intensity measures (mean and maximum) were ranked lower in terms of the percentage of data they applied to (62.57% and 61.82%).

These observations were submitted to a repeated-measure ANOVA with acoustic parameter as the within-subject factor. The results suggested that with an alpha level of .05, the differences of mean percentage were statistically significant among the acoustic parameters measured [$F(3, 27) = 5.876$, $P = .003$]. Follow-up pair-wise comparisons with Bonferroni adjustment were conducted. The results (shown in Figure 4-5) revealed that the difference in terms of their frequencies in focus realization among the four acoustic parameters was not significant [p= 1.00 between duration and mean F_0, as well as between mean and maximum intensity; p=.107 and.261 between duration and the intensity parameters (i.e., maximum and mean intensity); p=.155 and .185 between mean F_0 and the intensity parameters].

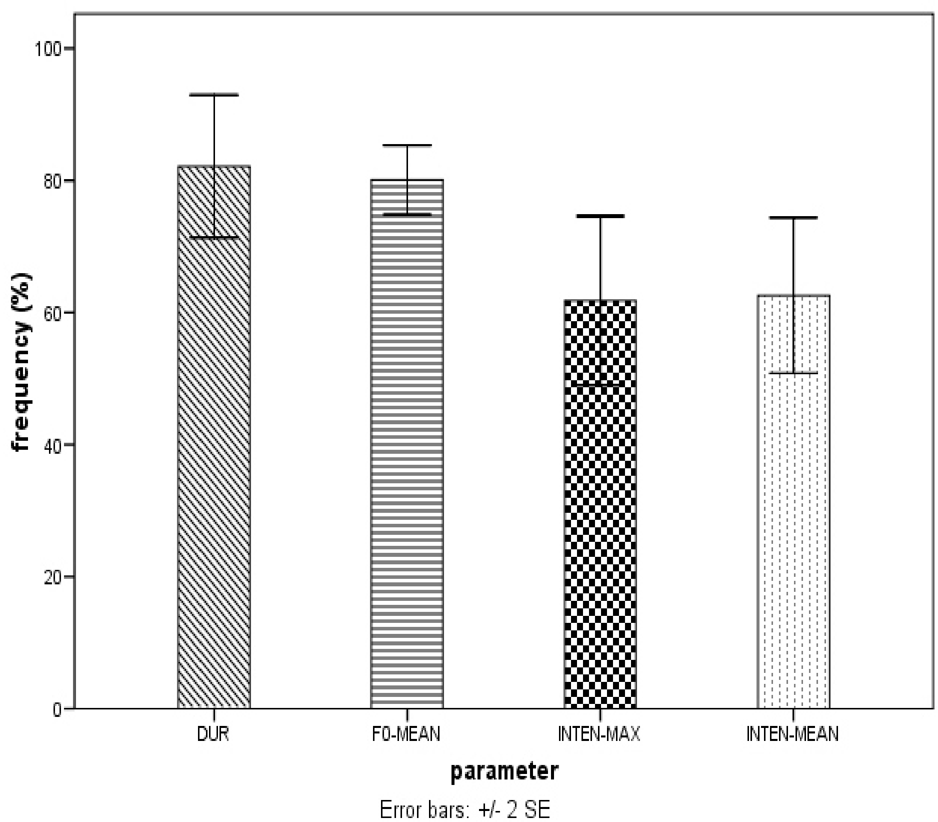


Figure 4-5. Acoustic parameters (and their frequencies) used in ‘focus’ realization of Tone 1.

Tone 2 Seven acoustic parameters were measured to examine focus realization in unaccented positions for Tone 2: duration, mean and maximum intensity, mean, maximum and minimum F_0, and the slope of F_0 from the onset to the offset (shown in Table 4-7 and Figure 4-6).

Table 4-7. Descriptive analysis of parameters used for focus realization in Tone 2

Parameters	Mean (%)	Std. Deviation
Duration (Dur)	83.75	16.72
Mean Intensity (Inten-mean)	39.11	8.66
Max Intensity (Inten-max)	43.75	10.62
Mean F_0 (F_0-mean)	21.43	15.99
Max F_0 (F_0-max)	66.06	23.59
Min F_0 (F_0-min)	35.71	12.68
F_0 Slope (Slope)	73.55	16.95

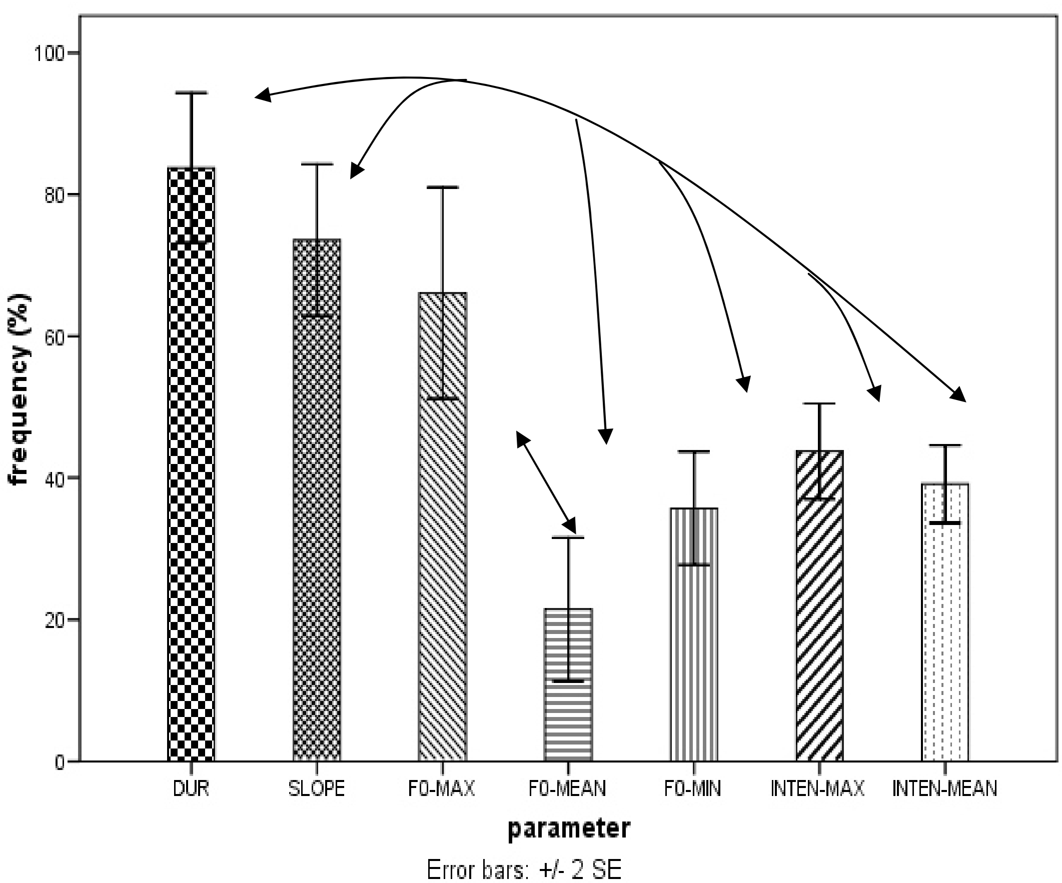


Figure 4-6. Acoustic parameters (and their frequencies) used in 'focus' realization of Tone 2. Arrows indicate significant difference in the frequency at which the two parameters were used

Duration Lengthening was used in83.75% of the data, numerically more than two F_0 measures (i.e., F_0 slope and F_0 maximum) which were used in 73.55% and 66.06% of the data. Other acoustic parameters, such as mean intensity, maximum intensity and minimum F_0 , appeared less frequently in 39.11%, 43.75% and 35.71% of the data respectively. MeanF_0 was used least frequently in 21.43% of the data to realize focus.

The repeated-measure analysis showed that, with an alpha level of .05, the frequency at which these acoustic parameters was used in 'focus' realization in Tone 2 was statistically significant different from each other [$F(6, 54) =23.63$, $P =.000$]. Follow-up pair-wise comparisons with Bonferroni adjustment suggested that duration and F_0 slope were used significantly more frequently than other acoustic parameters, such as mean intensity [$p=.001$between duration and mean intensity, and $p=.006$ between F_0 slope and mean intensity], maximum intensity [$p=.004$ and .014], mean F_0 [$p= .000$ and .000] and minimum F_0 [$p=.000$ and .001] for focus realization. Maximum F_0 was also used more frequently than mean F_0 [$p=.013$] to produce focus. There was no significant difference among duration, F_0 maximum and F_0 slope [$p=1.000$ between duration and F_0 maximum, duration and F_0 slope, and between F_0 maximum and F_0 slope]. Neither was there difference among intensity parameters (mean and maximum intensity), minimum F_0, and mean F_0 [$p=.086$~1.000] (Figure 4-6).

Tone 3 Six acoustic dimensions were measured to examine focus realization for Tone 3 (shown in Table 4-8 and Figure 4-7). The duration dimension was used in 81.25% of the data, followed by maximum intensity which was used in 63.57%of the data. Mean intensity and F_0 parameters (i.e., mean, maximum and minimum F_0) were used less frequently in 29.94% to 37.50% of the data.

Table 4-8. Descriptive analysis of parameters used for focus realization in Tone 3

Parameters	Mean (%)	Std. Deviation
Duration (Dur)	81.25	14.73
Mean Intensity (Inten-mean)	37.50	11.79
Max Intensity (Inten-max)	63.57	18.81
Mean F_0 (F_0-mean)	35.72	13.47
Max F_0 (F_0-max)	33.69	5.42
Min F_0 (F_0-min)	29.94	11.91

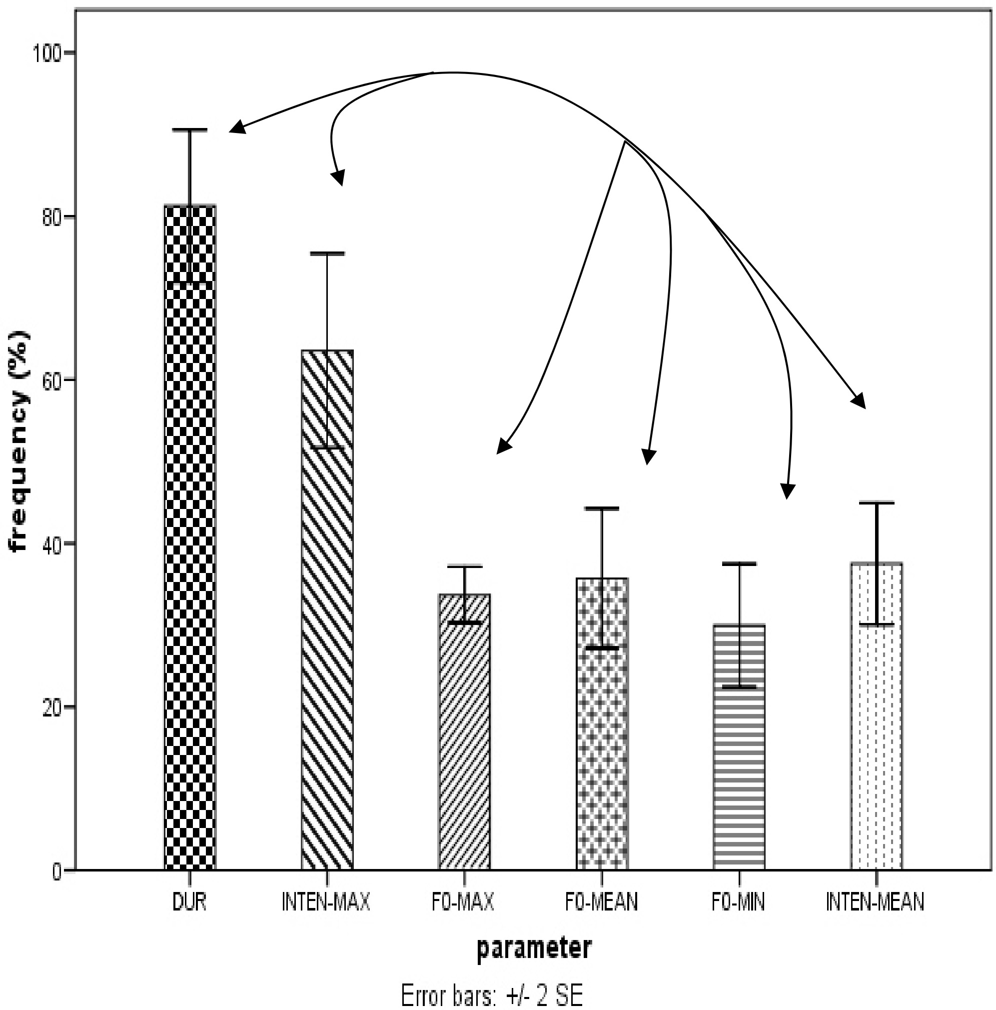


Figure 4-7. Acoustic parameters (and their frequencies) used in 'focus' realization of Tone 3. Arrows indicate significant difference in the frequency at which the two parameters were used

The repeated-measure analysis showed that, with an alpha level of .05, the frequency at which these acoustic parameters was used in 'focus' realization in Tone 3 was statistically significant different from each other [$F(5, 45) = 29.576$, $P = .000$]. Follow-up pair-wise comparisons suggested that duration and maximum intensity were the most frequently used parameters among all parameters measured to realize focus in Tone 3 [p=.000~.046]. There was no significant difference between duration and maximum intensity [p=.173]. Neither was there difference among mean intensity, mean F_0, maximum F_0 and minimum F_0 [p=1.000].

Tone 4 Seven acoustic dimensions were used to implement focus realization for Tone 4: lengthening the duration, increasing the mean and the maximum values of intensity and F_0, raising minimum F_0 and sharpening the slope from the F_0 onset to offset (shown in Table 4-9 and Figure 4-8).

Minimum F_0 was used in 42.69% of the data, less frequently than intensity dimensions (i.e, mean and maximum intensity) which appeared in 65.00% and 70.00% of the data. The intensity parameters were also used less frequently than the duration dimension in 86.25% of the data, and most F_0 dimensions (i.e., mean F_0, maximum F_0 and slope in 84.81%, 88.24% and 89.64% of the data respectively).

Table 4-9. Descriptive analysis of parameters used for focus realization in Tone 4

Parameters	Mean (%)	Std. Deviation
Duration (Dur)	86.25	12.43
Mean Intensity (Inten-mean)	65.00	28.14
Max Intensity (Inten-max)	70.00	17.87
Mean F_0 (F_0-mean)	84.81	12.28
Max F_0 (F_0-max)	88.24	14.51
Min F_0 (F_0-min)	42.69	13.16
F_0 Slope (Slope)	89.64	14.65

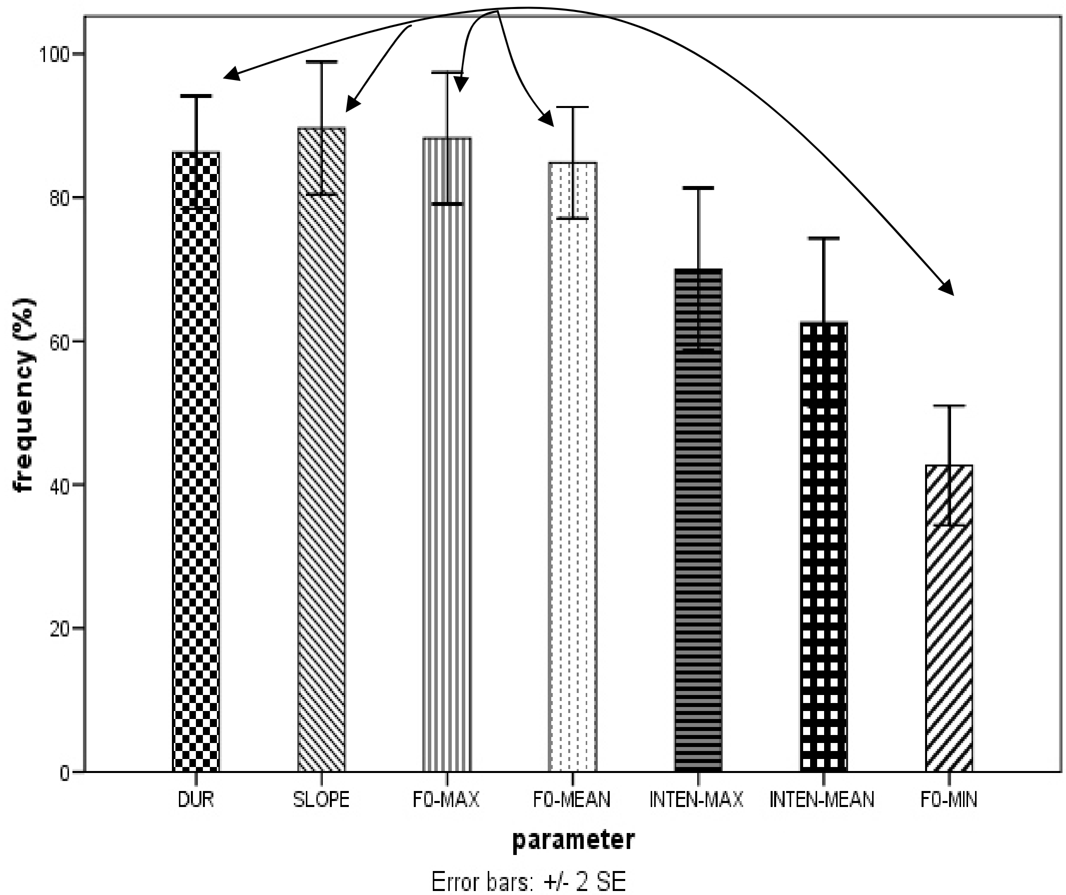


Figure 4-8. Acoustic parameters (and their frequencies) used in 'focus' realization of Tone 4. Arrows indicate significant difference in the frequency at which the two parameters were used

Results of the repeated-measure ANOVA analysis suggested that the presence differences across parameters were statistically significant [F (6, 54) =13.902, P=.000]. Follow-up pair wise comparison illustrated that duration, mean F_0 , maximum F_0 and F_0 slope were used significantly more frequently than minimum F_0 to implement focus [p = .000 between minimum F_0 and duration (or F_0 slope); p = .001 between minimum F_0 and mean F_0; p =.005 between minimum F_0 and maximum F_0.] The differences among duration and F_0 parameters (except minimum F_0) was not significant [p= 1.000]. Neither was the difference among mean intensity, maximum intensity and minimum F_0 significant. [p=1.000 between the two intensity parameters; p= .800 between mean intensity and minimum F_0; p = .065 between maximum intensity and minimum F_0].

Acoustic parameters for accent realization

In Table 4-5. (Repeated in Table 4-10), accent realization made use of the duration parameter in a dominant way.

Table 4-10. Acoustic parameters for accent realization in unfocused positions

Tones	DUR	INTEN-MEAN	INTEN-MAX	F_0-MEAN	F_0-MAX	F_0-MIN	SLOPE
T1	**81%**	37%	45%	12%			
T2	**81%**	31%	39%	21%	28%	15%	35%
T3	**83%**	22%	39%	39%	39%	26%	
T4	**85%**	34%	41%	41%	41%	7%	41%

Tone 1 Four acoustic parameters were used in the implementation of 'accent' in unfocused positions for Tone 1: duration, mean and maximum intensity, and mean F_0 (shown in Table 4-11and Figure 4-9). Numerically, the duration was the most frequent dimension adopted (used in 80.89% of the data), followed by intensity measures (present in 37.02% and 45.00% of the data) and mean F_0 (present in 11.67% of the data).

These observations were submitted to a repeated-measure ANOVA with acoustic parameter as the within-subject factor. The results suggested that with an alpha level of .05, the differences of mean percentage were statistically significant among the acoustic parameters measured [$F(3, 27) = 49.925$, $P = .000$]. Follow-up pair-wise comparisons with Bonferroni adjustment were conducted. The results (shown in Figure 4-9) revealed that duration was used more frequently than other parameters to implement accent [$p=.000$ between duration and mean intensity as well as between duration and mean F_0; $p=.001$ between duration and maximum intensity]. Mean and maximum intensity were also used more frequently than mean F_0 [$p=.004$ and .003] for accent realization. The difference between two intensity parameters was not significant [$p=1.000$].

Table 4-11. Descriptive analysis of parameters used for accent realization in Tone 1

Parameters	Mean (%)	Std. Deviation
Duration (Dur)	80.89	13.73
Mean Intensity (Inten-mean)	37.02	14.27
Max Intensity (Inten-max)	45.00	16.87
MeanF_0 (F_0-mean)	11.67	9.38

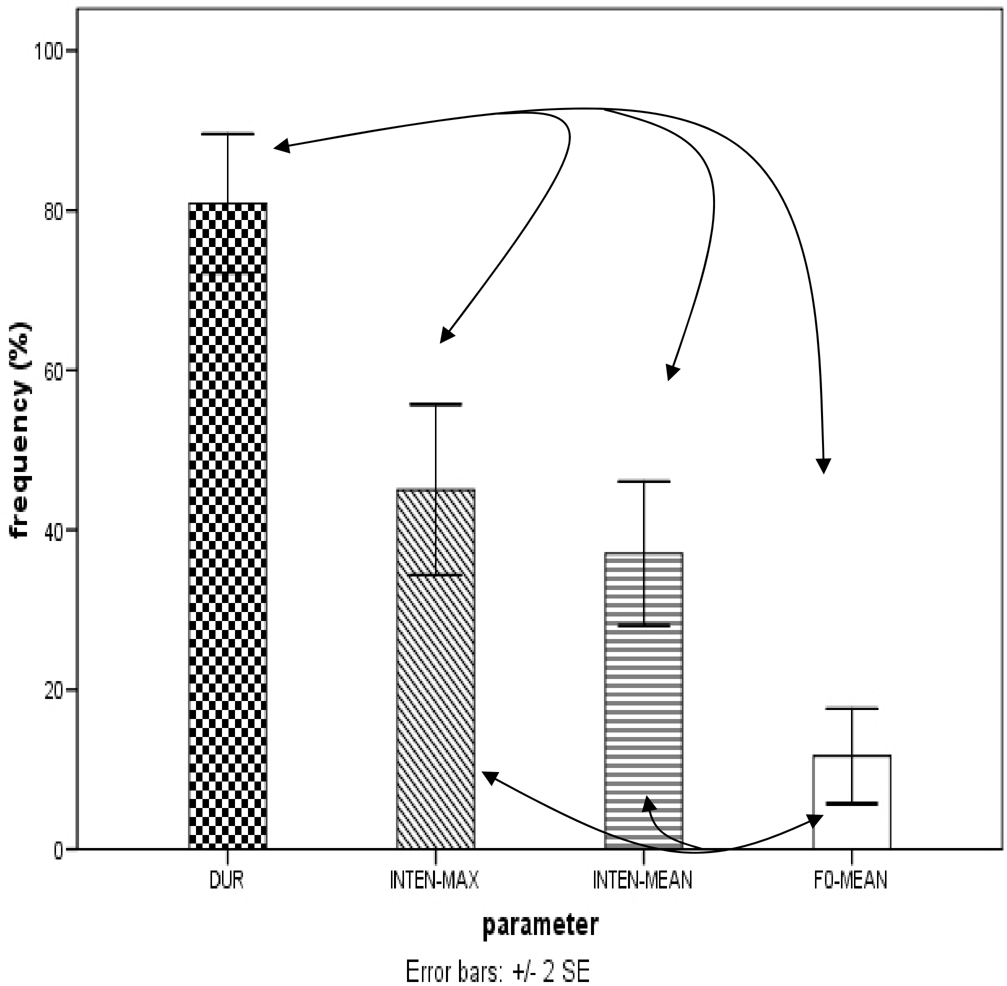


Figure 4-9. Acoustic parameters (and their frequencies) used in 'accent' realization of Tone 1. Arrows indicate significant difference in the frequency at which the two parameters were used

Tone 2 Seven acoustic parameters were measured to examine accent realization in unfocused positions for Tone 2: duration, mean and maximum intensity, mean, maximum and minimum F0, and the slope of F0 from the onset to the offset (shown in Table 4-12 and Figure 4-10).

Table 4-12. Descriptive analysis of parameters used for accent realization in Tone 2

Parameters	Mean (%)	Std. Deviation
Duration (Dur)	81.25	12.15
Mean Intensity (Inten-mean)	31.17	11.91
Max Intensity (Inten-max)	39.39	14.98
Mean F_0 (F_0-mean)	21.11	11.15
Max F_0 (F_0-max)	27.98	17.29
Min F_0 (F_0-min)	15.42	12.89
F_0 Slope (Slope)	35.45	12.95

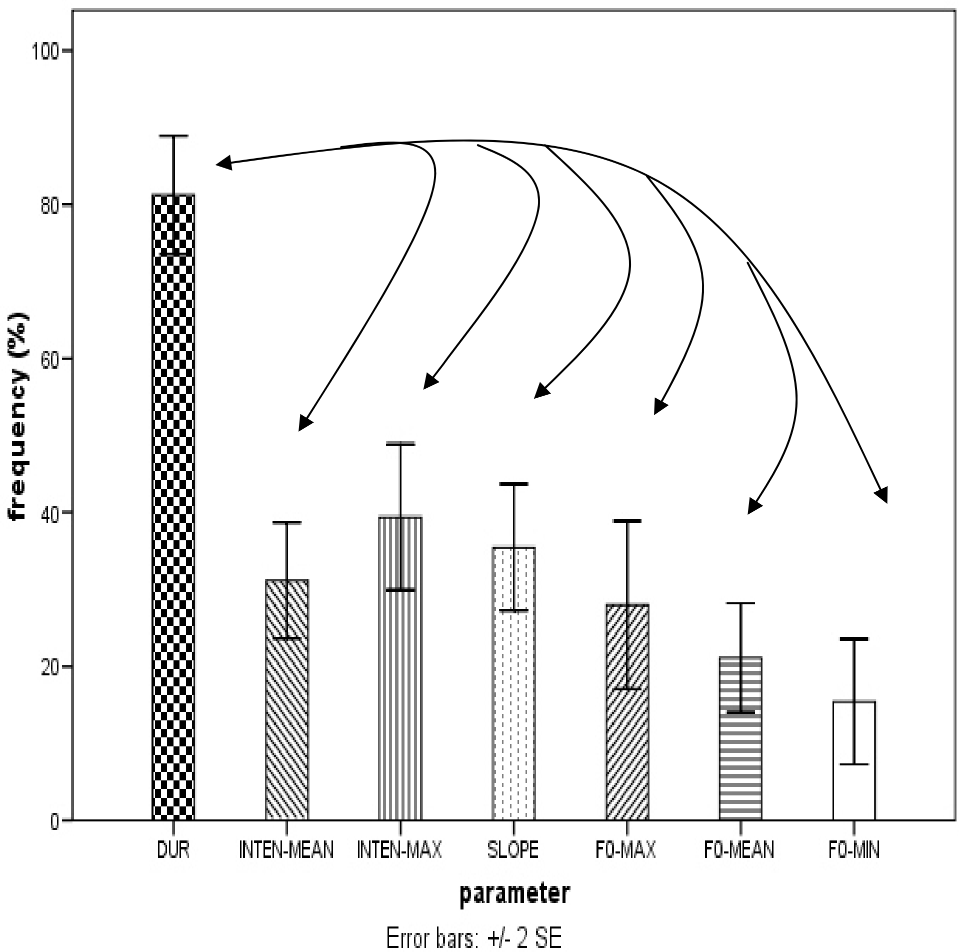


Figure 4-10. Acoustic parameters (and their frequencies) used in 'accent' realization of Tone 2. Arrows indicate significant difference in the frequency at which the two parameters were used

Duration Lengthening was used in81.25% of the data, numerically more than other acoustic parameters, such as mean intensity (present in 31.17% of the data), maximum intensity (present in 39.39% of the data), mean F_0 (present in 21.11% of the data), maximum F_0 (present in

27.98% of the data), minimum F_0 (present in 15.42% of the data), and F_0 slope (present in 35.45% of the data) to realize accent. The repeated-measure analysis showed that, with an alpha level of .05, the frequency at which these acoustic parameters was used in 'accent' realization in Tone 2 was statistically significant different from each other [$F(6, 54) = 24.44$, $P = .000$]. Follow-up pair-wise comparisons with Bonferroni adjustment suggested that duration were used significantly more frequently than other acoustic parameters [p=.000~.006]. There was no significant difference among intensity and F_0 parameters [p=1.000~.104].

Tone 3 Six acoustic dimensions were measured to examine accent realization for Tone 3 (shown in Table 4-13). The duration dimension was used in 82.50% of the data, followed by mean F_0, maximum F_0,, and maximum intensity which were used in 39.36%, 38.57% and 38.94% of the data. Mean intensity and minimum F_0 were used less frequently in 21.61% and 26.01% of the data.

Table 4-13. Descriptive analysis of parameters used for accent realization in Tone 3

Parameters	Mean (%)	Std. Deviation
Duration (Dur)	82.50	13.44
Mean Intensity (Inten-mean)	21.61	8.69
Max Intensity (Inten-max)	38.94	12.21
Mean F_0 (F_0-mean)	39.36	12.03
Max F_0 (F_0-max)	38.57	14.47
Min F_0 (F_0-min)	26.01	14.50

The repeated-measure analysis showed that, with an alpha level of .05, the frequency at which these acoustic parameters was used in 'accent' realization in Tone 3 was statistically significant different from each other [$F(5, 45) = 35.96$, $P = .000$]. Follow-up pair-wise comparisons suggested that duration was the most frequently used parameter among all parameters measured to realize accent in Tone 3 [p=.000~.046]. Maximum intensity and mean F_0

were also used more frequently than mean intensity to realize accent [p= .032 between maximum and mean intensity; p=.041between mean F_0 and intensity] (Figure 4-11).

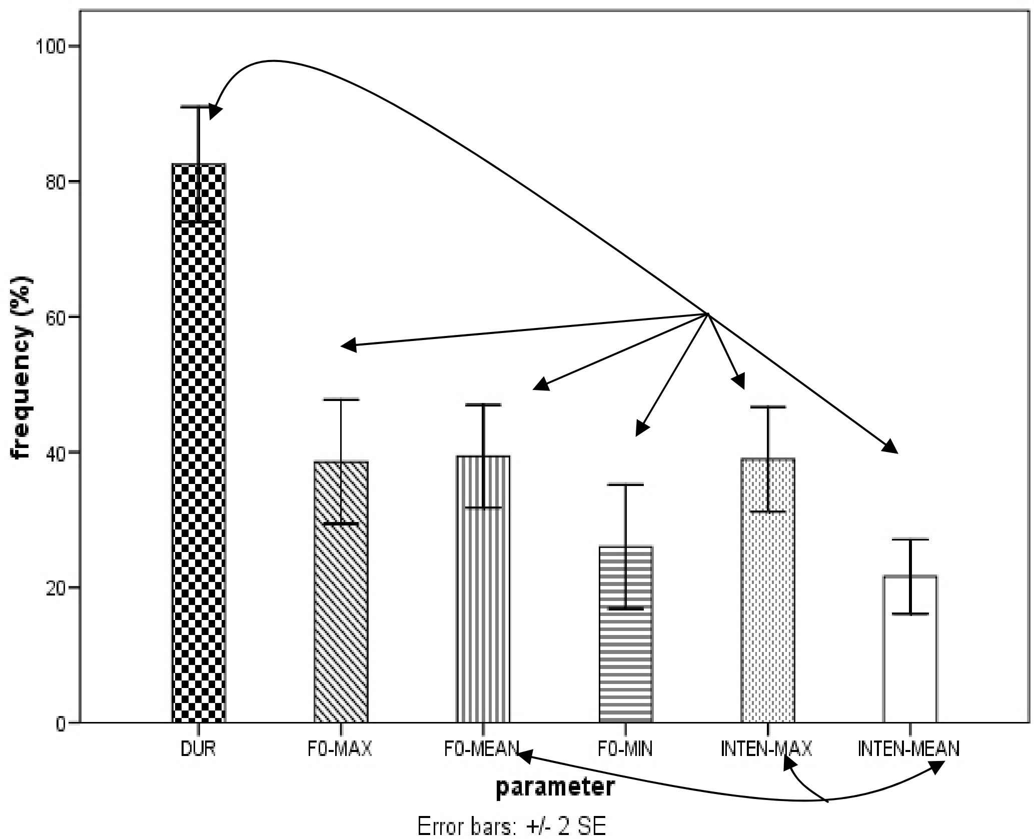


Figure 4-11. Acoustic parameters (and their frequencies) used in 'accent' realization of Tone 3. Arrows indicate significant difference in the frequency at which the two parameters were used

Tone 4 Seven acoustic dimensions were used to implement accent for Tone 4: lengthening the duration, increasing the mean and the maximum values of intensity and F_0, raising minimum F_0 and sharpening the slope from the F_0 onset to offset (shown in Table 4-14). Minimum F_0 was used in 6.85% of the data, less frequently than intensity and other F_0 dimensions (i.e, mean and maximum intensity, mean and maximum F_0, slope F_0) which appeared in between 33.75% and 41.25% of the data. The intensity and F_0 parameters were also used less frequently than the duration dimension in 84.82% of the data.

Table 4-14. Descriptive analysis of parameters used for accent realization in Tone 4

Parameters	Mean (%)	Std. Deviation
Duration (Dur)	84.82	7.86
Mean Intensity (Inten-mean)	33.75	13.24
Max Intensity (Inten-max)	40.89	13.79
Mean F_0 (F_0-mean)	41.25	14.49
Max F_0 (F_0-max)	41.01	16.07
Min F_0 (F_0-min)	6.85	9.40
F_0 Slope (Slope)	40.61	14.73

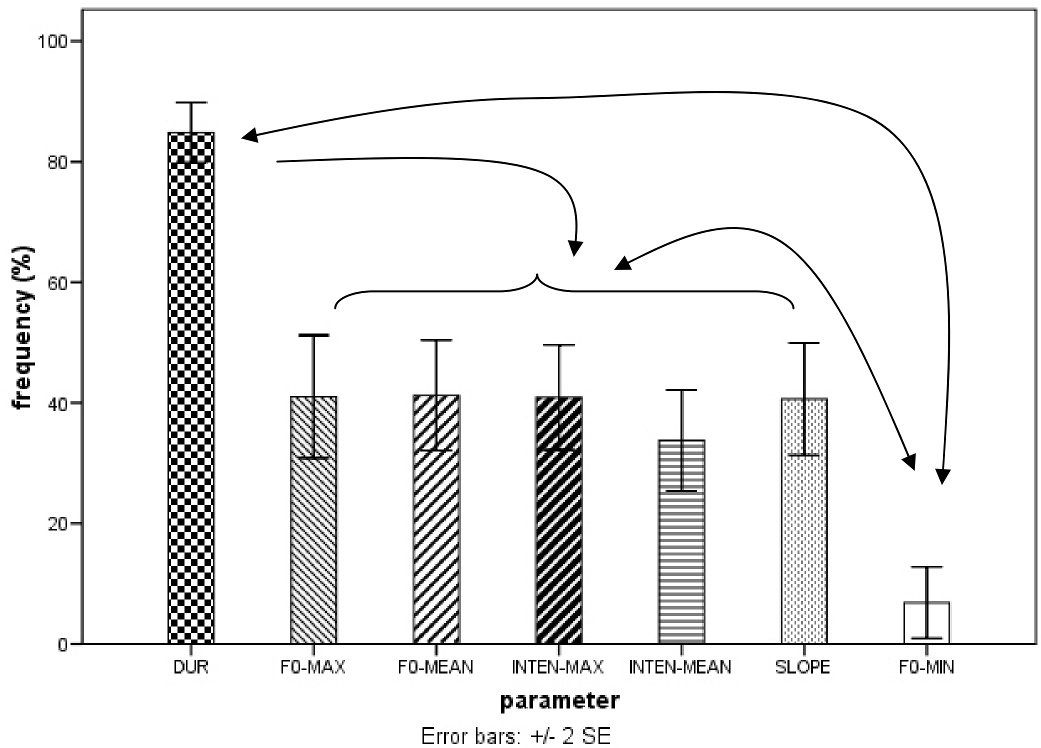


Figure 4-12. Acoustic parameters (and their frequencies) used in 'accent' realization of Tone 4. Arrows indicate significant difference in the frequency at which the two parameters were used

Results of the repeated-measure ANOVA analysis suggested that the presence differences across parameters were statistically significant [F (6, 54) =42.013, P=.000] (Figure 4-12). Follow-up pair wise comparison illustrated that duration were used significantly more frequently than all other parameters to implement accent [p = .000]. The difference among mean and maximum intensity, mean and maximum F_0, and slope F_0 was not significant [p= 1.000], while these parameters were also used more frequently than minimum F_0 [p=.000~.004].

Summary for Research Question 1

Seven acoustic parameters were measured for focus and accent realization. Both numerical rankings and statistical analyses suggested that acoustic parameters were differentially ranked in each tone and there was a boundary between parameters used in more than 60% of the data and those appearing in less than 45% of the data. Focus realization, in general, was implemented by six acoustic parameters: lengthening the duration, increasing the mean and the maximum values of intensity and F_0, and sharpening the slope from the F_0 onset to offset. Accent was mostly realized by duration.

For focus realization, tones differed in the main acoustic parameters used. Tone 1 used duration, mean F_0, mean and maximum intensity to implement focus (i.e., these four parameters were used in more than 60% of the data, and the difference among their frequencies for focus realization was not significant). Tone 2 used duration, maximum F_0 and F_0 slope to realize focus (i.e., these parameters were used in more than 60% of the data, significantly more frequently than other parameters measured). Tone 3 used duration and maximum intensity for focus implementation (i.e., these parameters were used in more than 60% of the data and appeared significantly more frequently than other parameters). Tone 4 used all parameters except minimum F_0 to produce focus (i.e., duration, mean and maximum intensity, mean, maximum F_0 and slope F_0 appeared in more than 60% of the data, and the difference between these six parameters and minimum F_0 was significant). In both Tone 1 and Tone 4, duration, intensity and F_0 parameters were used. Focus in Tone 2 was implemented by duration and F_0, and by duration and intensity in Tone 3. In other words, duration was the only parameter that was used by all four lexical tones to realize focus, whileF_0 and intensity parameters were used by some tones.

For accent realization, duration was the major parameter used by four tones (i.e., duration was the only parameter used in more than 60% of the data, and the difference in frequencies between duration and other parameters was significant), and other acoustic parameters such as intensity and F_0 also appeared in a certain percentage of data to implement accent.

In the next two chapters (i.e., Chapter Five ‘Interaction among tone, accent and focus in realization’ and Chapter Six ‘Acoustic cues for focus perception’), main acoustic parameters appearing in more than 60% of the data will be examined.

CHAPTER 5
INTERACTIONS AMONG TONE, ACCENT AND FOCUS IN REALIZATION

This chapter will discuss the interactions among tone, focus and accent (RQ2) in the production experiment. The methodology for the experiment was already described in Chapter Four and will not be repeated in this chapter. In this chapter, the effects of tone and accent on focus realization will be first described. In this section, focus realization implemented by six acoustic parameters (Duration, Mean and maximum of intensity and F_0, F_0 slope) is analyzed among different tones in both accented and unaccented positions. Next, the effects of tone and focus on accent realization will be presented in a similar fashion (accent implemented by the duration parameter is analyzed among lexical tones in focused and unfocused positions). A summary of RQ2 will be provided at the end of the chapter.

Research Question 2: Interactions among Tone, Accent and Focus in the Realization of Focus and Accent?

Acoustic parameters used in the implementation of 'focus' and 'accent' were discussed separately in the previous chapter. . Comparing prominence realizations where focus and accent were realized with the presence of the other (i.e., focus realization in accented positions and accent realization in focused positions) will be the focus of this chapter. Exploring interactions among tone, accent and focus provides answer(s) to Research Question 2 'What are the interactions among tone, accent and focus in the realization of focus and accent?'. For instance, the comparison between two focus realizations (in unaccented positions vs. accented positions) indicated effects of two main factors (tone and accent), and their interaction on focus realizations. Similarly, the comparison between accent realizations (in unfocused positions vs. focused positions) suggested the effects of tone and focus, as well as the interaction between accent realizations.

Comparisons were conducted on six acoustic parameters for focus realizations (e.g., duration, the mean and the maximum of intensity and F_0, and F_0 slope) and on the dominant duration parameter for accent realizations. Acoustic parameters were discussed in terms of their frequencies in the prominent data and their ratio values compared with non-prominent data. In other words, I compared the percentage of data that made use of a particular acoustic parameter, as well as the ratio increase in that parameter, in the realization of focus in two environments i.e., accented and unaccented positions. The same analysis was also conducted for accent realizations.

Effects of Tone and Accent on Focus Realizations

Parameter 1: duration

The effects of tone and accent on focus realizations were first examined through the duration parameter Figure 5-1 showed the percentage of data using this parameter to implement focus among tones in either unaccented or accented positions. The ratio increase was also listed Table 5-1. The results were submitted to repeated-measures with Accent (2 levels: Unaccented, Accented) as one within-subject factor and Tone (4levels: Tone1, Tone2, Tone 3, Tone 4) as the other.

Frequency data Analysis showed that the frequency at which an increase in duration was used to realize 'focus' was significantly affected by accent [Accent: $F(1, 9) = 5.646$, $p = .041$]. As shown in Figure 5-1, averaged across all 4 tones, the frequency at which increased duration was used to realize 'focus' in unaccented positions (in a solid line) was significantly higher than the one used to realize 'focus' in accented positions (in a dash line). However, the frequency at which this parameter was used to realize 'focus' among the four tones was not significantly different [Tone: $F(3, 27) = 0828$, $p = .490$] and the interaction between tone and accent was also insignificant [Tone x Accent: $F(3, 27) = .564$, $p = .643$]. These results suggested that an increase in vowel duration was used more frequently in the realization of 'focus' in an unaccented position than in an ‘accented’ position. That is, regardless of the tone it was produced with, ‘focused’ vowels in an unaccented position were more frequently to lengthen their duration than ‘focused’ vowels in an accented position.

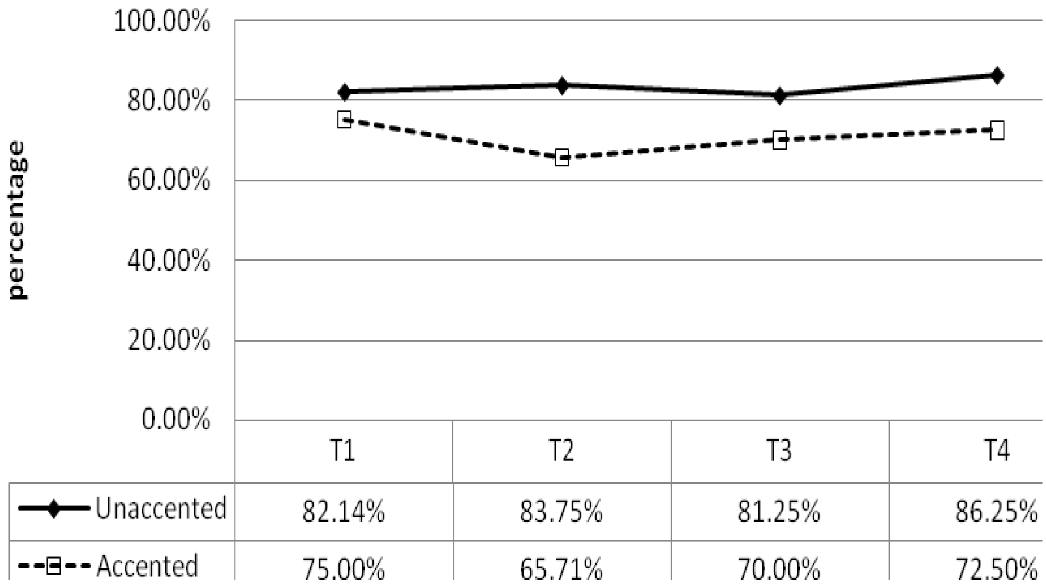


Figure 5-1. Percentages of data using duration as a parameter to realize focus

Ratio data Ratio values in Table 5-1 (except the marginal means listed in the last row) showed a ratio increase in duration produced by each speaker to implement focus in different sentence positions among tones. The values were generated by averaging repetitions (i.e., readings under the same category) provided by each speaker. The average mean among speakers in the last row indicated the ratio mean and the standard derivation (SD) for each category. For example, 1.35 (.21) revealed that focus realization for Tone 1 in unaccented positions was implemented by lengthening the duration to 1.35 times of its unfocused counterpart (so if the unfocused Tone1 was 100msec, the focused Tone 1 in unaccented positions would be 135msec), and the SD was .21. All average means and SDs were displayed in Figure 5-2. Repeated measures ANOVA performed on the data revealed that focused vowels in unaccented positions (shown as filled dark rectangular) had a significantly higher duration ratio than those produced in accented positions (shown as unfilled triangle). [Accent: $F(1, 9) = 15.309$, $p = .004$] and that duration ratio varied significantly among the four tones [Tone: $F(3, 27) = 4.969$, $p = .007$]. Follow up pair-wise comparisons suggested that focused vowels produced with Tone 3 had a significantly higher duration ratio increase than Tone 1 (shown in Table 5-2). No significant interaction was observed between tone and accent [Tone x Accent: $F(3, 27) = .888$, $p = .460$]. In sum, similar to the frequency data reported above, analyses performed on the duration ratio data revealed that 'focus' was more effectively realized in unaccented positions than in accented positions. Specifically, duration of focused vowels produced in unaccented positions was lengthened to a significantly greater extent than those produced in accented positions. In addition, averaged across both accented and unaccented conditions, focused vowels produced with tone 3 were lengthened to significantly larger extent than those produced with tone 1.

Table 5-1. Ratio means and the standard derivations of duration parameter for focus realizations*

	Unaccented Position				Accented Position			
Speakers	Tone 1	Tone 2	Tone 3	Tone 4	Tone1	Tone2	Tone3	Tone4
S1	1.50	1.51	1.43	1.49	1.26	1.13	1.58	1.25
S2	1.39	1.76	1.39	1.38	1.28	1.22	1.33	1.21
S3	1.60	1.82	1.92	1.85	1.22	1.40	1.38	1.26
S4	1.11	1.10	1.52	1.38	1.08	1.14	1.13	1.20
S5	1.27	1.54	1.19	1.23	1.12	1.16	1.24	1.32
S6	1.19	1.20	1.35	1.31	1.22	1.34	1.23	1.37
S7	1.19	1.32	1.51	1.35	1.09	1.17	1.32	1.21
S8	1.44	1.58	1.61	1.67	1.30	1.29	1.27	1.31
S9	1.69	1.81	1.80	1.58	1.42	1.33	1.56	1.14
S10	1.11	1.10	1.28	1.23	1.11	1.10	1.24	1.18
Average by Speakers	1.35 (.21)	1.47 (.28)	1.50 (.23)	1.45 (.20)	1.21 (.11)	1.23 (.10)	1.33 (.14)	1.25 (.07)

*The number in ( ) indicates standard deviation.

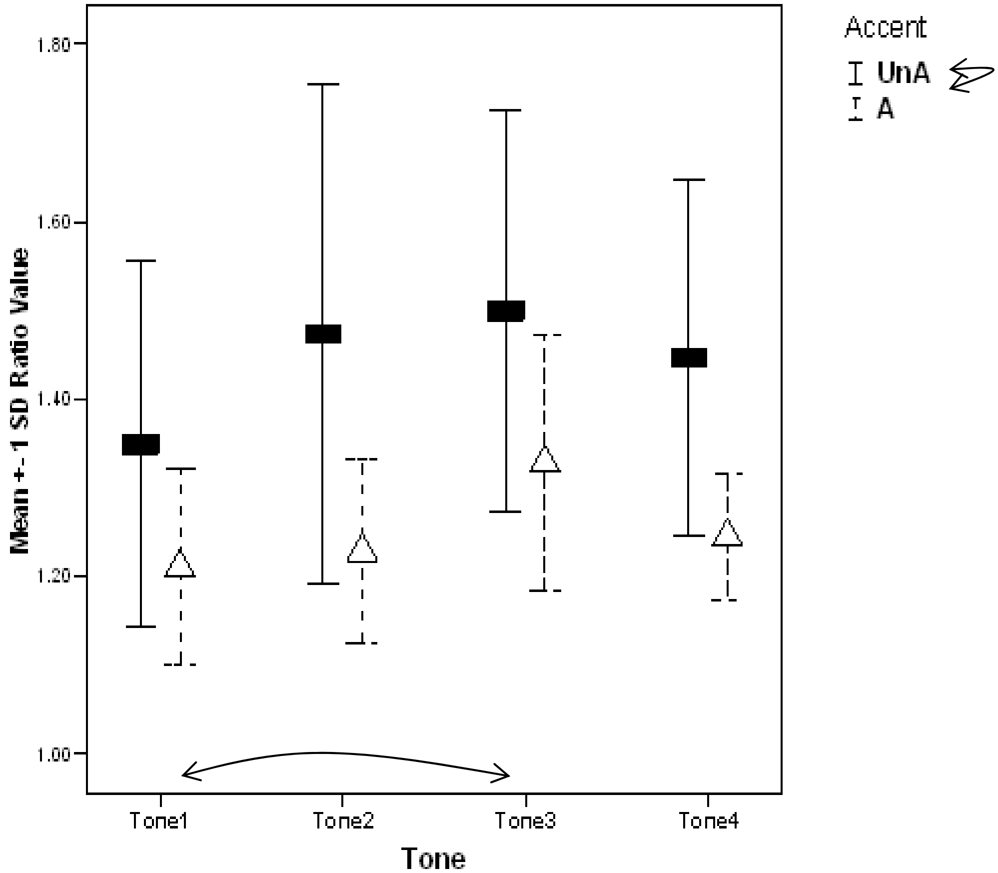


Figure 5-2. Ratio increase of the duration parameter in focus realizations. Arrow indicates a significant difference

Table 5-2. Pair wise comparisons of ratio means among tones

(I) tone	(J) tone	Mean Difference (I-J)	Std. Error	Sig.(a)
1	2	-0.072	0.026	0.135
	3	-.135(*)	0.028	0.006
	4	-0.067	0.037	0.628
2	1	0.072	0.026	0.135
	3	-0.063	0.039	0.870
	4	0.005	0.040	1.000
3	1	.135(*)	0.028	0.006
	2	0.063	0.039	0.870
	4	0.068	0.036	0.564
4	1	0.067	0.037	0.628
	2	-0.005	0.040	1.000
	3	-0.068	0.036	0.564

*. The mean difference is significant at the .05 level. a. Adjustment for multiple comparisons: Bonferroni.

Parameter 2: maximum intensity

Increasing the maximum value of intensity to realize focus was used in three tones: Tone 1, Tone 3 and Tone4. Figure 5-3 showed the percentage of data using this parameter to implement focus, and Table 5-3 listed the ratio increase. Both the frequency data and the ratio data were submitted to repeated-measures ANOVA with Accent (2 levels: Unaccented, Accented) and Tone (3levels: Tone1, Tone 3, Tone 4) as main factors.

Frequency data Analysis showed that the frequency at which an increase in maximum intensity was used to realize 'focus' was not significantly affected by main factors [Accent: F (1, 9) = 1.324, p =.279; Tone: F (2, 18) = 1.778, p =.197], or the interaction between accent and tone [Tone x Accent: F (2, 18) = .007, p =.993]. These results suggested that there was no significant difference among frequencies at which an increase in maximum intensity was used in the realization of 'focus'.

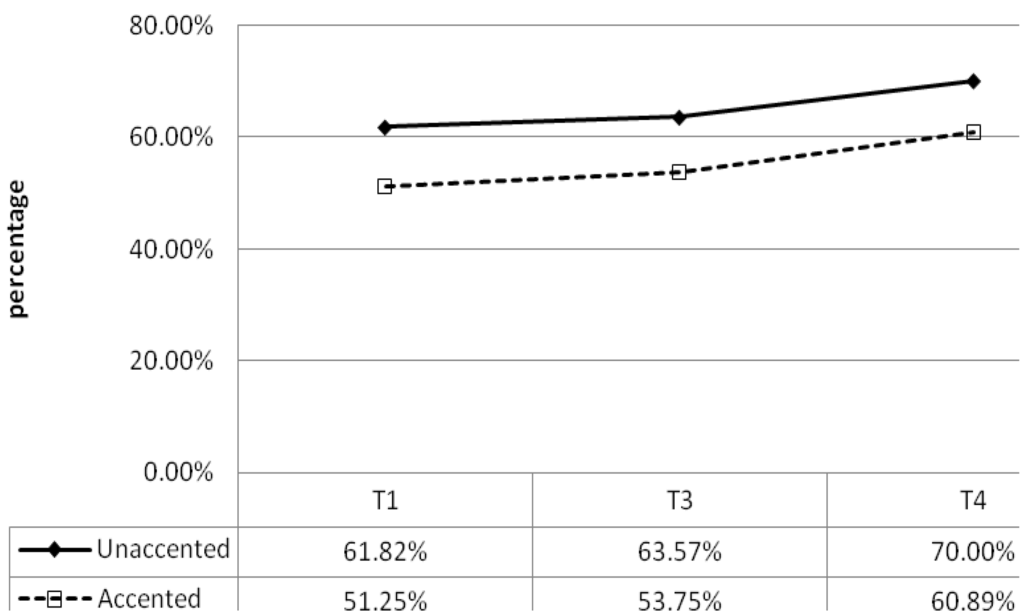


Figure 5-3. Percentages of data using intensity-max as a parameter to realize focus

Ratio data Ratio values in Table 5-3 showed a ratio increase in maximum intensity produced by each speaker to implement focus in different sentence positions among tones. All average means and SDs were displayed in Figure 5-4. Repeated measures ANOVA performed on the data revealed that focused tones in unaccented positions (shown as filled dark rectangular) had a significantly higher maximum intensity ratio than those produced in accented positions (shown as unfilled triangle). [Accent: $F(1, 9) = 8.148$, $p = .019$]. However, the ratio was not significantly different among the tones [Tone: $F(2, 18) = 1.208$, $p = .322$] and the interaction between tone and accent was also insignificant [Tone x Accent: $F(2, 18) = 1.122$, $p = .347$]. In sum, analyses performed on the maximum intensity ratio data revealed that 'focus' was more effectively realized in unaccented positions than in accented positions. Specifically, maximum intensity of focused vowels produced in unaccented positions was increased to a significantly greater extent than those produced in accented positions.

Table 5-3. Ratio means and the standard derivations of maximum intensity parameter for focus realizations

	Unaccented Position			Accented Position		
Speakers	Tone 1	Tone 3	Tone 4	Tone 1	Tone 3	Tone 4
S1	4.41	3.37	3.19	3.03	1.40	2.37
S2	4.89	1.97	1.46	2.05	3.61	2.78
S3	5.19	2.62	4.94	3.62	3.18	4.14
S4	1.60	1.76	2.96	2.48	2.08	2.70
S5	2.40	2.87	3.14	2.95	2.69	2.18
S6	3.31	2.37	4.33	2.76	1.96	3.98
S7	4.00	4.66	4.93	3.56	1.54	5.85
S8	6.44	2.74	5.65	2.05	3.12	7.67
S9	2.24	4.18	2.43	2.13	2.72	1.98
S10	5.26	5.98	4.82	2.33	4.87	1.76
Average by Speakers	3.97 (1.56)	3.25 (1.32)	3.79 (1.34)	2.70 (.59)	2.72 (1.05)	3.54 (1.92)

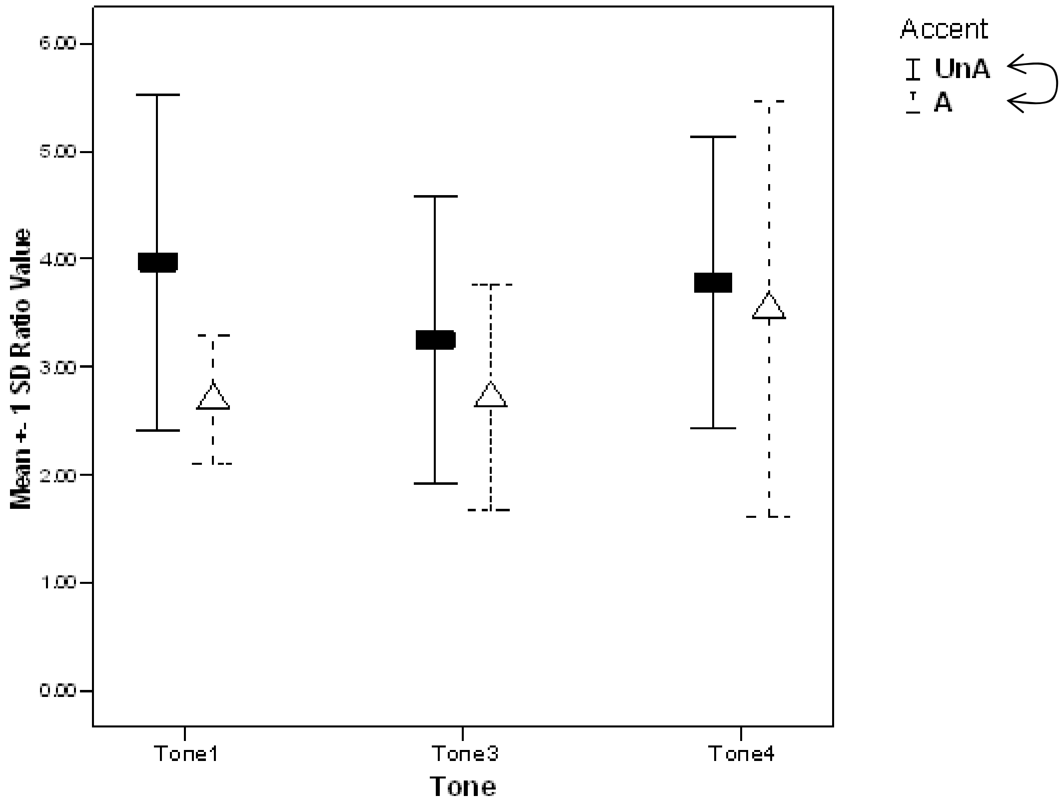


Figure 5-4. Ratio increase of the maximum intensity parameter in focus realizations Arrow indicates a significant difference

Parameter 3: mean intensity

Increasing the mean value of intensity to realize focus was used in Tone 1 and Tone4. Similarly, both the frequency data (in Figure 5-5) and the ratio data (in Table 5-4) were submitted to repeated-measures ANOVA with Accent (2 levels: Unaccented, Accented) and Tone (2levels: Tone1, Tone 4) as the within-subject factors.

Frequency data Analysis showed that the frequency at which an increase in mean intensity was used to realize 'focus' was not significantly affected by main factors [Accent: F (1, 9) = 3.379, p =.099; Tone: F (1, 9) = 2.262, p =.167], or the interaction between accent and tone [Tone x Accent: F (1, 9) = .557, p =.475]. These results suggested that there was no significant difference among frequencies at which an increase in mean intensity was used in the realization of 'focus'.

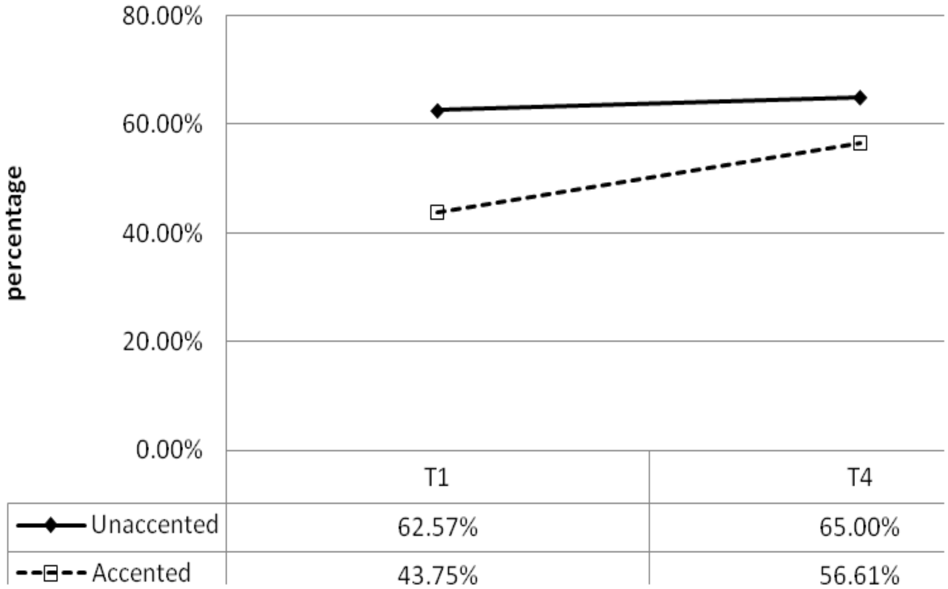


Figure 5-5. Percentage of data using intensity-mean as a parameter to realize focus

Ratio data The ratio values in Table 5-4 showed a ratio increase in mean intensity produced by each speaker to implement focus in different sentence positions among tones. All average means and SDs were displayed in Figure 5-6. Repeated measures ANOVA performed on the data revealed that focused tones in unaccented positions (shown as filled dark rectangular) had a significantly higher mean intensity ratio than those produced in accented positions (shown as unfilled triangle). [Accent: $F(1, 9) = 6.230$, $p = .034$]. However, the ratio was not significantly different between the tones [Tone: $F(1, 9) = 2.063$, $p = .185$] and the interaction between tone and accent was also not significant [Tone x Accent: $F(1, 9) = .730$, $p = .415$]. These results suggested that 'focus' was more effectively realized in unaccented positions than in accented positions. Specifically, mean intensity of focused vowels produced in unaccented positions was increased to a significantly greater extent than those produced in accented positions.

Table 5-4. Ratio means and the standard derivations of mean intensity parameter for focus realizations

Speakers	Unaccented Position Tone 1	Unaccented Position Tone 4	Accented Position Tone1	Accented Position Tone4
S1	4.53	3.16	2.55	1.68
S2	4.47	2.78	2.56	2.00
S3	7.03	4.22	2.81	2.14
S4	3.44	3.48	2.34	2.76
S5	2.60	2.48	2.77	3.87
S6	3.89	3.43	4.13	3.99
S7	4.00	4.53	5.63	3.32
S8	5.23	3.93	2.13	4.63
S9	2.38	4.06	2.89	2.48
S10	5.04	4.15	2.31	1.99
Average by Speakers	4.26 (1.35)	3.62 (.67)	3.01 (1.07)	2.89 (1.01)

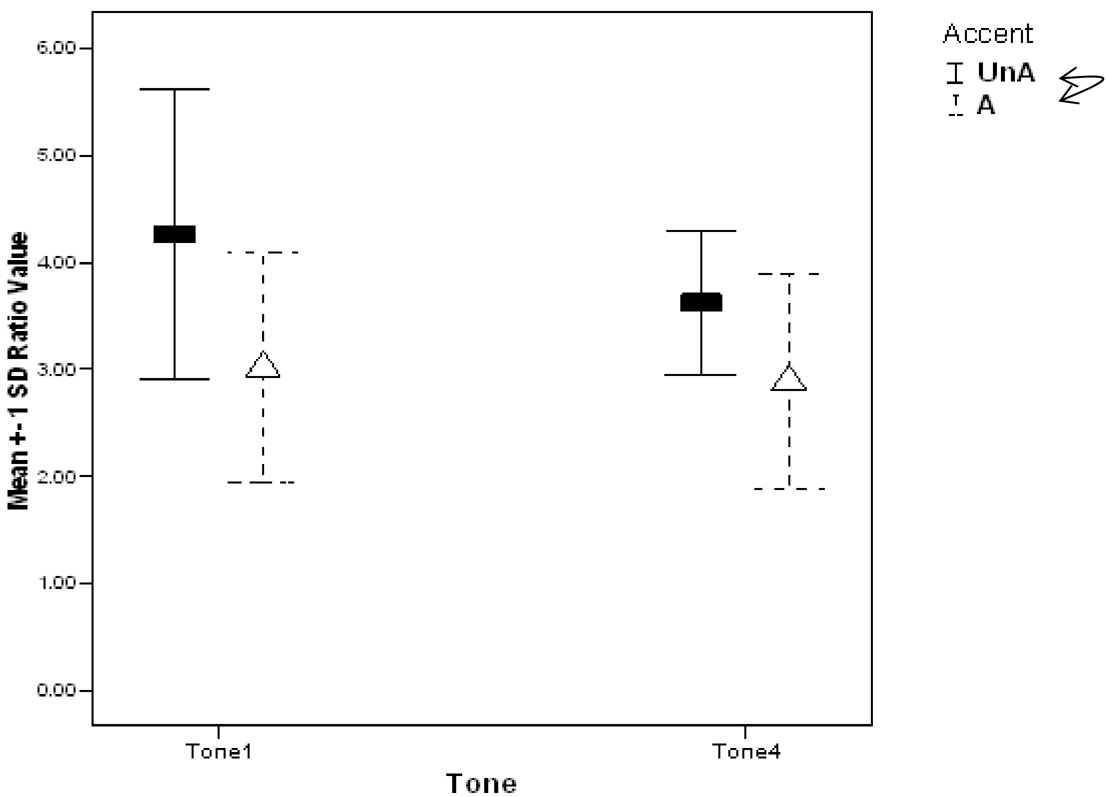


Figure 5-6. Ratio increase of the mean intensity parameter in focus realizations. Arrow indicates a significant difference

Parameter 4: mean F_0

Figure 5-7 showed the percentage of data using mean F_0 to implement focus in Tone 1 and Tone 4 in either unaccented or accented positions. The ratio increase was also listed Table 5-5. The results were submitted to repeated-measures with Accent (2 levels: Unaccented, Accented) as one within-subject factor and Tone (2levels: Tone1, Tone 4) as the other.

Frequency data Analysis showed that the frequency at which an increase in mean F_0 was used to realize 'focus' was significantly affected by accent [Accent: $F(1, 9) = 6.587$, $p = .030$].

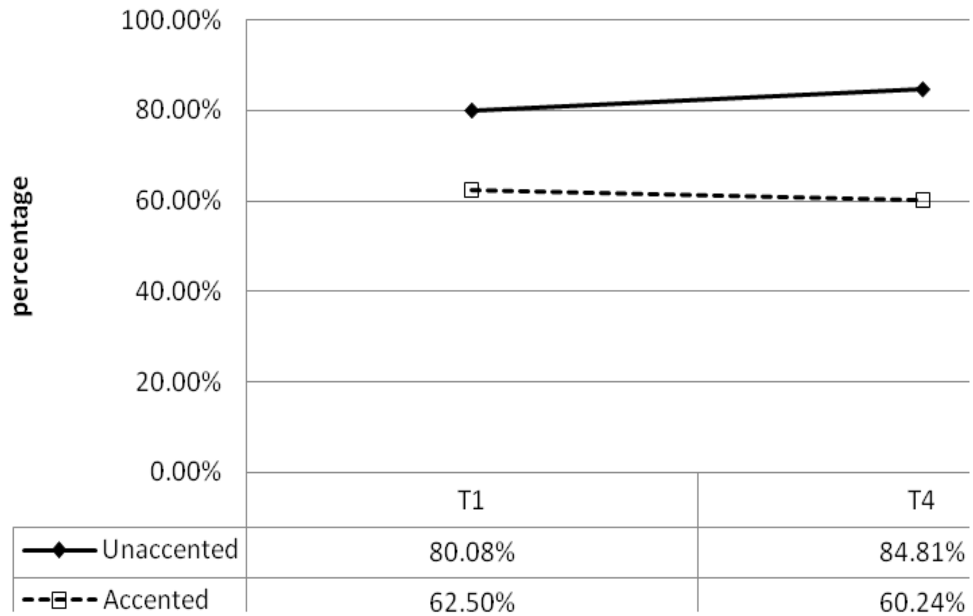


Figure 5-7. Percentages of data using F_0-mean as a parameter to realize focus

As shown in Figure 5-7, averaged between the two tones, the frequency at which increased mean F_0 was used to realize 'focus' in unaccented positions (in a solid line) was significantly higher than the one used to realize 'focus' in accented positions (in a dash line). However, the frequency at which this parameter was used to realize 'focus' between Tone 1 and Tone 4 was not significantly different [Tone: $F(1, 9) = .063$, $p = .808$] and the interaction between tone and accent was not significant [Tone x Accent: $F(1, 9) = .779$, $p = .400$]. These results suggest that an increase in mean F_0 was used more frequently in the realization of 'focus' in an unaccented position than in an 'accented' position. That is, regardless of the tone it was produced with, 'focused' vowels in an unaccented position are more frequently to have higher mean F_0 than 'focused' vowels in an accented position.

Ratio data Ratio values in Table 5-5 showed a ratio increase in mean F_0 produced by each speaker to implement focus in different sentence positions among tones. Average means and SDs were displayed in Figure 5-8.

Table 5-5. Ratio means and the standard derivations of mean F_0 parameter for focus realizations

	Unaccented Position		Accented Position	
Speakers	Tone 1	Tone 4	Tone1	Tone4
S1	20.03	29.86	20.10	25.08
S2	27.30	31.78	11.71	12.46
S3	37.31	44.47	29.12	28.75
S4	31.61	40.91	21.84	27.50
S5	21.52	22.38	24.47	43.45
S6	39.36	46.66	18.16	34.65
S7	27.76	41.41	16.52	20.19
S8	48.55	43.13	27.55	43.14
S9	42.78	28.52	36.54	40.51
S10	20.87	21.37	22.68	12.87
Average by Speakers	31.71 (9.93)	35.05 (9.38)	22.87 (7.04)	28.86 (11.56)

Repeated measures ANOVA performed on the data revealed that focused tones in unaccented positions (shown as filled dark rectangular) had a significantly higher mean F_0 ratio than those produced in accented positions (shown as unfilled triangle). [Accent: $F(1, 9) = 6.081$, $p = .036$], and focused vowels produced with Tone 4 had a significantly higher mean F_0 ratio increase than Tone 1[Tone: $F(1, 9) = 6.522$, $p = .031$]. However, the interaction between tone and accent was not significant [Tone x Accent: $F(1, 9) = .442$, $p = .523$]. In sum, similar to the frequency data reported above, analyses performed on the mean F_0 ratio data revealed that 'focus' was more effectively realized in unaccented positions than in accented positions. Specifically, mean F_0 of focused vowels produced in unaccented positions was increased to a significantly greater extent than those produced in accented positions. In addition, averaged across both accented and unaccented conditions, mean F_0 of focused vowels produced with Tone 4 was increased to significantly larger extent than those produced with Tone 1.

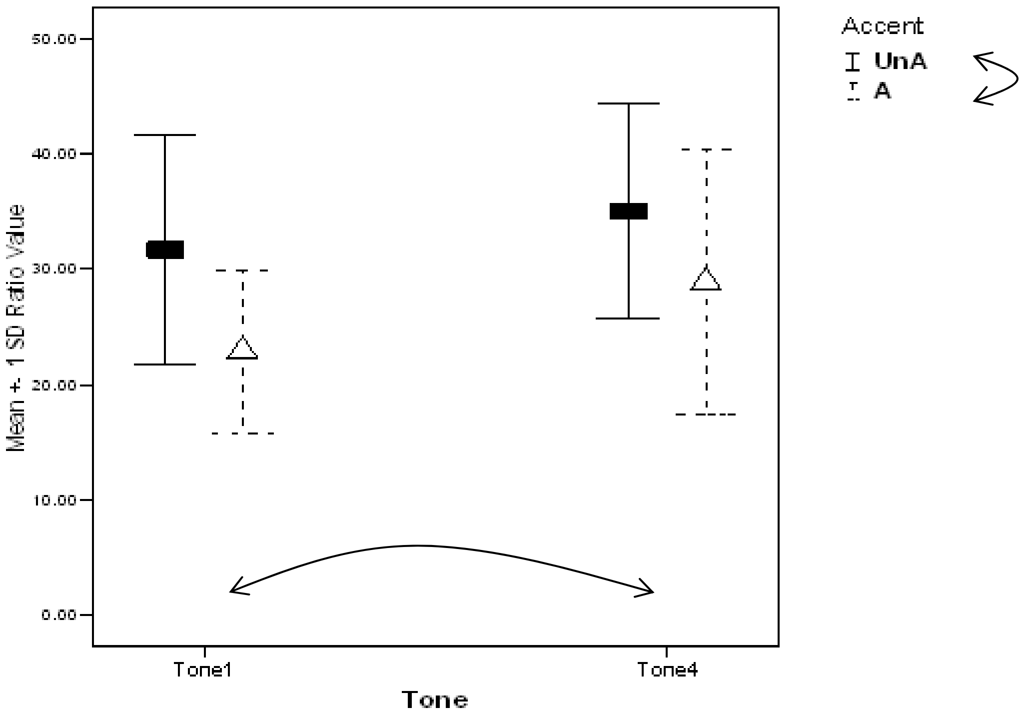


Figure 5-8. Ratio increase of the mean F_0 parameter in focus realizations. Arrow indicates a significant difference

Parameter 5: maximum F_0

Increasing the maximum value of F_0 was used in Tone 2 and Tone4 to realize focus. . Figure 5-9 showed the percentage of data using this parameter to implement focus, and Table 5-6 listed the ratio increase. Both the frequency data and the ratio data were submitted to repeated-measures ANOVA with Accent (2 levels: Unaccented, Accented) and Tone (2levels: Tone2, Tone 4) as main factors.

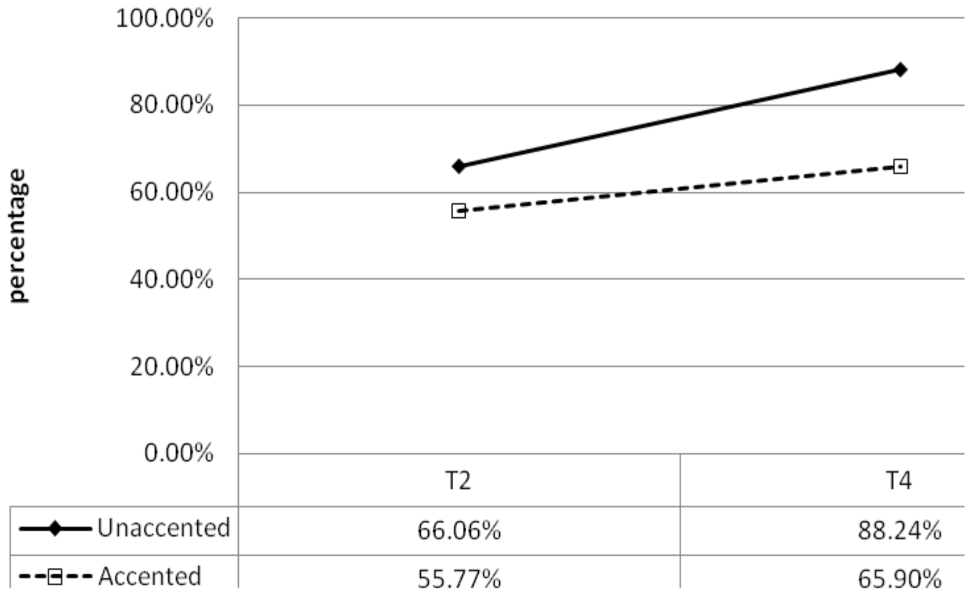


Figure 5-9. Percentage of data using F_0-max as a parameter to realize focus

Frequency data Analysis showed that 'focused' vowels produced with Tone 4 were more frequently to obtain higher maximum F_0 than 'focused' vowels produced with Tone 2 [Tone: $F(1, 9) = 12.961$, $p = .006$]. However, the frequency at which this parameter was used to realize 'focus' between accented and unaccented positions was not significantly different [Accent: $F(1, 9) = 4.671$, $p = .059$] and the interaction between tone and accent was also not significant [Tone x Accent: $F(1, 9) = 2.724$, $p = .133$]. These results suggest that an increase in maximum F_0 was used more frequently in the realization of 'focus' in Tone 4 than in Tone 2. That is, regardless of the position it was placed to, 'focused' vowels in Tone 4 are more frequently to have higher maximum F_0 than 'focused' vowels in Tone 2.

Ratio data Ratio values in Table 5-6 showed a ratio increase in maximum F_0 produced by each speaker to implement focus in different sentence positions among tones. Average means and SDs were displayed in Figure 5-10.

Table 5-6. Ratio means and the standard derivations of maximum F_0 parameter for focus realizations

	Unaccented Position		Accented Position	
Speakers	Tone 2	Tone 4	Tone2	Tone4
S1	32.46	35.18	19.05	19.16
S2	36.82	38.86	37.53	17.83
S3	33.57	51.67	48.02	54.40
S4	28.51	60.63	44.68	40.80
S5	41.45	51.11	47.60	29.48
S6	37.20	56.71	31.06	35.12
S7	28.64	53.54	23.67	23.09
S8	57.19	50.85	35.58	55.19
S9	45.00	32.59	30.43	20.18
S10	20.61	43.27	37.02	35.96
Average by Speakers	36.15 (10.15)	47.44 (9.44)	35.46 (9.72)	33.12 (13.83)

Repeated measures ANOVA performed on the data in Figure 5-10 revealed that focused tones in unaccented positions (shown as filled dark rectangular) had a significantly higher maximum F_0 ratio than those produced in accented positions (shown as unfilled triangle). [Accent: $F(1, 9) = 7.475$, $p = .023$]. However, no significant difference was observed between Tone 2 and Tone 4 [Tone: $F(1, 9) = 2.193$, $p = .173$]. The analysis also showed a significant interaction between accent and tone factors [Tone x Accent: $F(1, 9) = 5.670$, $p = .041$]. Follow up pair-wise comparisons suggested that Tone 4 had a significant higher ratio than Tone 2 in unaccented positions [$t(9) = 2.461$, $p = .036$], but the tonal difference was not significant in accented positions [$t(9) = .635$, $p = .541$]. Similarly, the accent effect was significant for Tone 4 (where the maximum F_0 ratio in unaccented positions was significantly higher than in accented positions [$t(9) = 4.03$, $p = .003$]), but not for Tone 2 [$t(9) = .157$, $p = .879$].

In sum, analyses performed on the maximum F_0 ratio data revealed that, averaged across both Tone 2 and Tone 4, maximum F_0 of focused vowels produced in unaccented positions was increased to a significantly greater extent than those produced in accented positions. In addition, maximum F_0 of focused vowels produced with Tone 4 was increased to significantly larger extent than those produce d with Tone 2 in unaccented positions.

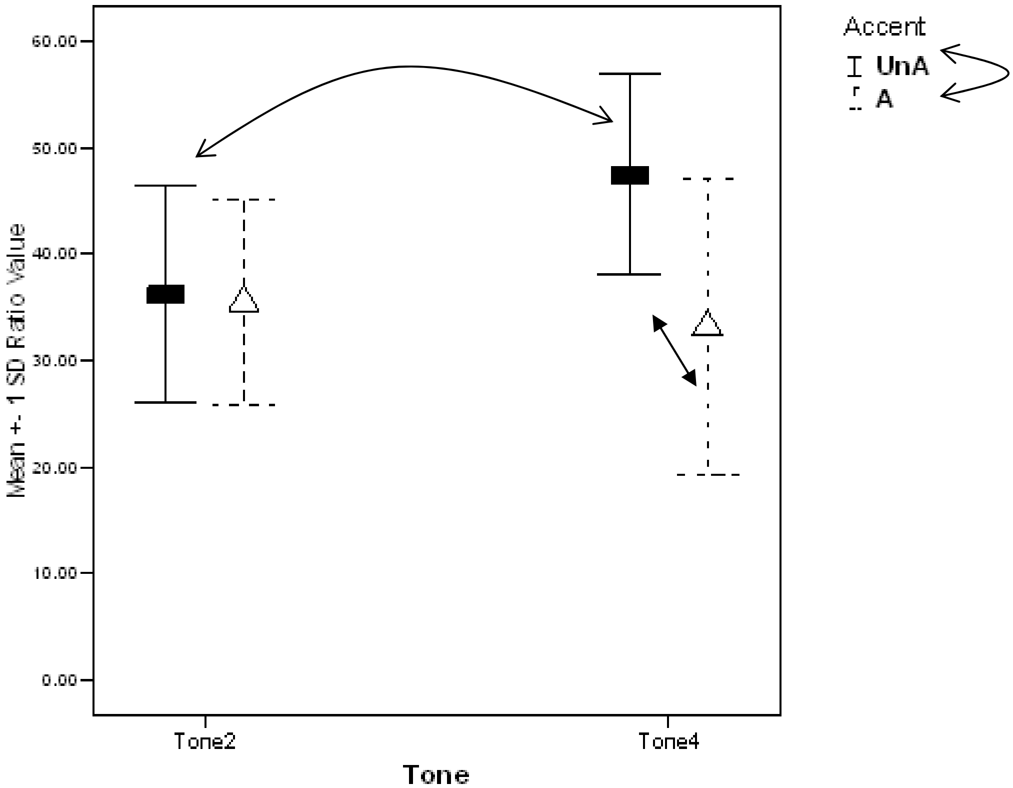


Figure 5-10. Ratio increase of the maximum F_0 parameter in focus realizations. Arrow indicates a significant difference

Parameter 6: F_0 slope

Increasing the slope of F_0 was also used in Tone 2 and Tone 4 to realize focus. Figure 5-11 showed the percentage of data using this parameter to implement and Table 5-7 listed the ratio increase. The results were submitted to repeated-measures with Accent (2 levels: Unaccented, Accented) as one within-subject factor and Tone (2levels: Tone2, Tone 4) as the other.

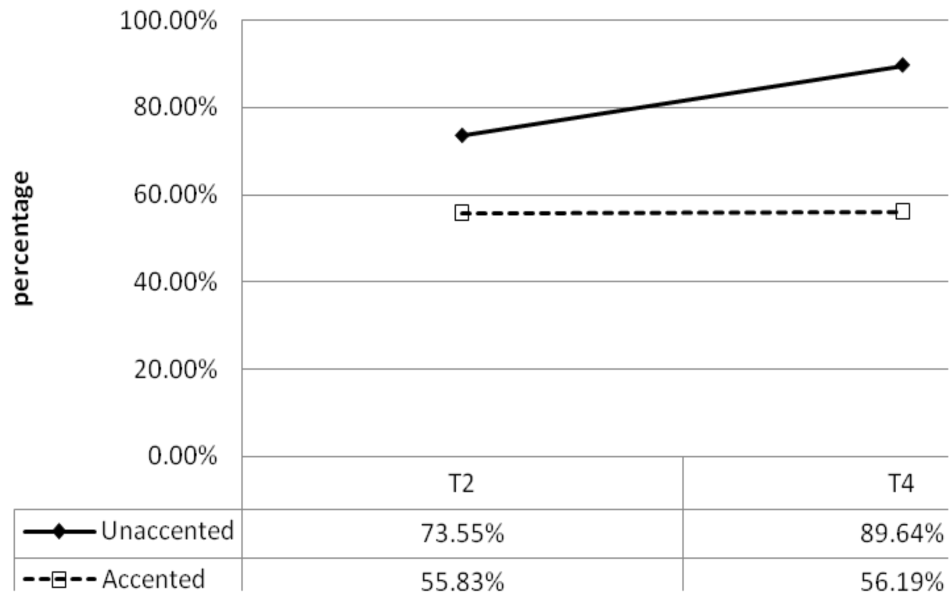


Figure 5-11. Percentage of data using F_0-slope as a parameter to realize focus

Table 5-7. Ratio means and the standard derivations of F_0 slope parameter for focus realizations

	Unaccented Position		Accented Position	
Speakers	Tone 2	Tone 4	Tone2	Tone4
S1	1.38	1.86	1.38	1.42
S2	1.65	2.35	1.15	1.52
S3	2.09	1.71	1.65	2.28
S4	1.27	1.73	1.16	1.36
S5	1.48	1.63	1.35	1.84
S6	1.61	2.95	1.25	1.19
S7	1.58	1.61	1.76	1.53
S8	1.37	2.32	1.05	2.41
S9	1.41	1.82	1.08	1.70
S10	1.24	1.89	1.35	1.23
Average by Speakers	1.51 (.25)	1.99 (.43)	1.32 (.24)	1.65 (.42)

Frequency data Analysis showed that the frequency at which an increase in F_0 slope was used to realize 'focus' was significantly affected by accent [Accent: $F(1, 9) = 26.901$, $p = .001$]. As shown in Figure 5-11, averaged across 2 tones, the frequency at which increased F_0 slope was used to realize 'focus' in unaccented positions (in a solid line) was significantly higher than the one used to realize 'focus' in accented positions (in a dash line). However, the frequency at which this parameter was used to realize 'focus' between Tone 2 and Tone 4 was not significantly different [Tone: F (1, 9) = 1.903, p =.201] and the interaction between tone and accent was also insignificant [Tone x Accent: F (1, 9) = 1.753, p =.218]. These results suggested that an increase F_0 slope was used more frequently in the realization of 'focus' in an unaccented position than in an 'accented' position.

Ratio data Repeated measures ANOVA performed on the data in Figure 5-12 revealed that focused tones in unaccented positions (shown as filled dark rectangular) had a significantly higher slope F_0 ratio than those produced in accented positions (shown as unfilled triangle). [Accent: F (1, 9) = 5.622, p =.042], and focused vowels produced with Tone 4 had a significantly higher slope F_0 ratio increase than Tone 2[Tone: F (1, 9) = 14.247, p =.004]. However, the interaction between tone and accent was not significant [Tone x Accent: F (1, 9) = .485, p =.504].

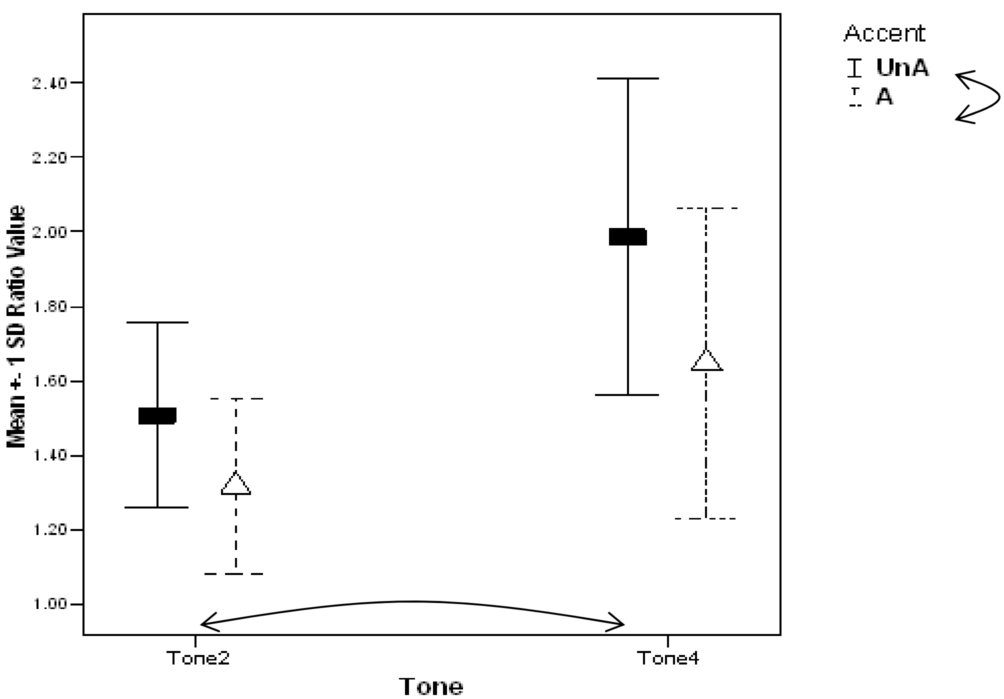


Figure 5-12. Ratio increase of the F_0 slope parameter in focus realizations. Arrow indicates a significant difference

In sum, similar to the frequency data reported above, analyses performed on slope F_0 ratio data revealed that 'focus' was more effectively realized in unaccented positions than in accented positions. Specifically, slope F_0 of focused vowels produced in unaccented positions was increased to a significantly greater extent than those produced in accented positions. In addition, averaged across both accented and unaccented conditions, slope F_0 of focused vowels produced with Tone 4 was increased to significantly larger extent than those produce d with Tone 2.

Effects of Tone and Focus on Accent Realizations

Duration was a dominant parameter used for accent realization among all tones. Similar to the analyses for focus realizations, the frequency data (in Figure 5-13) and the ratio data (in Table 5-8) were submitted to repeated-measures ANOVA with Focus (2 levels: Unfocused, Focused) as one within-subject factor and Tone (4levels: Tone1, Tone2, Tone 3, Tone 4) as the other factor.

Frequency data Analysis showed that the frequency at which an increase in duration was used to realize 'accent' was significantly affected by focus [Focus: $F(1, 9) = 5.308$, $p = .047$]. As shown in Figure 5-13, averaged across all 4 tones, the frequency at which increased duration was used to realize 'accent' in unfocused positions (in a solid line) was significantly higher than the one used to realize 'accent' in focused positions (in a dash line). However, the frequency at which this parameter was used to realize 'accent' among the four tones was not significantly different [Tone: $F(3, 27) = 2.806$, $p=.080$] and the interaction between tone and focus was also insignificant [Tone x Focus: $F(3, 27) = 2.402$, $p = .090$]. These results suggested that an increase in vowel duration was used more frequently in the realization of 'accent' in an unfocused position than in a 'focused' position. That is, regardless of the tone it was produced with, 'accented' vowels in an unfocused position were more frequently longer than 'accented' vowels in a focused position.

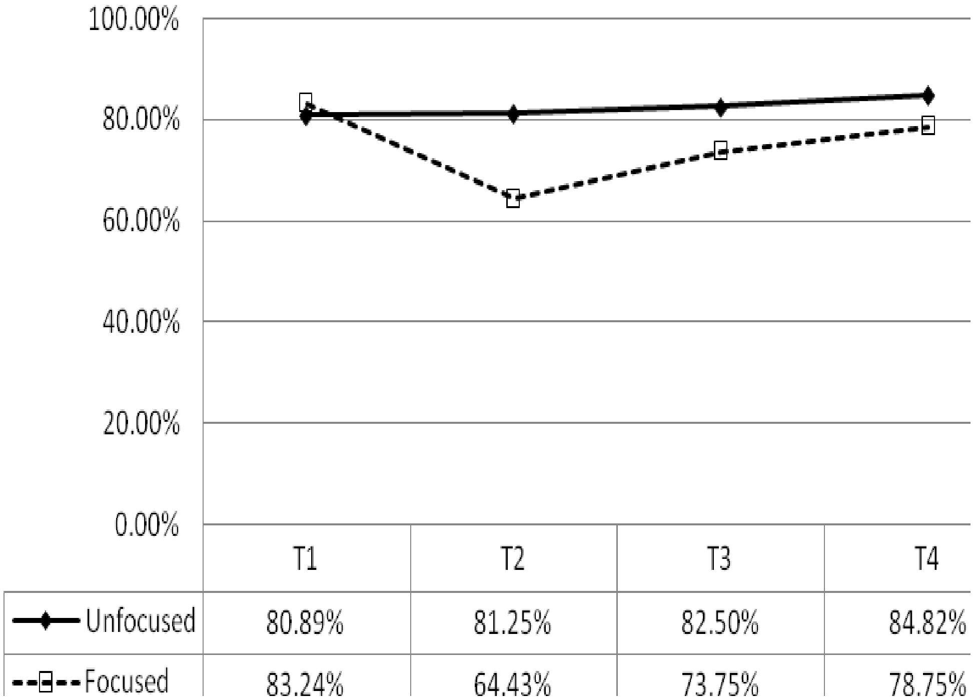


Figure 5-13. Percentages of data using duration as a parameter to realize accent

Ratio data Ratio values in Table 5-8 showed a ratio increase in duration produced by each speaker to implement accent in different sentence positions among tones. Average means and SDs were displayed in Figure 5-14.

Table 5-8. Ratio means and the standard derivations of duration parameter for accent realizations

	Unfocused Position				Focused Position			
Speakers	Tone 1	Tone 2	Tone 3	Tone 4	Tone1	Tone2	Tone3	Tone4
S1	2.03	1.67	2.50	1.85	1.59	1.61	1.80	1.38
S2	2.84	2.34	2.39	1.81	1.68	1.53	2.16	1.81
S3	2.48	3.13	1.57	2.02	1.80	1.59	1.55	1.24
S4	2.73	2.57	2.47	2.54	2.21	2.34	1.91	2.11
S5	2.51	2.79	2.11	2.95	1.99	2.16	1.68	2.12
S6	3.04	2.74	1.95	2.36	2.90	2.86	2.86	1.97
S7	2.43	2.44	2.07	2.11	2.04	2.37	1.97	1.67
S8	2.24	3.13	1.97	1.67	2.02	2.70	1.73	1.35
S9	2.09	2.72	1.56	1.94	1.60	1.78	1.91	1.37
S10	1.80	1.98	2.33	1.42	1.71	2.24	1.88	1.43
Average by Speakers	2.42 (.39)	2.55 (.47)	2.09 (.34)	2.07 (.45)	1.95 (.39)	2.12 (.47)	1.95 (.36)	1.65 (.34)

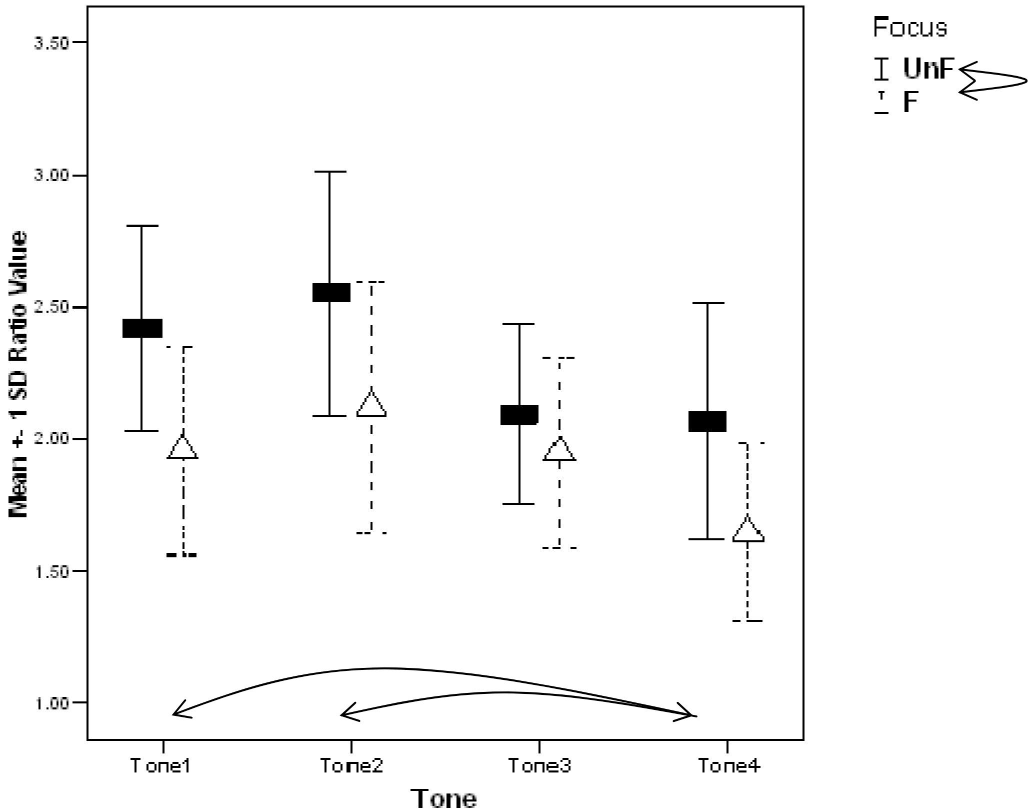


Figure 5-14. Ratio increase of the duration parameter in accent realizations. Arrow indicates a significant difference

Repeated measures ANOVA performed on the data revealed that accented vowels in unfocused positions (shown as filled dark rectangular) had a significantly higher duration ratio than those produced in focused positions (shown as unfilled triangle) [Focus: $F(1, 9) = 20.231$, $p = .001$], and that duration ratio varied significantly among the four tones [Tone: $F(3, 27) = 5.831$, $p = .003$]. Follow up pair-wise comparisons suggested that accented vowels produced with Tone 1 and Tone 2 had a higher duration ratio than Tone 4 [$p = .041, .045$] (shown in Table 5-9). However, no significant interaction was observed between tone and focus [Tone x Focus: $F(3, 27) = 1.487$, $P = .240$]. These results revealed that 'accent' was more effectively realized in unfocused positions than in focused positions. Specifically, duration of accented vowels produced in unfocused positions was lengthened to a significantly greater extent than those produced in focused positions. In addition, averaged across both focused and unfocused conditions, accented vowels produced with Tone 1 and Tone 2 were lengthened to significantly larger extent than those produced with tone 4.

Table 5-9. Pair wise comparisons of ratio means among tones

(I) tone	(J) tone	Mean Difference (I-J)	Std. Error	Sig.(a)
1	2	-0.148	0.104	1.000
	3	0.168	0.103	0.822
	4	.331(*)	0.095	0.041
2	1	0.148	0.104	1.000
	3	0.316	0.154	0.420
	4	.479(*)	0.139	0.045
3	1	-0.168	0.103	0.822
	2	-0.316	0.154	0.420
	4	0.163	0.122	1.000
4	1	-.331(*)	0.095	0.041
	2	-.479(*)	0.139	0.045
	3	-0.163	0.122	1.000

Summary for Research Question 2

Table 5-10. and Table 5-11. summarized the significant results of main factors and their interactions on focus and accent realizations in terms of the frequencies at which acoustics parameters were used to implement prominence and their ratio values.

Table 5-10. Interaction among tone, accent and focus: frequency data

	Focus realizations			Accent realizations		
	Main factors		Interaction	Main factors		Interaction
	Accent	Tone	Accent^Tone	Focus	Tone	Focus^Tone
DUR	*	-------	--------	*	-----	-------
INTEN-MAX	------	------	------			
INTEN-MEAN	-----	-------	------			
F_0-MEAN	*	------	--------			
F_0-MAX	*	*	------			
SLOPE	*	-----	---------			

"*" and "___"indicated significant and insignificant results respectively. The cells in shade were not analyzed.

Table 5-11. Interaction among tone, accent and focus: ratio data

	Focus realizations			Accent realizations		
	Main factors		Interaction	Main factors		Interaction
	Accent	Tone	Accent^Tone	Focus	Tone	Focus^Tone
DUR	*	*	---	*	*-	-----
INTEN-MAX	*	-----	-----			
INTEN-MEAN	*	-----	------			
F_0-MEAN	*	*	---			
F_0-MAX	*	----	*			
SLOPE	*	*	----			

“*” and “___”indicated significant and insignificant results respectively. The cells in shade were not analyzed.

Generally, there were fewer significant results in Table 5-10 than Table 5-11, which indicated that the frequency data was less affected by the interation among focus, accent and tone than the ratio data. In other words, fewer differences were noticed in the percentage of data that made use of a particular parameter than the actual ratio increase in that acoustic dimention to implement prominence.

For focus realization (including both the frequency data and the ratio data), it was affected by both accent and tone categories. Moreover, the effect of accent was observed in more cases than tonal effects. Focus gained significantly higher frequncies and ratios when realized in unaccented positions. The only exception lied on the intensity parameters (i.e., the frequency at which intensity parameters were used to realize focus was not significantly higher in unaccented positions than in accented positions). Regarding the tonal effects on focus realization, they were observed exclusively in duration and F_0 parameters, not in intensity parameters. Tone 4 had significantly higher frequncies and ratios in F_0 parameters (e.g., The frequency at which maximum F_0 was used to implement focus was significantly higher in Tone 4 than Tone2; the ratio increase in mean F_0 and F_0 slope was to significantly larger extent in Tone 4 than in Tone 1 and Tone 2 respectively). Tone 3 had a significantly higher ratio than Tone 1 when increased duration was used to realize focus. Focus realization was seldom affected by the interaction between accent and tone. This suggested that focus realized in accented and unaccented positions seldom varied on the basis of which tone it was assigned to.

Similarly, accent realizations were affected by main factors of focus and tone. There was, however, no interaction between the two factors. Increased duration used to realize accent appeared significantly more frequently and to a larger extent in unfocused positions. Moreover, the ratio of duration lengthening in Tone 1 and Tone 2 was significantly higher than Tone 4.

CHAPTER 6
ACOUSTIC CUES FOR FOCUS PERCEPTION

In this chapter, the perception experiment will be described. The goal of the perception experiment was to investigate, for each lexical tone, how acoustic cues were ranked in terms of their importance in prominence perception. The research question addressed was RQ3: Among acoustic parameters used to produce focus and accent, which ones are used in the perception of prominence?

The experiment focused on the acoustic parameters adopted most frequently (present in 60% or more of the data) in prominence realizations (mentioned in Table 4-4 and Table 4-5, Chapter Four). The acoustic parameters adopted by the same tone were compared in pairs to test native speakers' preferable cues in prominence perception. To operationalize the comparisons, target words were digitally modified with one acoustic parameter fully and exclusively enhanced at a time to signal prominence (i.e., the modification was performed to only one acoustic parameter and shown a full degree of prominence for each token, all other parameters were intact as in non-prominent positions). The modified tokens were then embedded in the sentence frames. In other words, for each token, the original target tone produced in a prominent condition in the production experiment described in Chapter 4 was replaced by its own modified version with only one 'prominent acoustic parameter'. The two tokens in each comparison trial, both having a different acoustic parameter matching the prominent version of the same tone, were played to native Mandarin Chinese listeners for 'preference' (or 'naturalness') judgment.

For example, in the production experiment, it was found that for Tone 1, focus was realized using four acoustic parameters: duration, mean and maximum of intensity, and mean F_0 (Table 4-4, chapter Four). To test relative perceptual importance between duration and mean F_0, the unfocused Tone 1 was modified by increasing its duration to the same length as the focused

Tone 1 counterpart to generate one token for the perception test. The other token was generated by shifting the mean F_0 of unfocused Tone 1 to the same level as the focused Tone1 (so, only one prominent parameter, either duration or F_0, was present in each token). The two tokens were embedded in the same focused position (replacing the original focused Tone 1) to generate two utterances and played to listeners who were asked to decide which modified token they preferred in that focused position.

The cue selected most of a tone was considered the most frequently adopted acoustic cue to perceive prominence for that particular tone. Since duration was the dominant parameter used for accent realization (no comparison could be made with other parameters), the perception experiment was conducted to study focus perception only**.**

The chapter will be organized as follows. First, the design of the perception experiment will be described. In this section, we will explain how target words were modified. Next, results will be presented and analyzed tone by tone to rank acoustic cues in the perception of prominence.

Methods

Subjects

Twenty native speakers of Mandarin Chinese (10 female and 10 male), ages between 25 and 33, participated in this experiment. They were born in Beijing and its neighboring areas (sharing the same Beijing Mandarin Dialect and using Standard Mandarin Chinese for daily communication), and had stayed in the US for less than three years at the time of testing. All reported normal language and speech development and passed a bilateral hearing screen in the range of 250 to 8,000 Hz measuring at 25 dB HL (by DSP Pure Tone Audiometer).

Stimuli

The stimuli used in this experiment were the same disyllabic proper names produced with all possible combination of the four Chinese tones (16 in total) used in the production experiment. Multiple tokens were generated based on each disyllabic word naturally produced in 'unfocused' environment ([-A-F]) by a female speaker using the same recording procedure with a sampling rate of 44.1 kHz and 16-bit PCM as in the production experiment. Digital modification for each token was conducted on only one of the acoustic parameters found in the production experiment to have been used to realize 'focus'.

Table 6-1 (the original Table 4-4 from Chapter Four was repeated here for convenience) showed acoustic parameters used to realize 'focus' for each tone (accounting for over 60% of the data).

Table 6-1. Acoustic parameters for focus realization

Tones	DUR	INTEN-MEAN	INTEN-MAX	F_0-MEAN	F_0-MAX	F_0-MIN	SLOPE
T1	82%	63%	62%	80%			
T2	84%				66%		74%
T3	81%		64%				
T4	86%	65%	70%	85%	88%		90%

As shown in Table 6-1, four parameters were used to signal focus in Tone 1; three parameters in Tone 2, two parameters in Tone 3 and six parameters for Tone 4. Duration was used most frequently to implement focus. Parameters such as F_0 and intensity were present in fewer data. Table 6-2 listed the ranking of these acoustic parameters in focus realization. The ranking was based on the how frequently these acoustic parameters were used (i.e., the percentage data) to realize focus. As could be seen from this table, parameters were ranked decreasingly from Parameter 1 to Parameter 6 (if existed). Duration was ranked the highest, followed by F_0 and intensity parameters in a decreasing order.

Table 6-2. Rank of acoustic parameters in focus realization

Tones	Para 1	Para2	Para 3	Para 4	Para 5	Para 6
T1	DUR	F_0 - MEAN	INTEN- MAX	INTEN - MEAN		
T2	DUR	F_0 SLOPE	F_0-MAX			
T3	DUR	INTEN- MAX				
T4	DUR	F_0-MAX	F_0 SLOPE	F_0- MEAN	INTEN- MAX	INTEN- MEAN

To test the relative perceptual importance of these parameters in each tone, the target disyllabic word 'LiZhi' produced in unfocused positions was modified using the Praat software to generate 15 tokens (four for Tone 1, three for Tone 2, two for Tone 3 and six for Tone 4). For example, to test the perceptual importance of duration, maximum F_0 and F_0 slope in Tone 2, three tokens were generated. The first token, increased the duration to the same length of its 'focused' counterpart without any modifications to F_0-max and F_0 slope (the calculation of how long the modified duration would be was based on the prominence ratio described in Table 5-1, Chapter Five). The second token raised the pitch maximum to the 'focused' level without modifying duration and F_0 slope intentionally (The modification of maximum F_0 value might affect the F_0 slope simultaneously, but the changes in F_0 slope was ignored in this study). The last token increased the F_0 slope without affecting duration and maximum F_0.

There were two issues to be noted. First, the prominence ratio (described in Chapter Five) was generated after the across-talker normalization (mentioned in Chapter Four), where all actual values of unfocused and focused tones were adjusted by the overall speaking rate and vocal F_0 ranges of the sentences they belonged to. To generate the actual value of a modified 'focused' tone in a particular sentence, the prominence ratio was adjusted (or retrieved back) by the difference between the sentence where the unfocused tone was extracted to serve as a basis for modification and the sentence where the modified 'focused' tone replaced the real focused version. An example of duration modification was illustrated in Figure 6-1.

Review: Formula for Duration Ratio mentioned in Chapter 4

$$(4.1)\ \text{Normalized Duration} = \frac{\text{Msec (Target Tone)}}{\text{Speaker's Speaking Rate}}$$

$$(4.2)\ \text{Speaker's Speaking Rate} = \frac{\text{Msec (Sentence)}}{\text{Num. of Syllables}}$$

$$(4.6)\ \text{Duration Prominence Ratio} = \frac{\text{Normalized Duration in Condition P}}{\text{Normalized Duration in Condition NP}}$$

What we had for Duration Modification:

Duration Prominence Ratio: listed in Table 5-1, Chapter Five

Msec (Sentence) in Condition NP: the length of a sentence where the unfocused tone anchored could be measured

Msec (Sentence) in Condition P: the length of a sentence where the real focused tone anchored (and where the modified tone would be embedded) could be measured

Msec (Target Tone) in Condition NP: the length of the unfocused tone could be measured

What we tried to get in Duration Modification:

Msec (Target Tone) in Condition P: the length of the focused tone

---

Figure 6-1. Example of duration modification

When applying (4.2) to (4.1), we got an alternative formula for normalized duration as shown in Figure 6-2.

$$\text{Normalized Duration in Condition P} = \frac{\text{Msec (Target Tone) in Condition P} * \text{Num. of Syllables}}{\text{Msec (Sentence) in Condition P}}$$

$$\text{Normalized Duration in Condition NP} = \frac{\text{Msec (Target Tone) in Condition NP} * \text{Num. of Syllables}}{\text{Msec (Sentence) in Condition NP}}$$

Figure 6-2. Alternative formula for normalized duration

The Duration Prominence in (4.6) was interpreted as follows (in Figure 6-3) after 'Normalized Duration in Condition P' and 'Normalized Duration in Condition NP' were replaced by the alternative formula in Figure 6-2. 'Num. of syllables' in Condition P and NP were omitted since both sentences shared the same number of syllables.

$$\text{Duration Prominence Ratio} = \left(\frac{\text{Msec (Target Tone) in Condition P}}{\text{Msec (Sentence) in Condition P}}\right) \Big/ \left(\frac{\text{Msec (Target Tone) in Condition NP}}{\text{Msec (Sentence) in Condition NP}}\right)$$

Figure 6-3. Alternative formula for duration prominence ratio.

In Figure 6-3, all items were calculated or measured, except the length of focused tone in 'Condition P' (as mentioned in 'What we had for duration modification' and 'What we tried to get for duration modification', Figure 6-1). To obtain duration value of the newly created 'focused' tone [i.e., Msec (Target Tone) in Condition P], the following formula (Figure 6-4 below) was used.

$$\text{Msec (Target Tone) in P} = \text{Duration Ratio*} \ \text{Msec (Target Tone) in NP} * \left(\frac{\text{Msec (Sentence) in P}}{\text{Msec (Sentence) in NP}} \right)$$

Figure 6-4. Formula for duration modification manipulated by prominent ratio

Second, lexical tones used more than one acoustic parameter at a time to realize focus. For example, six parameters were used by Tone 4 (i.e., duration, mean and maximum of intensity and F_0, and F_0 slope). Although not all of these parameters were used every time when focus was realized in Tone 4, it was safe to say (assume) that the majority of parameters were adopted simultaneously for focus realization (because four parameters out of the six were present (used) in more than 80% of the focused Tone 4 data and the other two parameters also occurred in more than 65% of the data). Assuming equal degree of contribution among the six parameters in the focus realization of Tone 4, each parameter would account for one-sixth of the total realized prominence.

An increase in one acoustic parameter for a modified 'focused' tone based on prominence ratio (e.g., duration modification in Figure 6-4) represented the fullest extent that that particular parameter contributed to prominence realization. However, since more than one acoustic parameters were used in the realization of focus in each tone, the fullest contribution of each parameter merely represented a fraction of the total contribution of all parameters combined For example, the duration modification of Tone 4 resulted in an increase in duration of that tone to the same value as that of the original focused version, and represented the full amount of contribution of made by duration (among other cues) to the realization of focus for that tone.

However, since six acoustic parameters were used in focus realization for this tone, the contribution of duration alone would account for only one-sixth of the total amount of prominence realized in original Tone 4. So, a question was raised: would the modified 'focused' tone with only one of the six acoustic parameters approximated its value in the original 'focused' tone stand out from its neighboring context in the utterance and be perceived as prominence? To find an answer to this question, the modified tokens embedded in the focused position of sentences were played to five native speakers in a pilot study to guarantee the tokens sounded as what they labeled: 'focused'. Listeners heard one sentence at a time and judged whether the token in the focused position sounded 'focused/prominent'. The results showed that listeners could not differentiate the modified focused tone from its environment. They commented that the so-called 'focused' tone did not stand out from the sentence, and the prominence was not perceived. The comment indicated that, alone, the contribution of each individual acoustic parameter in focus realization was not sufficient for 'prominence' to be perceived. A realization in just one acoustic dimension could not reflect the magnitude of prominence. It also supported the substantial idea of research question 3 which was the weight difference of each acoustic cue in focus perception presuming more than one acoustic cues would be perceived. In other words, RQ3 focused on relative importance or 'weight' among acoustic cues in focus perception instead of what cues were used for perception (which was a matter of either all or nothing).

Since the modified tokens generated by prominence ratios were not prominent in the utterances where they were embedded into, and the research question aimed at the relative weight among modified tokens instead of the absolute value of the 'prominent' parameter in each token, a weight factor was introduced into the modification process (Figure 6-5 showed the duration modification formula taking the 'weight' factor into consideration). Acoustic parameters used in the same tone were assigned the same weight factor, and the value of the weight factor was equal to the number of parameters used in that tone. For example, the weight factor for Tone 1 was 4, because four acoustic parameters were used to signal focus in Tone 1. Similarly, the 'weight' for Tone 2 was 3, for Tone 3 was 2, and for Tone 4 was 6, because these were the number of parameters in focused Tone 2, Tone 3 and Tone 4 respectively.

$$\text{Msec (Target Tone) in P} = \text{Duration Ratio} * \text{Msec (Sentence) in P} * \left(\frac{\text{Msec (Target Tone) in P}}{\text{Msec (Sentence) in NP}} \right) * \text{Weight}$$

Figure 6-5. Formula for duration modification

The inclusion of the weight factor into the modification process was inspired by the compensatory lengthening, which referred to a set of phonological phenomena wherein the disappearance of one element of a representation is accompanied by a lengthening of another element (Kavitskaya, 2002). For instance, the loss of coda in a closed syllable triggered the lengthening of the vowel in Lithuanian (the 3rd person singular form of 'decide' was [sprɛn-dʒa] and its infinitive form was [spræ;-sti] where the vowel of the first closed syllable was lengthened as a consequence of the loss of the nasal coda [n]. The same was true for the word 'send', where the 3rd person singular form was [sun-tʃɛ] and the infinitive form was [su:-sti]). Therefore, the lengthening was compensatory insofar as it was crucially dependent on the deletion of some element. In other words, either a consonant coda (to form a CVC structure) or a long vowel (to form a CVV structure) served the function of keeping a heavy syllable.

Take Tone 1 as an example, in modified Tone 1 tokens, an increase in the weight of one 'focused' parameter (i.e., duration) also led to the absence of other acoustic parameters (mean intensity, maximum intensity and mean F_0). Hence, if the contribution of four parameters in Tone 1 were the same (e.g., each weighed 1 in the real focused tone), a modified token with one prominent parameter (e.g., duration) needed to quadruple its value to compensate for the absence of the other three parameters. The four modified Tone 1 tokens (each had duration, mean intensity, maximum intensity and mean F_0 modified respectively), though quadruple in their absolute values, still maintained relative values or prominence among each other.

All modified tokens (after the weight adjustment) were embedded in sentences where they replaced the original focused tones and presented to three native speakers of Mandarin Chinese (other than the five listeners in the pilot study or the twenty participants in the perception experiment). The listeners agreed that the stimuli were acceptable exemplars of focus realization.

Procedure

Stimuli were presented binaurally, one at a time over head phones to participants. The participant heard a sequence of two different stimuli A and B with a 1 sec inter-stimulus interval (ISI). 1 sec ISI was adopted based on studies of optimizing measures of perception experiment (Harnsberger et. al, 2004; Wayland et. al, 2004, 2005, 2006). In Harnsberger et al.'s (2004) ASA presentation, they used 1sec ISI for both categorical AXB discrimination test and categorical AX discrimination test. Wayland et. al (2004, 2005) investigated the ability of native English (NE) and native Chinese (NC) speakers to identify and discriminate 'the mid versus the low' tone contrast in Thai before and after auditory training. The variables under investigation were language background and the ISI of the presentation (500 ms vs. 1500 ms). In the NC group, a significant improvement in identification from the pretest to the posttest was observed under both ISI conditions, and the improvement was not significantly different, which suggested that the training procedure was superior to ISI effects in the perception of Thai among Chinese listeners. Their later (2006) study on native Thai speakers' acquisition of English word stress patterns used the longer ISI (1500ms) because the presented stimuli were two sentences.

The modified tones stimuli A and B were always from the same tone category. The stimuli were presented in random order for a total of 125 trails (25 trials * 5 repetitions=125 trials.) The 25 trails included 6 trials for Tone 1 including all possible comparisons between two acoustic parameters out of the four used in focused Tone 1, 3 trials for the three parameters used in Tone2, 1 trial for the two parameters in Tone 3 and 15 trials for the six parameters in Tone 4).The participants was asked to respond which utterance they preferred by clicking a button labeled 'A' or 'B'. They were allowed to replay each trial two times. If they didn't have preference between the two stimuli, they clicked "same' button and the next trial was started. Responses labeled as "same" were omitted from analysis. This amounted to 5.24% of the data.

Results and Analyses

In this section, results of the perception experiment described above will be presented. As mentioned, the experiment was conducted to address the third research question: Among acoustic parameters used to produce focus, which ones are used in the perception of prominence?

Research Question 3: Among Acoustic Parameters Used to Produce Focus, Which Ones are Used in the Perception of Prominence?

Tone 1

Four acoustic parameters were used to signal focus for Tone 1: lengthening the duration, increasing the mean and the maximum values of intensity, and raising the mean value of F_0. Numerically, the modified Tone 1 with maximum intensity as the only cue was preferred most frequently in focus perception (used in 77.20% of the data). Duration was the second important cue in Tone 1 focus perception (present in 54.74% of the data), followed by mean intensity (43.79%) and mean pitch (24.27%) (shown in Table 6-3 and Figure 6-6).

Table 6-3. Descriptive analysis of acoustic cues used in focus perception for Tone 1

Cues	Mean (%)	Std. Deviation
Duration (Dur)	54.74	22.71
Mean Intensity (Inten-mean)	43.79	19.56
Max Intensity (Inten-max)	77.20	19.41
Mean Pitch (Pitch-mean)	24.27	23.00

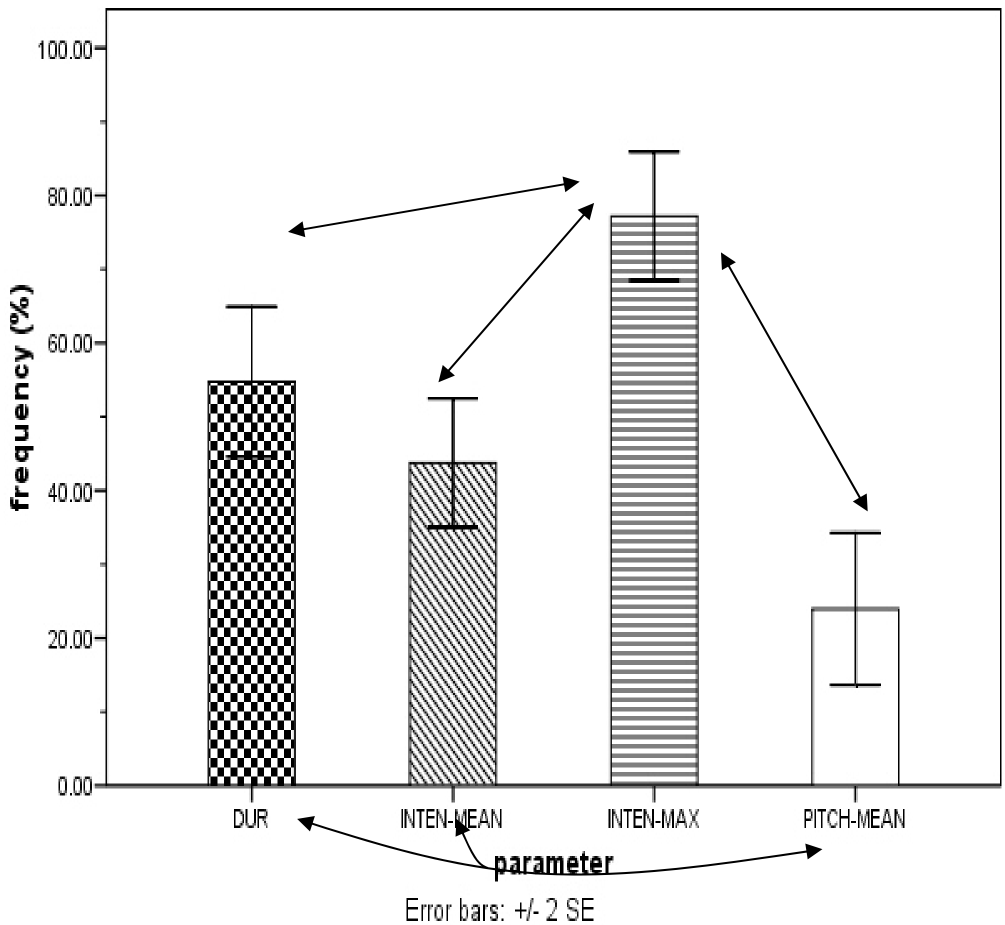


Figure 6-6. Acoustic cues (and their frequencies) used in focus perception for Tone 1. Arrow indicates significant difference

These data were submitted to a repeated-measure ANOVA with acoustic cues as the within-subject factor (shown in Figure 6-6). The results suggested that with an alpha level of .05, the frequency at which each acoustic cue was preferred in prominent perception was significantly different from one another [$F(3, 57) = 16.411$, $P = .000$]. Follow-up pairwise (2 tailed) T tests were conducted. The results suggested that maximum intensity was more frequently preferred to perceive focus in Tone 1 than other cues [t (19) = 2.877, p =.010 between maximum intensity and duration; t (19) = 5.011, p =.000 (2-tailed) between maximum and mean intensity; t (19) = 7.014, p =.000 (2-tailed) between maximum intensity and mean pitch]. Duration and mean intensity were weighted significantly more heavily than mean pitch to perceive focus [t (19) = 3.608, p =.002 between mean pitch and duration; t (19) = 2.514, p =.021 between mean pitch and mean intensity]. No significant difference was observed between duration and mean intensity [t (19) = 1.385, p =.182].

Tone 2

Three acoustic cues were used for focus perception in Tone 2: duration, maximum pitch and pitch slope. Among them, the duration cue was selected in 87.81% of the data to perceive focus, numerically more than the two pitch cues (which was chosen in 37.42% and 24.77% of the data) (shown in Table 6-4).

Table 6-4. Descriptive analysis of acoustic cues used in focus perception for Tone 2

Cues	Mean (%)	Std. Deviation
Duration (Dur)	87.81	17.39
Max Pitch (Pitch-max)	37.42	17.38
Pitch Slope (Slope)	27.77	20.25

The repeated-measures ANOVA showed that with an alpha level of .05, the difference among cues was statistically significant [F (2, 38) =46.608, P=.000] (shown in Figure 6-7.). Results of follow-up pair-wise comparisons revealed that the differences between duration and pitch cues were significant [t (19) = 8.038, p =.000 between duration and maximum pitch; t (19) = 8.690, p =.000 between duration and slope]. However, the difference between maximum pitch and pitch slope was not significant [t (19) = 2.032, p =.056].

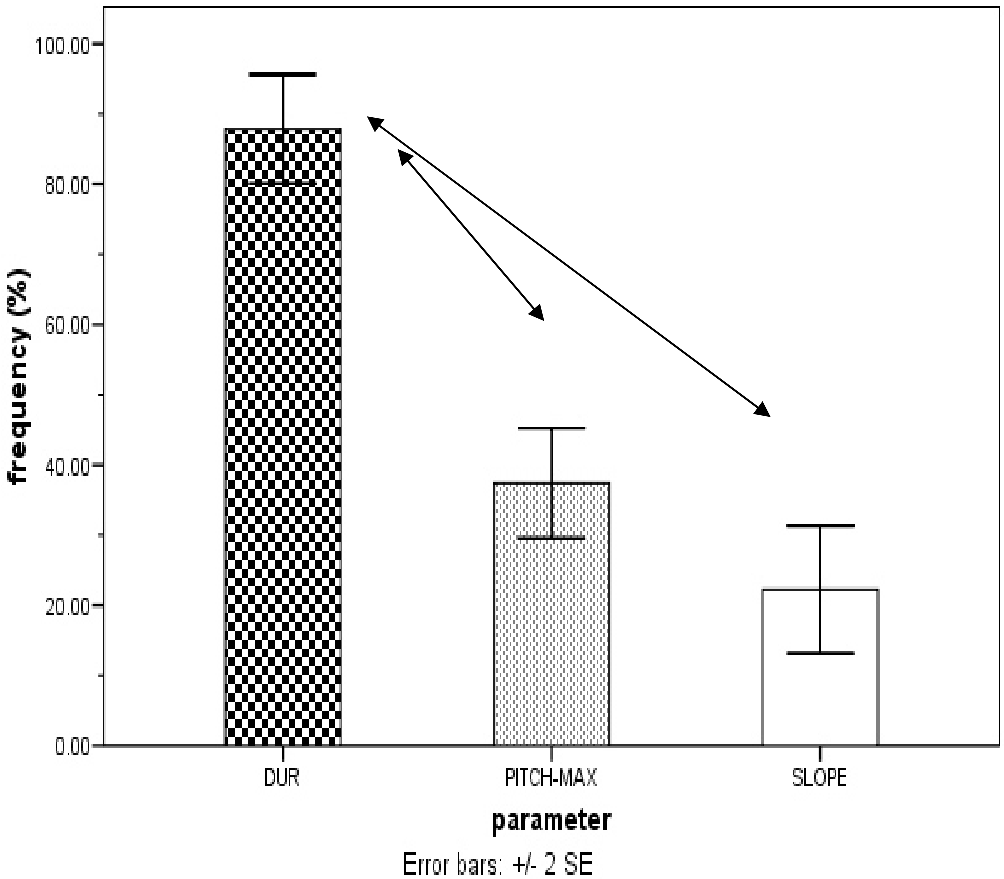


Figure 6-7. Acousitc cues (and their frequencies) used in focus perception for Tone 2.Arrow indicates significant difference

Tone 3

Among the two acoustic cues for focus perception in Tone 3, duration was selected in 67.67% of the data, significantly more preferred than maximum intensity (which was chosen in 32.33%of the data) to perceive focused Tone 3 [$F(1, 19)=5.665$, $P=.028$] (shown in Table 6-5 and Figure 6-8).

Table 6-5. Descriptive analysis of acoustic cues used in focus perception for Tone 3

Cues	Mean (%)	Std. Deviation
Duration (Dur)	67.67	33.19
Max Intensity (Inten-max)	32.33	33.19

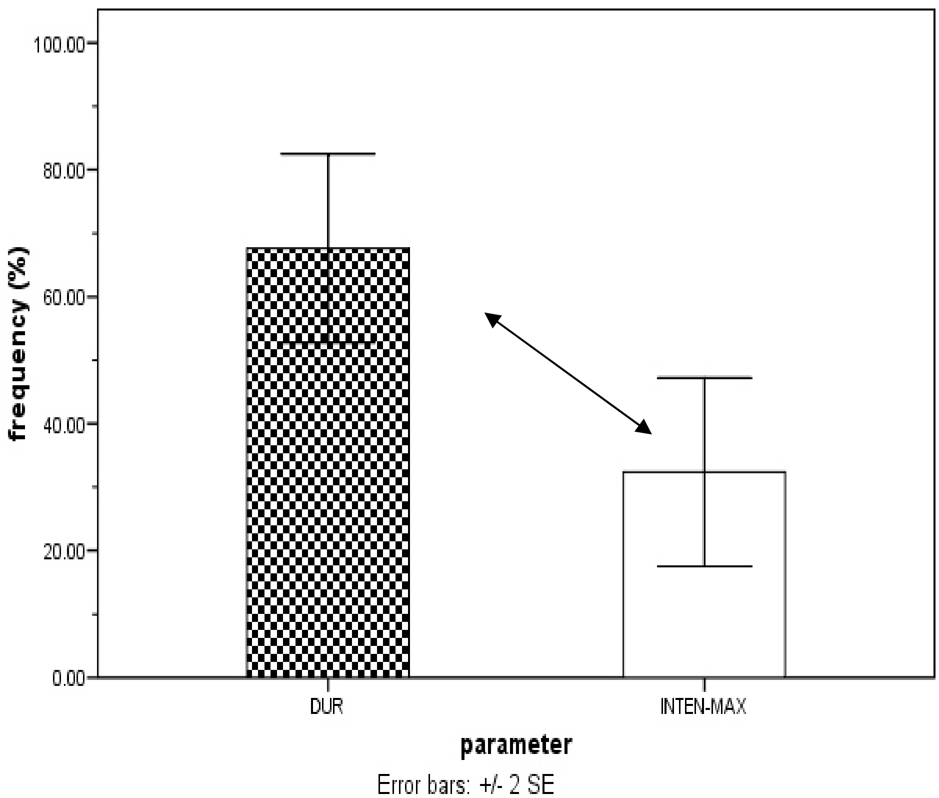


Figure 6-8. Acousitc cues (and their frequencies) used in focus perception for Tone 3. Arrow indicates significant difference

Tone 4

Six acoustic parameters were used to implement focus in Tone 4. From Table 6-6, the intensity cues were selected more frequently to perceive focus than other cues (i.e., maximum and mean intensity were selected in 85.78% and 69.71% of the data, more frequently than duration and pitch cues which were chosen in 66.80%, 36.93%, 21.94% and 18.93% of the data respectively).

Table 6-6. Descriptive analysis of acoustic cues used in focus perception for Tone 4

Cues	Mean (%)	Std. Deviation
Duration (Dur)	66.80	15.99
Mean Intensity (Inten-mean)	69.71	14.16
Max Intensity (Inten-max)	85.78	23.63
Mean Pitch (Pitch-mean)	36.84	7.25
Max Pitch (Pitch-max)	18.93	13.45
Pitch Slope (Slope)	21.94	11.98

The repeated-measures ANOVA suggested that the difference in frequencies at which acoustic cues were selected to perceive focus were significant [F (5, 95) =56.401, P=.000] (shown in Figure 6-9).

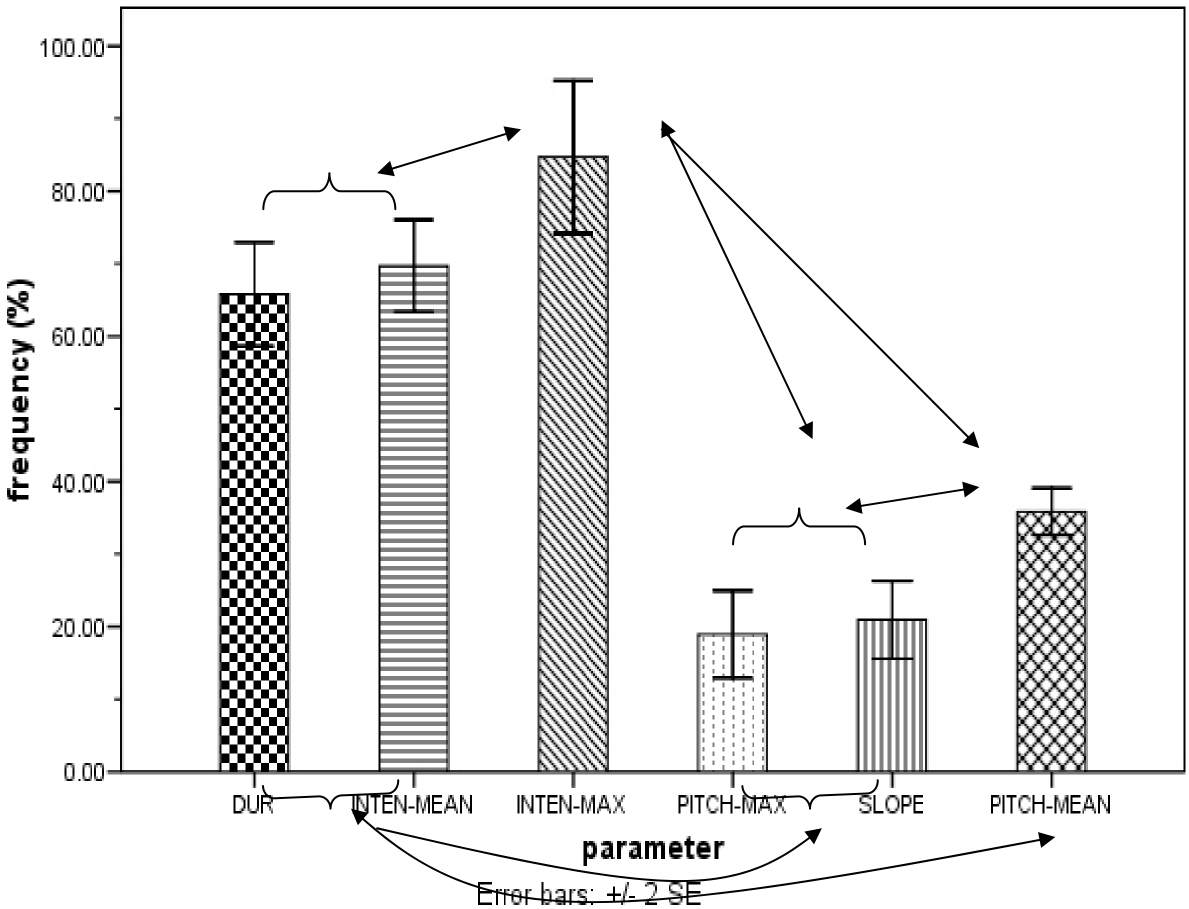


Figure 6-9. Acousitc cues (and their frequencies) used in focus perception for Tone 4. Arrow indicates significant difference

Follow-up t-tests illustrated that maximum intensity was significantly more frequently selected to perceive focus than other cues [t(19) = 2.281, p = .034 between maximum intensity and duration; t(19) = 5.708, p = .000 between maximum and mean intensity; t(19) = 8.841, p = .000 between maximum intensity and mean pitch; t(19) = 8.692, p = .000 between maximum intensity and slope; t(19) = 8.682, p = .000 between maximum intensity and maximum pitch]. Duration and mean intensity were also selected significantly more frequently than mean pitch [t

(19) = 7.937, p = .000 between duration and mean pitch; t (19) = 8.874, p = .000 between mean intensity and pitch], but the difference between duration and mean intensity was not significant [t (19) = .620, p = .543]. Moreover, the mean pitch was also significantly higher than the rest of the pitch cues (i.e., slope, maximum pitch) [t (19) = 4.227, p = .000 between mean pitch and slope; t (19) = 4.200, p = .000 between mean and maximum pitch]. However, no significant difference was observed between pitch slope and maximum pitch [t (19) = 1.290 p = .213].

Summary of Research Question 3

Acoustic cues were differentially ranked in terms of how frequently listeners selected them to perceive focus. For focus in Tone 1 and Tone 4, listeners preferred maximum intensity to perceive focus, followed by duration and mean intensity cues and made the least use of pitch cues. Among the pitch cues in Tone 4, mean pitch was preferred than maximum pitch and pitch slope in focus perception. For focused Tone 2, consistent results with Tone 1 and Tone 4 were found that the ranking of duration was significantly higher than pitch cues (i.e., maximum pitch and pitch slope) to perceive focus. For focus perception in Tone 3, the result was different from Tone 1 and Tone 4, and listeners preferred duration to maximum intensity in focus perception. Since the modified tokens for intensity cues did not completely separate from each other (i.e. modification in mean intensity could not avoid affecting maximum intensity, and vice versa), it was too early claim that intensity cues were more/less preferred than duration in focus perception based on the current results (e.g., no preference between duration and mean intensity in Tone 1 and Tone 4; maximum intensity was preferred (than duration) in Tone 1 and Tone 4, but duration was preferred (than maximum intensity) in Tone 3), but it was safe to conclude that duration and intensity cues in general were more preferred than pitch cues in focus perception.

CHAPTER 7
GENERAL DISCUSSION AND CONCLUSIONS

In this chapter, I will first summarize the results of the production and perception experiments to answer the three research questions I proposed. Next, results obtained from this current study will be discussed and compared with those found in previous relevant studies. In this section, the mismatches between acoustic parameters used to signal prominence and the cues in perception will be presented. Explanations in terms of trading relations will be provided for the perception results. Phonological account of prominence realization will be proposed under tone geometry and OT frameworks. Finally limitations and potential directions for future exploration will be addressed.

Summary of Results

In this dissertation, I have investigated linguistic prominence caused by accent and/or focus in the environment of longer utterances to examine the interactions among tone, accent and focus in Mandarin Chinese in seven acoustic dimensions: duration, mean intensity, maximum intensity, mean F_0, maximum F_0, minimum F_0, and F_0 slope. Research questions 1 and 2 were addressed -in a production experiment designed to study the acoustic parameters used to signal prominence manifested as ‘focus’ and ‘accent’, and the interactions among tone, focus and accent in prominence realization. Research question 3 was addressed in a follow-up perception study of focus designed to explore relative importance among acoustic cues in focus perception.

Summary for Research Question 1: What are the Acoustic Parameters Used to Realize Focus and Accent among Lexical Tones of Mandarin Chinese?

In chapter Four, I have demonstrated with a production experiment that focus and accent differed in terms of the number of acoustic parameters used in their realizations. Focus, in general, was mainly realized by duration lengthening, F_0 slope sharpening, as well as an increase in mean and maximum of intensity and F_0 (i.e., these parameters were used in more than 60% of the data to realize focus and appeared significantly more frequently than other parameters). In other words, focus realization made use of all acoustic parameters measured except the minimum F_0. Accent was produced mainly with an increase in duration, though increase in intensity and F_0 was also observed in a small proportion of data to realize accent.

However, it was found that different tones used different acoustic parameters to realize focus. For Tone 1(the level tone) and Tone 4 (the falling tone), duration, intensity and F_0 parameters were used. Specifically, Tone 1 made use of four acoustic parameters while Tone 4 used six: an increase in duration, mean and maximum intensity was observed for both Tone 1 and Tone 4. Besides, focused Tone 4 was also implemented by an increase in mean F_0, maximum F_0 and F_0 slope, while focused Tone 1 only exhibited an increase in mean F_0. For Tone 2 (the rising tone) and Tone 3 (the dipping tone), it was found that an increase in duration was used in focus realization. In addition, an increase in maximum F_0 and F_0 slope was also found in focused Tone 2, while an increase of maximum intensity was found in Tone 3. In sum, duration was found to have been used in focus implementation in all lexical tones. F_0 and intensity parameters were used in some tones, but not in others (e.g., F_0 parameters were adopted by Tone 1, Tone 2 and Tone 4, but not Tone 3; intensity parameters were adopted by Tone 1, Tone 3 and Tone 4, but not Tone 2).

The difference in frequencies at which these acoustic parameters were used to implement focus by a particular tone was not significant. In other words, the parameters used in more than 60% of the data for a focused tone did not differentiate each other in terms of their frequencies. Specifically, duration, mean and maximum intensity, mean F_0 were used in more than 60% of the data to realize focus in Tone 1, and the frequency differences among them were not significant. Duration, maximum F_0 and F_0 slope were used significantly more frequently than intensity parameters in focused Tone 2, but the frequencies at which duration and F_0 parameters were used did not differ significantly among each other. The same was true for Tone 3 and Tone 4. For example, duration and intensity were used significantly more frequently than F_0 in the manifestation of focus for Tone 3, but no significant frequency difference was observed between the main parameters (i.e., duration and intensity).

To explore accent realization, an increase in duration was used to realize accent in all lexical tones. For each particular tone, duration was used significantly more frequently than other acoustic parameters measured.

Summary for Research Question 2: What are the Interactions among Tone, Accent and Focus in the Realization of Focus and Accent?

In chapter Five, I have demonstrated that there were interactions among tone, accent and focus when they were realized cocurrently. The interactions were explored in two ways: in terms of how often they were used to implement prominence (i.e., the percentage of data showing an increase in a particular acoustic dimension) and the extent of the increase (i.e., the ratio between non- prominent and prominent conditions).

Generally speaking, how frequent a parameter was used to signal prominence (showing an increase in an acoustic parameter) was less affected by the interactions, while more interactions among focus, accent and tone were revealed regarding the the extent of the increase (the ratio of an increase in a particular acoustic dimension). In other words, for an acoustic parameter used to realize focus and accent, more differences were observed in ‘the extent of its increase’ than in ‘the frequency’ it was used in promience realization.

Focus realization was significantly affected by accent and tonal categories. Moreover, effects of accent was greater than that of tones. Among acoustic parameters used to realize focus, most of them were adopted more frequently when focused tones were realized in unaccented positions than in accented positions. The only exception lied in the intensity parameters whose frequency was not significantly differerent for focus realized in unaccented and accented positions. Similarly, the extent of the increase was also significantly higher in unaccented positions, which indicated more increase (or modifications) in acoustic parameters when focus was realized in unaccented positions than accented positions. In other words, focus was more fully realized in unaccented positions(than accented positions): acoustic parameters were more frequently used; and the increase in these paramters displayed a greater extent. Regarding tone effects on focus realization, singnificant differences among tones were observed in duration and F_0 parameters, but not in intensity parameters. Both Tone 4 and Tone 2 made use of maximum F_0 to realize focus, but maximum F_0 was used more frequently in Tone 4 than Tone 2. Tone 4 also exhibited a greater extent of the increase than Tone 1 and Tone 2 in mean F_0 and F_0 slope respectively. The extent of duration lengthening was significantly higher in Tone 3 than in Tone 1. Although the implementation of focus varied as a fuction of accent and tones, no interaction between tone and accent was found in the realization of focuswhich implied that focus realized in accented and unaccented positions seldom varied on the basis of which tone it was assigned to and vice versa.

Accented was realized in the same fashion as focus. It was significantly affected by focus and tone, but no the interaction between the two factors on accent realization was found. Specially, duration was used more frequently and with greater exent of modification when accent was realized in unfocused positions than in focused positions. The extent of duration lengthening was significantly higher in Tone 1 and Tone 2 than in Tone 4.

Summary for Research Question 3: Among Acoustic Parameters used to Produce Focus, Which Ones are Used in the Focus Perception?

In chapter Six, I have demonstrated with a perception experiment that acoustic parameters used for focus realization were differentially ranked in focus perception (Since duration was the only parameter used in more than 60% of the data to realize accent, no perceptual ranking was generated). Overall, duration and intensity cues were ranked significantly higher than pitch cues among all tones, which suggested that duration and intensity cues were used more often than pitch cues in focus perception. To be more specific, in Tone 1 and Tone 4, listeners most frequently used maximum intensity cue to perceive focus, followed by duration and mean intensity cues and made the least use of pitch cues. A consistent result was found in Tone 2, where listeners preferred the duration cue to pitch cues to perceive focus. Duration was also preferred in Tone 3, when compared to the maximum intensity cue.

General Discussion

New Findings

The results found in this dissertation were consistent with previous studies regarding general realizations of prominence in Mandarin Chinese in three respects: (i) F_0, duration and intensity were used together to realize focus; (ii) Changes in F_0 were mainly observed in Tone 1, Tone 2 and Tone 4, but not Tone 3 (i.e., Tone 1 raised the mean F_0, Tone 2 and Tone 4 raised maximum F_0). (iii) Focus was more fully realized without the presence of accent. Besides, the following findings were first addressed from the production and perception experiments conducted in this study:

- Focus realization made use of more acoustic parameters (including different facets of duration, F_0 and intensity) than accent (which was realized mainly by duration lengthening).
- Accent was also more fully realized without the presence of focus.

- Lexical tones differed in terms of acoustic parameters signaling prominence.
- For an acoustic parameter adopted by more than one lexical tone, tones differed in terms of how often that parameter was adopted to signal prominence (i.e., the percentage of data showing modifications in that acoustic dimension) and the extent of the modification (i.e., the ratio between non- prominent and prominent conditions in that parameter).
- Acoustic cues used for focus perception were not ranked in a same fashion as in focus realization. Duration and intensity cues were selected more frequently than pitch cues in focus perception, while duration, F_0 and intensity parameters were equally important in production.

Mismatches between Realization and Perception of Focus

Comparing results generated from RQ1 and RQ3, I argued that there existed mismatches between acoustic parameters used in focus realization and the cues for perception. In focus realization, no significant difference was observed among duration, F_0 and intensity (used in more than 60% of the data) to implement focus. In other words, for acoustic parameters used in a majority of the data, their frequencies were not significant different. In focus perception, however, duration and intensity cues were ranked significantly higher than pitch cues. Specifically, listeners preferred duration and intensity to pitch cues to perceive focus in Tone 1, Tone 2 and Tone 4. A comparison between focus realization and perception suggested that duration and intensity were important for both focus realization and perception, while F_0 parameters were only primary for focus realization.

The results were consistent with previous literature on prominence in Mandarin Chinese (mentioned in Chapter Three) that tones were modified in F_0, duration and intensity parameters to realize a prominent syllable (Chen, 2004; Hsu, 2006; Jin, 1996; Shen, 1985; Shih, 1988; Tseng, 1988; Xu, 1999, 2004; Yip, 1993), while duration and intensity cues were sufficient in (word-level) prominence perception (Shen, 1993).

Cross linguistically, the results were also consistent with Gussenhoven and Blom's (1978) proposal in their study about perception of prominence by Dutch listeners that the acoustic parameters measured in speech production were not necessarily perceptual cues for listeners. Many studies argued that pitch was more often adopted in speech production, while intensity in perception.

For example, Erber and Witt (1977) investigated effects of stimulus intensity on speech perception by deaf children. They presented monosyllabic, trochaic (disyllabic words with stress syllable followed by unstressed syllable), and spondaic (disyllabic words with two stressed syllables) words to profoundly (over 95 dB HTL) hearing-impaired children at sensation levels (SL) ranging from near detection to near discomfort. The result showed that the profoundly deaf children's stress pattern perception improved as a function of increasing intensity. In some cases, the maximum perception was obtained at the highest intensity level that the children would tolerate.

Studies on normal hearing subjects also demonstrated an important role of intensity in speech perception. Tanner and Rivette (1964) compared the efficiency of human observers in amplitude-discrimination tasks to their efficiency in frequency-discrimination tasks. The behavior of one of the four observers suggested that he was completely insensitive to frequency differences, while he could distinguish amplitude differences. A language background check indicated he was a native speaker of Punjabi, a language with lexical tones. Therefore the authors quoted Liberman et al.'s (1957, 1961) hypothesis that observers were less efficient at discriminating differences that occurred within the same phoneme, and proposed that the results reflected a cultural–bound condition, which in this case, was the phonemic function of pitch. Lehiste and Fox (1992) investigated perception of prominence by Estonian and English Listeners in both speech and nonspeech materials. In their study, stimuli were lengthened to 425, 450, or 500 msec and/or increased in amplitude by 3 or 6dB. The subjects were asked to indicate which token in each trial was 'most prominent'. The results showed that, for English-speaking listeners, amplitude cues overrode duration cues to perceive word prominence. Also, Vainio and Jarvikivi (2006) explored tonal features, intensity, and word order in the perception of prominence in Finnish. Listeners judged the relative prominence of two consecutive nouns in a three-word utterance, where the accentuation of the nouns was systematically varied. Intensity was found to affect the perception judgment. The study suggested that lowering the intensity of the accented word led to fewer responses to sentence stress on the last word.

Trading Relations in Focus Perception

A trading relation (or perceptual equivalence) was described as 'when two or more cues contribute to a given phonetic distinction, they can be traded against each other' (Repp, 1982). In other words, the acoustic cues were perceptually equivalent. For example, In Fitch et. al's (1980) study, they investigated the perceptual equivalence of two acoustic cues (i.e. silent closure duration and vocalic formant transition onsets) for stop manner in the 'slit'-'split' distinction. In a phonetic identification task, they synthesized stimuli that consisted of an [s]- like noise, followed by a variable amount of silence (cue 1), and then by either of two vocalic syllables [lit] or [plit] which were modified only to have formant onset differences (cue 2). The result showed the [p] stop preferred long silence and low formant onset frequency. As the silence was longer, less low onset was needed to hear the stop, and similarly when the onset was lower, less silence was needed. Hence, there was a trading relation (an equivalence in perception) between silence and the formant onset for stop distinction.

There were three explanations accounted for trading relations from auditory, phonetic and informational perspectives respectively. An auditory explanation relied on a description of the way auditory system processed the sound, regardless of whether or not the sound was perceived as speech. The process could either be cues integrated into a unitary auditory percept at an early stage in perception (the auditory integration hypothesis), or some kind of functional interaction at 'higher' levels (the auditory interaction hypothesis, which argued that the selective attention was directed to one of the cues, and the perception of that cue was affected by the setting of other cues) (Blumstein & Stevens, 1979, 1980; Ganong, 1978; Pastore, 1981; Stevens & Blumstein, 1978).

The auditory terms had problem to explain why trading relations only occur in stimuli from phonetic boundary regions, and disappeared when listeners tried to discriminate stimuli that unambiguously belong to the same phonetic category (Best, 1981; Fujisaki & Kawashima, 1969, 1970; Hodgson & Miller, 1996; Repp, 1982, 1983). A phonetic explanation was provided that speech was produced by a vocal tract, and the production of a phonetic segment had complex and temporally distributed acoustic consequences. Therefore, the information supporting the perception of the same phonetic segment was acoustically diverse and spread our over time. Listeners recovered the abstract units of speech by integrating the multiple cues that resulted from their production. The basis for the perceptual integration was conceptualized in a way that listeners knew from experience what a given phonetic segment 'ought' to sound 'like' in a given context. Insofar as phonetic contrasts involved more than one acoustic parameter, trading relations among these parameters resulted when the stimulus was ambiguous because it was being evaluated with reference to idealized representations or 'prototypes': a 'conflicting' change in one parameter could be offset (or compensated) by a 'cooperating' change in another so that the perceptual distances from the prototypes remain constant.

The phonetic account also had its own problem to explain why intensity, a phonetically irrelevant cue for the presence vs. absence of a stop, participated in a trading relation that was supposed to be a byproduct of phonetic categorization (Wright, 1993). In an informational explanation, the increase in sensitivity at the crossover point or the boundary region was due to subject uncertainty at the point where the signal produced an equally ‘good’ (or ‘bad’) fit for mental representation at the either side of the boundary. Therefore, variations in the signal that were not phonetically relevant could be involved in trading relations if they did heighten the uncertainty of a particular feature. A continuous value between 0 and 1 was assigned to an acoustic cue depending on the perceptual system’s certainty of the cue being present in the signal. The greater the certainty, the higher the value was. To achieve a greater certainty of a signal, when the value of one acoustic cue was lowered, other cues tried to make compensation by increasing their certainty values. For example, when a stimulus that had a quiet burst also had an adequately long preceding duration of silence, it could still be an equally good fit to a stored representation of the phoneme /p/ as a stimulus with a loud burst but a shorter preceding silent duration. Thus the manipulation of the burst could be compensated for by equivalent manipulation of the silence duration. The informational explanation of trading relation also allowed an acoustic cue to affect the certainty of a particular signal, but exerted no effects to the perception of other signals. Moreover, it extended the study of trading relations to domains larger than a single sound, for example, to intonation. McRoberts et.al. (1995) investigated fundamental frequency (F0) of the voice under two conditions. In one condition, F_0 was used to convey linguistic distinction (a Y/N question vs. a statement distinction), and in the other condition it was used to affective distinction (a positive affect vs. negative affect distinction). The results claimed that a trading relation was obtained between F_0 peak and terminal rise when F_0 was used to convey Y/N question intonation: a significant negative correlation was found between stressed-syllable peak F_0 and the amount of final rise for questions produced. However, no trading relation was found when F_0 was used to express emotions.

In this study, trading relations were found in the modified ‘prominent’ tokens in focus perception. A real natural focused tone used more than one acoustic parameter to signal prominence (in Chapter Four). In the perception experiment, each modified token had only one ‘prominent’ parameter fully realized (in Chapter Six). To compensate the disappearance of other acoustic cues, the single ‘prominent’ cue in a modified token was weighed more heavily than that in a real focused token. For example, a real focused Tone 3 lengthened the duration, and increased the intensity to signal its prominence. The two modified focused Tone 3 had an increase in duration but not in intensity in one token, and an increase in intensity but not in duration in the other token. The increase in duration or intensity in modified tokens was much greater than that in real focused tokens to achieve the same perceptual prominence judged by native listeners. Thus perceptual equivalence showed among real focused Tone 3 with an increase in both duration and intensity cues, and modified ‘single –cue’ Tone 3 with much greater extent either to lengthen duration or to increase intensity.

The results of the perception experiment also supported informational explanation of trading relations in two ways. First, duration and intensity were involved in trading relations when focus was realized. A stimulus with longer duration could compensate the absence of intensity increase and a stimulus with greater intensity could have no duration lengthening to be good fits to a representation of focus. Listeners preferred intensity and duration cues in focus perception and the prominence conveyed by greater intensity or longer duration alone could be equivalent to the saliency of multiple cues in a real focused tone. The phenomena could be explained in terms of informational module as listeners’ certainty of duration and intensity cues in prominent or focused tones, which was a subjectively derived description in memory through experience with the native language. It could not be explained by the phonetic module, because the two acoustic parameters were not phonetically significant in Mandarin Chinese (i.e., differences in duration and intensity were not gesturally relevant. Neither were they used to distinguish phonemes in the language). Second, the informational explanation of trading relations allowed an acoustic cue to affect perception of some signals, but not others. In Mandarin Chinese, pitch was an important cue in tone perception, but it was not preferred in focus perception. In tonal perception, pitch height and contour were primary cues to distinguish tones in the system. However, in focus perception, increase in pitch cues could not counterbalance the absence of other acoustic cues. As a result, the modified 'focused' token with only pitch cues were not selected as prominent as other tokens with modifications on duration and intensity. The fewer effects of pitch in focus perception did not exclude the possibility that pitch was perceived in listeners' auditory system. A possible explanation could be found in listeners' (un)certainty of pitch cues in focus perception. It was likely that pitch played such an important role in tonal perception that listeners became less sensitive when pitch played other roles or functions.

Phonological Implications of Prominence Realization

From the 'summary of results' section earlier in this chapter, it was concluded that focus realization was signaled by six acoustic parameters: duration, the mean and the maximum values of intensity and F_0, and the F_0 slope; while accent was mostly realized by duration (RQ1 in Chapter Four). Regarding the interaction among tone, accent and focus, focus and accent were significantly affected by each other when they coincided, but the difference of focus realization in accented and unaccented positions (or the difference of accent realization in focused and unfocused positions) seldom varied on the basis of which tone it was assigned to (RQ2 in Chapter Five). Given these conclusions, how could the prominence realized in target words be modeled in Mandarin Chinese? I proposed a suprasegmental account (shown in Figure 7-1) where focus was manifested via the phonetic encoding of the segmental contents (focused tones were fully implemented in its F_0) and suprasegmental contents (focused tones had an increase in duration and an optional increase in intensity), and accent was phonetically encoded with suprasegmental contents (accented tones were lengthened). The suprasegmental account was consistent with findings from the focus perception experiment (RQ3 in Chapter Six) that duration and intensity was more preferred to perceive focus. A possible explanation was that listeners preferred to use suprasegmental codes (which were duration and intensity) to perceive focus (or information in larger domains, such as sentences), while keeping segmental codes (which were pitch cues) to perceive tones (or local information).

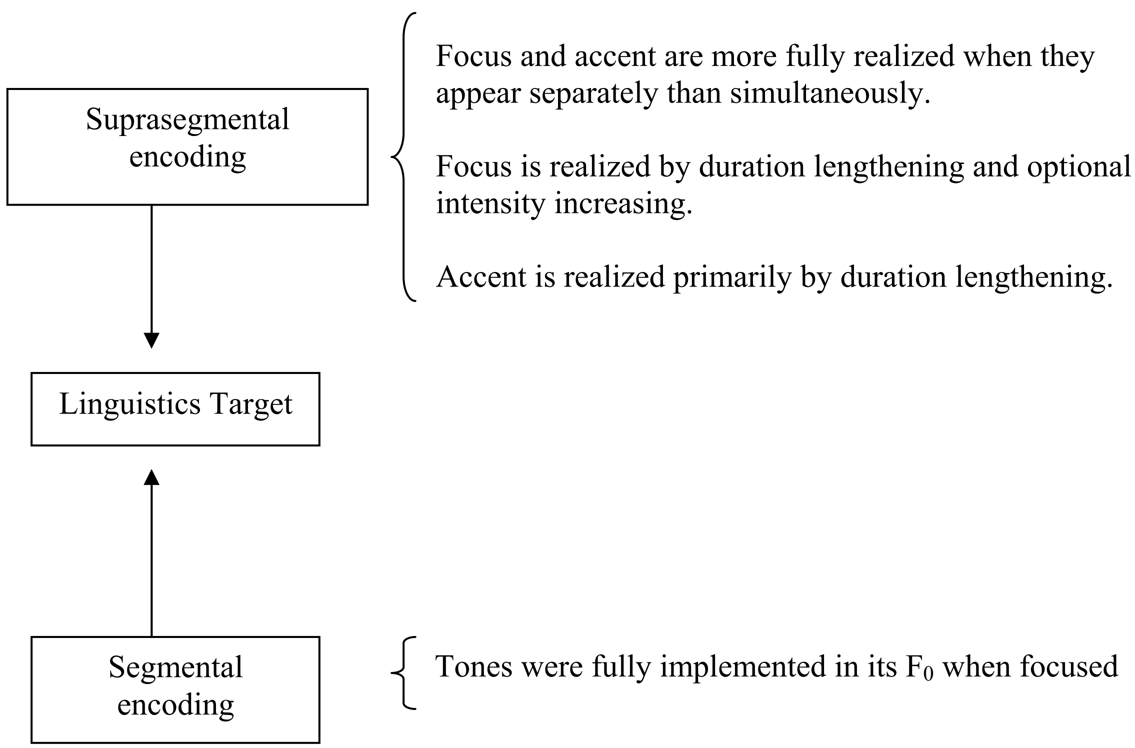
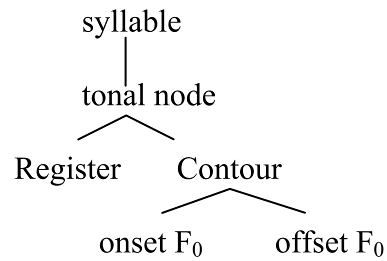
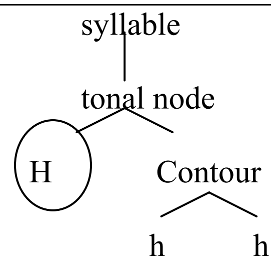
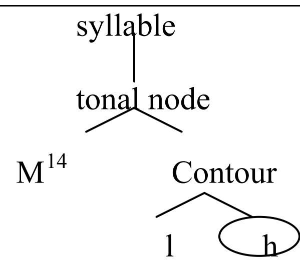
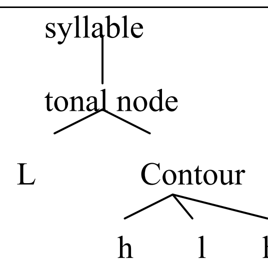
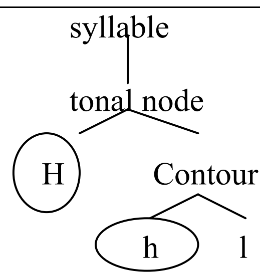


Figure 7-1. Suprasegmental account for prominence realization in Mandarin Chinese

Table 7-1. Tone geometry model used to explain focus realization among lexical tones

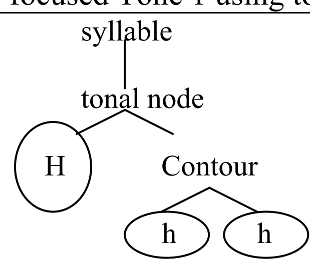
Focus realization	Tone geometry explanation
	
Focused Tone 1: raise mean F_0	
Focused Tone 2: raise max F_0 and F_0 slope	
Focused Tone 3: no changes	
Focused Tone 4: raise mean F_0, max F_0 and slope F_0	

[14] Tone 2 in Mandarin Chinese is a Mid-high rising tone (labeled as '35' in Chao's five-scale system). Its register is different from Tone 1 (labeled as '55') and Tone 4 (labeled as '51'). In many phonological descriptions, its register was labeled as 'H' (a high tone) and considered as a rising tone starting from lower F0 in the Higher register and raised to a higher F0. These descriptions had no problem to distinguish Tone 2 from other lexical tones in Mandarin Chinese, because no other tones had the same contour as Tone 2. However, from a phonetic- based point of view, the register of Tone 2 is [-high, -low] in Woo's system or [+central] in Sampson's system. Register 'M' is used in Table 7-1 to emphasize its register difference from high tones: Tone 1 and Tone 4.

To further discuss the full implementation of lexical tones when focused, I summarized the modifications in F_0 among focused tones. Tone 1 raised the overall F_0 mean, Tone 2 raised the maximum F_0 and changed F_0 slope, Tone 3 didn't have F_0 modifications when focused, Tone 4 raised the overall F_0 mean, as well as maximum F_0 and F_0 slope.

To explain the phenomena in terms of tone geometry, Tone 1 made changes in its register without affecting contour, Tone 2 made changes in its contour without affecting register, Tone 3 was intact, and Tone 4 changed in both register and contour. Table 7-1 provided explanations for focus realization in Mandarin Chinese with Bao (1999)'s tone geometry model (shown in Table 2-7, Chapter Two), since it was the only model where the contour and the register could change independently. Inside the table, the 'H' register nodes in Tone 1 and Tone 4 were affected by changes in mean F_0. It was likely that the 'h' F_0 values in the contour node of Tone 1 were also raised, but the raise in both 'h's didn't change the level contour (as shown in Table 7-2). The contour node in Tone 2 and Tone 4 was affected by changing maximum 'h' values and retaining the 'l' values, and changes in F_0 slope could be considered as a result of contour modification.

Table 7-2. Alternative explanation for focused Tone 1 using tone geometry model

Focused Tone 1: raise mean F_0	

From Table 7-1 and Table 7-2, it was noticed that 'H' register and 'h' F_0 value attracted focus, while 'L' registered tone rejected focus. To explain these findings using constraints in OT treatment, the faithfulness and markedness constraints were described as follows:

- IDENT-T: Correspondent tones are the same.
- *Low tone/F: Focus is not realized in Low tone

- *High tone/F: Focus is not realized in High tone
- *Low tone/UF: Non-Focus is not realized in Low tone
- *High tone/UF: Non-Focus is not realized in High tone

The markedness constraints listed above were relevant to the tonal node. ‘Focus prefers High tone and avoids Low tone’ was achieved by ranking *Low tone/F higher than *High tone/F (i.e., *Low tone/F>> *High tone/F), and ‘Non-Focus prefers Low tone and avoids High tone’ was achieved by ranking *High tone/UF higher than *Low tone/UF (i.e., *High tone/UF>> *Low tone/UF).

- *L, l/F: Focus is not realized in Low register or low F_0
- *H, h/F: Focus is not realized in High register or high F_0
- *L, l/UF: Non-Focus is not realized in Low register or low F_0
- *H, h/UF: Non-Focus is not realized in High register or high F_0

This part of markedness constraints were related to terminal tonal features. ‘Focus prefers High register and high F_0, and avoids Low register and low F_0’ were described as ranking *L, l/F higher than *H, h/F (i.e., *L, l/F>>*H, h/F), and ‘Non-Focus prefers Low register and low F_0, and avoids High register and high F_0’ were described as ranking * H, h/UF higher than * L, l/UF (i.e., * H, h/UF >>* L, l/UF).

OT tableaux (shown from Table 7-3 to Table 7-6) indicated focus implementation on the segmental level. The winner candidates among all lexical tones violated *High tone/F, *Low tone/UF, *H, h/ F, and * L, l/ UF, so these constraints were ranked lowest. Similarly, all of them satisfied IDENT-T, so this constraint was ranked highest. The constraints for the tonal node (*Low tone/F, *High tone L, l/ F) were ranked higher than constraints for terminal tonal features (*L, l/F, *H, h/ UF). One explanation for this ranking was that non-focus in low Tone 3 was realized with an ‘h’ feature, which indicated that the constraint * H, h/ UF (i.e., Non-Focus is not realized in High register or high F_0) was violated to satisfy the constraint *Low tone/F (i.e., Focus is not realized in Low tone).

Table 7-3. OT treatment for Tone 1 focus realization

Tone 1 H(hh) F	IDENT -T	*Low tone/F	*High tone/UF	*L, l/ F	* H, h/ UF	*High tone/F	*Low tone/UF	*H,h/ F	* L, l/ UF
☞H (h h) F						*		***	
H (h h) F					**!				
L(h h) F	*!								

Table 7-4. OT treatment for Tone 2 focus realization

Tone 2 M(l h) F	IDENT -T	*Low tone/F	*High tone/UF	*L, l/ F	* H, h/ UF	*High tone/F	*Low tone/UF	*H,h/ F	* L, l/ UF
☞M (l h) F								*	*
H (l h) F	*!								
M (l h) F				*!				*	

Table 7-5. OT treatment for Tone 3 focus realization

Tone 3 L (h l h) F	IDENT -T	*Low tone/F	*High tone/UF	*L, l/ F	* H, h/ UF	*High tone/F	*Low tone/UF	*H,h/ F	* L, l/ UF
☞L (h l h) F					**		*		**
L (h l h) F		*!							
H (h l h) F	*!								

Table 7-6. OT treatment for Tone 4 focus realization

Tone 4 H (h l) F	IDENT -T	*Low tone/F	*High tone/UF	*L, l/ F	* H, h/ UF	*High tone/F	*Low tone/UF	*H,h/ F	* L, l/ UF
☞H (h l) F						*		**	*
H (h l) F				*!					
H (h l) F			*!						

Future Directions

There is space for improvement for this study. One improvement for future studies would be to include more tokens and subjects for both production and perception experiments to increase the reliability of the statistical analysis. In other words, a higher variability in stimuli and subjects will enhance conclusions concerning the focus and accent realizations, and the focus perception among lexical tones. Secondly, in order to further testify the mismatches between focus production and perception, the modified single-cue tokens could be more separately from each other. Methods, for example to modify F_0 and intensity maximum without affecting other F_0 and intensity cues respectively, need to be proposed. Moreover, current tokens were embedded to sentences through concatenation technique using Praat software without smoothing the transitions when connected. More natural speech synthesis methods shall be applied to enhance listeners' natural perceptual behaviour.

Besides, some results generated in this study need further exploration. The trading relations among cues in focus perception were implied in a pilot study in this dissertation, both identification and discrimination tasks could be incorporated in a focus perception experiment to examine the relations among perceptual cues and find justifications for current trading relation modules. Also, the competition between accent and focus observed in the production experiment needs more investigation. Duration is the major parameter used in the manifestation of accent, however focus implemented with F_0 and intensity parameters is less realized when appearing together with accent. Questions are left open such as 'what are the explanations for the less realized focus implemented with F_0 and intensity parameters (in accented positions)?' 'Is it because of F_0 and intensity parameters used in a small percentage of data to realize accent?' 'Is it because of the interaction among acoustic parameters (i.e., duration, F_0 and intensity) used in focus realization (i.e., when duration lengthening is not fully realized to implement focus in accented positions, other parameters also become less effective in focus realization)?' Moreover, studies using other methodologies could be conducted to study prominence in Mandarin Chinese, such as ERP studies on brain activities when prominence is perceived attentively or inattentively.

The study could also be expanded to non-native speakers of Mandarin Chinese. Pitch was considered as a less frequently used acoustic cue to perceive focus in Chinese among native speakers. Was it universal among all speakers or caused by Chinese speakers' tonal language background? Speakers with different language backgrounds might adopt different acoustic dimensions in their focus perception. For example, Min speakers had a significantly greater maximum range of speaking intensity than Mandarin speakers; while both Mandarin and Min speakers had a greater maximum range of speaking F_0 and intensity than English speakers (Chen, 2005). German speakers used pitch cues to perceive focal accent (Batliner, 1991), while Estonian and Swedish speakers were more responsive to duration cues than amplitude cues to perceive English prominence (Lehiste & Fox, 1992, 1993). Hence, studies can be conducted among speakers of different language groups (such as English speakers without any tonal background and Thai speakers with similar tonal system to Chinese) to investigate their perception of Chinese prominence.

Similarly, studies could also aim at second language learners of Mandarin Chinese to find possible influences of their native languages and Chinese proficiency levels on the perception and production of prominence in Mandarin Chinese. Short-term, as well as long term training effects could be included to explore possible changes from acoustic parameters used in native languages for prominence production and perception among L2 learners gradually to more native-like ones in Chinese.

In a conclusion, results from the study not only provide insight into the understanding of prominence realization and perception among native speakers in Mandarin Chinese, but also provide valuable information in pedagogical domains. In Chinese L2 teaching, language teachers can use such information in their teaching methodology, such as how to make emphases or focus on important contents in the classroom. Teachers also need to be aware of students' language background differences in the perception of such emphases or focus, instead of assuming what has been emphasized is perceivable by all students. Moreover, Chinese prosody should also be taught intentionally inside the class. Lexical tones, though very important, is not the whole part of spoken Chinese. Currently, L2 Chinese teachers have made a lot of effort in the accurate pronunciation of isolated words or syllables with correct tones. However, to make expressions and deliver information, isolated words need to be combined to larger speech domains such as sentences and paragraphs, and intonation is an indispensable part at this level. Thus, more listening activities regarding Chinese prosody could serve as an input in the beginning level class, and more speaking tasks can be added from the intermediate level class when learners are able to produce sentence-length utterances.

LIST OF REFERENCES

Akinlabi, A., & Liberman, M. (2000). The tonal phonology of Yoruba Clitics. In B. Gerlach and J. Grijzenhout (Eds.), *Clitics in Phonology, Morphology and* Syntax (pp. 31-62). Amsterdam: Benjamins.

Archangeli, D., & Langendoen, T. (Eds.). (1997). *Optimal theory: An overview*. Oxford: Oxford University Press.

Bao, M-Z., Chu, M., and Wang, Y. J (2007). The influence of reading styles on accent assignment in Mandarin. *Computational Linguistics and Chinese Language Processing, 12(1),* 91-106.

Bao, Z. (1990). *On the nature of tone*. Ph.D. dissertation, MIT.

Bao, Z. (1999). *The structure of tone*. Oxford: Oxford University Press.

Batliner, A. (1991). Deciding upon the relevancy of intonational features for the marking of mocus: a statistical approach. *Journal of Semantics, 8(3),* 171-189.

Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht: Foris publications.

Beckman, M. E. (2006). Tone inventories and tune-text alignments. Paper presented at the annual meeting of the Society for Pidgin and Creole Linguistics, Albuquerque, 6-7 January 2006.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics, 29(3),* 191-211.

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production. *Journal of the Acoustical Society of America, 66*, 1001-1017.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environment, *Journal of the Acoustical Society of America, 67,* 648-662.

Boersma, P., & Weenink D. (2004). *Praat: a system for doing phonetics by computer*. Amsterdam: Institute of Phonetic Sciences of the University of Amsterdam.

Brown, K. (1980). Grammatical incoherence. In H. W. Dechert & M. Raupach (Eds.), *Temporal Variables in Speech*. The Hague: Mouton.

Buekers, R. & Kingma, H. (1997). Impact of phonation intensity upon pitch during speaking: a quantitative study in normal subjects. *Logopedics Phoniatrics Vocology, 22,* 71—77.

Buring, D. (1997). *The meaning of topic and focus The 59th Street Bridge Accent.* London and New York: Routledge Studies in German Linguistics.

Cao, J. F. (1995). Basic temporal structure of a sentence in Standard Chinese. *Journal of Chinese Linguistics, 7*.

Cao, J. F. (1999). Acoustic-phonetic characteristics on the rhythm of Standard Chinese. In *the Proceedings of 4th National Conference on Modern Phonetics.* Beijing, August 25-27.

Cao, J. F. (2004). Restudy of segmental lengthening in Mandarin Chinese. In *the proceedings of Speech Prosody 2004*. Nara, Japan. March 23-26.

Cao, J. F. (2004). Tonal aspects in spoken Chinese: Global and local perspectives. Paper presented at the International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages. Beijing, 28-30 March 2004.

Cao, J. F., Lv, S. N., & Yang, Y. F. (2000). Prosody and a proposed phonetic model. *Report of Phonetic Research 2000*, 27-31.

Cassimjee, F., & Kisseberth, C. W. (1998). Optimality domains theory and Bantu tonology: a case study from Isixhosa and Shingazidja. In L. M.Hyman and C. W. Kisseberth (Eds.), *Theoretical Aspects of Bantu Tone* (pp. 33-132). Stanford, Calif.: CSLI.

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

Chao, Y.R. (1930). A system of tone letters. *Le maître phonétique, 45,* 24-27.

Chen, H. (2004). Tone and prominence in Standard Chinese. Paper presented at the International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages. Beijing, 28-30 March 2004.

Chen, S. H. (2005). The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects. *The Journal of the Acoustical Society of America, 117(5),* 3225-3230.

Chomsky, N. (1971). Deep structure, surface structure, and semantic interpretation. In D. Steinberd & L. Jakobovis (Eds.), *Semantics-an Interdisciplinary Reader in Philosophy, Linguistics and Psychology.* Cambridge: CUP.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Chu, M.& Bao, M-Z. (2004). Comparison of sentential-stress allocation within base phrases among different reading styles. Paper presented in Speech Prosody 2004. Nara, Japan. March 23-26.

Chu, M., Wang, Y.J., & He, L. (2003). Labeling stress in continuous Mandarin speech perceptually. In *the Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, Spain, August 3-9.

Clements, G. N. (1981). The hierarchical representation of tone features. In I. R. Dihoff (Ed.), *Current Approaches to African Linguistics* (pp. 145-176). Dordrecht: Foris.

Cruttenden, A. (1986). *Intonation.* Cambridge: Cambridge University Press.

Culter, A. & Ladd, D. R. (1983). Prosody: models and measurements. Springer-Verlag Berlin Heidelberg.

de Lacy, P. (1999). *Tone and prominence*. MS, University of Massachusetts, Amherst. ROA#333.

Deng, D., Chen, M., & Lu, S.N. (2004). Study on stress models of Chinese disyllable. Paper presented in the International Symposium on Tonal Aspects of Languages with Emphasis on Tone Languages. Beijing, March 28-30.

Dogil, G. (1999). The phonetic manifestation of word stress. In H. van der Hulst (Ed.), *Word Prosodic Systems in the Languages of Europe* (pp. 273-334) Berlin: de Gruyter.

Downing, L. J. (2003). Stress, tone and focus in Chichewa and Xhosa. In R. Anyanwu (Ed.), *Stress and Tone – the African Experience. Frankfurter Afrikanistische Blätter, 15,* 59-81.

Drubig, H. B. & Schaffar, W. (2001). Focus construction. In M. Haspelmath et, al (Eds.), *Language Typology and Language Universal*. Berlin: Walter de Gruyter.

Duanmu, S. (1990). *A formal study of syllable, tone, stress and domain in Chinese languages*. Ph. D. dissertation. M.I.T.

Duanmu, S. (1994). Against contour tone. *Linguistic Inquiry, 25*, 555-608.

Duanmu, S. (1999). Stress and the development of disyllabic words in Chinese. *Diachronica, 16(1),* 1-35.

Duanmu, S. (2000). *The phonology of Standard Chinese*. Oxford: Oxford University Press.

Duanmu, S. (2004). Left-headed feet and phrasal stress in Chinese. *Cahiers de linguistique Asie Orientale, 33 (1)*, 65-103.

Duanmu, S. (2006). Chinese (Mandarin): phonology. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (2nd ed.) (pp. 351-355). Oxford, UK: Elsevier Publishing House.

Erber, N. P., & Witt, L. H. (1977). Effects of stimulus intensity on speech perception by deaf children. Journal of Speech and Hearing Disorders, 42(2), 271-278.

Face, T. L. (2001). Focus and early peak alignment in Spanish intonation. *Probus, 13,* 223-246.

Féry, C., & Samek-Lodovici, C. (2006). Focus projection and prosodic prominence in nested foci. *Language, 82 (1),* 131-150.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception and Psychophysics, 27(4),* 343-350.

Fox, A. (2000). *Prosodic features and prosodic structures: the phonology. of suprasegmentals.* Oxford: Oxford University Press.

Frota, S. (2000) *Prosody and focus in European Portuguese: Phonological phrasing and intonation*. New York: Garland Publishing, Inc.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1,* 126-152.

Fujisaki, H., & Kawashima, T. (1969). On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute, 28*, 67-73.

Fujisaki, H., & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute, 29*, 207-214.

Gandour, J. (1981). Perceptual dimensions of tone: evidence from Cantonese. *Journal of Chinese Linguistics, 9(1),* 20-36.

Gandour, J. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics, 12(2),* 235-261.

Ganong, W. F. (1978). The selective adaptation effects of burst-cued stops. *Perception and Psychophysics, 24*, 71-83.

Garde, P. (1968). *L'Accent.* Paris: Presses Universitaires de France.

Garding, E. (1983). A generative model of intonation. In A. Cutler, & D. R. Ladd, (Eds.), *Prosody: Models and Measurements* (pp. 11-25). Springer-Verlag Berlin Heidelberg.

Gordon, M. (2005). An autosegmental/metrical model of Chickasaw intonation. In S-A. Jun (Ed.), Prosodic Typology: The Phonology of Intonation and Phrasing (pp. 301-330). Oxford: Oxford University Press.

Gruber, J. (1964). *The distinctive features of tone*. Manuscript.

Gussenhoven, C. (1983). Focus, mode, and the nucleus. *Journal of Linguistics, 19*, 377-417.

Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.

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

Gussenhoven, C., & Blom, J. G. (1978). Perception of prominence by Dutch listeners. *Phonetica, 35(4),* 216-230.

Guthrie, M. & Carrington, J. F. (1988). Lingala: Grammar and dictionary. London: Baptist Missionary Society.

Halford, B. K. & Pilch, H. (1994). *Intonation*. Tübingen: Gunter Narr Verlag Tubingen.

Harnsberger, J. D., Yeon, S.-H., & Silver, J. (2004). Optimizing measures of the perceptual assimilation of stop consonants. Presented at the 148th Meeting of the Acoustical Society of America, San Diego, November 15-19.

He, Y., & Jin, S. (1992). Intonations of Beijing dialect: an experimental exploration. *Yuyan Jiaoxue Yu Yanjiu, 2*, 71-96.

Hockett, C. (1955). A manual of phonology. *International Journal of American Linguistics Memoir 11*. Baltimore: Waverly Press.

Hockett, C. (1958). *A course in modern linguistics*. New York: MacMillan.

Hodgson, P., & Miller, J. L. (1996). Internal structure of phonetic categories: Evidence for within-category trading relations. *The Journal of the Acoustical Society of America, 100(1),* 565-576.

Hombert, J. M., Ohala J., & Ewan, W. (1979). Phonetic explanations for the development of tones. *Language, 55*, 37-58.

Hsu, H. C. (2006). Revisiting tone and prominence in Chinese. *Language and Linguistics, 7(1),* 109-137.

Hyman, L. (1993). Register tones and tonal geometry. In H. van der Hulst, & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 75-108). Berlin: Mouton de Gruyter.

Hyman, L. (2006). Word-prosodic typology. *Phonology, 23*, 225-57.

Jin, S. (1996). *An acoustic study of sentence stress in Mandarin Chinese*. Ph. D. dissertation, Ohio State University.

Johnston, H. M. (2005). The influence of frequency and intensity patterns on the perception of pitch. Unpublished dissertation.

Jones, D. (1950). *The phoneme: Its natural and use*. Cambridge: W. Heffner and Sons.

Kager, René. (1999). *Optimality theory*. Cambridge: Cambridge University Press.

Kavitskaya, D. (2002). *Compensatory lengthening: phonetics, phonology, diachrony.* New York: Routledge.

Ke, J. Y., Ogura, M., & Wang, W. S-Y. (2003). Optimization models of sound systems using genetic algorithms. *Computational Linguistics, 29 (1),* 1-18.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics, 35(1),* 104-117.

King, P. H. (1995). *Configuring topic and focus in Russian*. Stanford: CSLI Publications.

Kiss, K. (1995). Introduction. In K. Kiss (Ed.), *Discourse Configurational Languages*. New York, Oxford: Oxford University Press.

Komiyama, S., Watanabe, H., & Ryu, S. (1984). Phonetographic relationship between pitch and intensity of the human voice. *Folia Phoniat, 36,* 1–7.

Ladd, R. D. (1980). *The structure of intonational meaning: evidence from English*. Bloomington: Indiana University Press.

Ladd, R. D. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Ladefoged, P. (2000). *A course in phonetics*. (4th ed.). Thomson Wadsworth.

Leben, W. R., Inkelas, S., & Cobler, M. (1989) Phrases and Phrase Tones in Hausa. In P. Newman and R. Botne (Eds.) *Current Approaches to African Linguistics* (pp. 45–61). Dordrecht: Foris.

Lehiste, I (1970). Suprasegmentals. Cambridge, MA: MIT press.

Lehiste, I., & Fox, R. A. (1992). Perception of prominence by Estonian and English listeners. *Language and Speech, 35(4),* 419-434.

Lehiste, I., & Fox, R. A. (1993). Influence of duration and amplitude on the perception of prominence by Swedish listeners. Speech Communication, 13, 149-154.

Liao, R. (1994). *Pitch contour formation in Mandarin Chinese: A Study of Tone and Intonation.* Ph.D. dissertation, Ohio State University.

Liberman, A. M., Harris, K. S., Eimas, P., Lisker, L., & Bastlan, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech, 4,* 175-195.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358-368.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry 8 (2)*, 249-336.

Lin, M. C., Yan, J. Z., & Sun, G. H. (1984). A primary experiment on the stress pattern of normal disyllabic words in Mandarin. *Dialect, 1,* 57-73.

Liu, F. and Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica, 62,* 70-87.

Luksaneeyanawin, S. (1993). Thai. In D. Hirst, & A. Di Cristo (Eds.), *Intonation Systems* (pp. 376-94). Cambridge: Cambridge University Press.

Luo, C., & Wang, J.(1957). *The outline of phonetics of Standard Chinese*. Beijing: Sciences Publish House.

Martinet, A. (1954). Accent et tons. *Miscellanea phonetica, 2*, 13-24.

Mckie, M. (1996). Semantic Rhyme: A Reappraisal. *Essays in Criticism*, *46*, 340-58.

McRoberts, G. W., Studdert-Kennedy, M., & Shankweiler, D. P. (1995). The role of fundamental frequency in signaling linguistic stress and affect: Evidence for a dissociation. *Perception and Psychophysics, 57(2),* 159-174.

Merala, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception and Psychophysics, 48 (2),* 169-178.

Moore, C.B. (1993). Some observations on tones and stress in Mandarin Chinese. *Working Papers of the Cornell Phonetics Laboratory, 8*, 82-117.

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

Myers, S. (1997). OCP effects in optimality theory. *Natural Language and. Linguistic Theory, 15*, 847-892.

Newman. S. (1946). On the stress system of English. *Word, 2*,171-187.

Odden, D. (1995). Tone: African languages. In J. Goldsmith (Ed.), *Handbook of Phonological Theory* (pp. 444-75). Oxford: Blackwell.

Ohala, J.J. (1978). Production of tone. In V.A. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 5-40). New York: Academic Press.

Pastore, R. E. (1981). Possible psychoacoustic factors ion speech perception. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the Study of Speech*. Hillsdale, NJ: Erlbaum.

Pike, E. V. (1974). A multiple stress system versus a tone system. *International Journal of American Linguistics, 40*,169-175.

Pike, K. L. (1948). *Tone languages*. Ann Arbor: University of Michigan Press.

Potisuk, S., Gandour, J., & Harper, M. P. (1996). Acoustic correlates of stress in Thai. *Phonetica, 53(4),* 200-220.

Prince, A., & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar*. Rutgers University Center for Cognitive Science Technical Report 2.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92(1),* 81-110.

Repp, B. H. (1983). Categorical perception: issue, methods, findings. In N. J. Lass (Ed.), *Speech and Language: Advances in Theory and Practice*. New York: Academic Press.

Rosen, S. M. (1977). Speech perception and speech synthesis: the effect of fundamental frequency patterns on perceived duration. *Speech Transmission Laboratory Quarterly Progress and Status Report, 1,* 17-30.

Samek-Lodovici, V. (2005). Prosody syntax interaction in the expression of focus. *Natural Language and Linguistic Theory, 23*, 687-755.

Schmerling, S. (1976). *Aspects of English sentence stress*. Austin: University of Texas Press.

Selkirk, E. (2002). Contrastive FOCUS vs. presentational focus: Prosodic evidence from right node raising in English. *Speech Prosody 2002: Proceedings of the 1st International Conference on Speech Prosody*, 643-646.

Shen, J. (1985). Tonal register and intonation of the Beijing dialect. In *Collection of Experiments on Beijing Phonetics*. Beijing: Beijing University Press.

Shen, T. (1981). Tone sandhi in old Shanghai. *Fangyan, 2,* 131-144.

Shen, X. N. (1990). *The prosody of Mandarin Chinese*. Berkeley: University of California Press.

Shen, X. N. (1993). Relative duration as a perceptual cue to stress in Mandarin. *Language and Speech, 36(4),* 415-433.

Shih, C. (1986). *The prosodic domain of tone sandhi in Chinese*. Ph.D. dissertation, University of California, San Diego.

Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory, 3*, 83-109.

Silverman, D. (1997). Tone sandhi in Comaltepec Chinantec. *Language, 73 (3),* 473-492.

Snider, K. (1999). *The geometry and features of tone*. Dallas: SIL and University of Texas, Arlington.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America, 64*, 1358-1368.

Sun, S. H. (1997). *The development of a lexical tone phonology in American adult learners of Standard Mandarin Chinese*. Honolulu: University of Hawaii Press.

Surendran, D., Levow, G.-A., & Xu, Y. (2005). Tone recognition in Mandarin using focus. In *the Proceedings of Interspeech 2005*. Lisbon, Portugal, September 4-8.

Sweet, H. (1906). *A primer of phonetics*. Oxford: Claredon Press.

Tanner, W. P., & Rivette, G. L. (1964). Experimental study of `tone deafness'. *Journal of the Acoustical Society of America, 36*, 1465-1467.

Tekman, H. G. (1995). Cue trading in the perception of rhythmic structure. *Music Perception, 13,* 17–38.

Tekman, H. G. (1997). Interactions of perceived intensity, duration, and pitch in pure tone sequences. *Music Perception, 14*, 281–294.

Terken, J. (1994). Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America, 95(6),* 3662-3665.

Thompson, L. (1987). A Vietnamese reference grammar. Hawaii: University of Hawaii.

Trager, G. L. (1941). The theory of accentual systems. In L. Spier (Ed.), *Language, Culture, and Personality* (pp. 131-45). Menasha, WI: Sapir Memorial Publications Fund.

Tseng, C. (1981). *An acoustic phonetic study on tones in Mandarin Chinese.* Ph.D. dissertation, Brown University.

Tseng, C. (1988). Some stress related acoustic features of disyllabic words in Mandarin Chinese. *Bulletin of the Institute of History and Philology Academia Sinica, 59 (3),* 577-615.

Vainio, M., & Jarvikivi, J. (2006). Tonal features, intensity, and word order in the perception of prominence. *Journal of Phonetics, 34(3),* 319-342.

Wang, S. Y. (1967). Phonological features of tone. *International Journal of American Linguistics, 33*(2), 93-105.

Wang, Y.J., Chu, M., & He, L. (2003). Pilot study of semantic stresses in Mandarin. In *the Proceedings of the 6th National Conference of Modern Phonetics*. Tianjing, China, October 18-20.

Wang, Y.J., Chu, M., He, L. (2003). Location of sentence stresses within disyllabic words in Mandarin. In *the Proceedings of the 15th International Congress of Phonetic Sciences.* Barcelona, Spain, August 3-9.

Waterson, N. (1976). Perception and production in the acquisition of phonology. *Neurolinguistics, 5,* 294-322.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning, 54 (4),* 681-712.

Wayland, R. P., Guion, S. G., Landfair, D., & Li, B. (2006). Native Thai speakers' acquisition of English word stress patterns. *Journal of Psycholinguistic Research, 35 (3),* 285-304.

Wayland, R. P.,& Li, B.(2005).Training native Chinese and native English listeners to perceive Thai tones. Presented at the ISCA Workshop on Plasticity in Speech Perception. London, June 15-17.

Wayland, R.P., & Guion, S.G. (2003). Perceptual discrimination of Thai tones by naïve and experienced learners of Thai. *Applied Psycholinguistics, 24*, 113-129.

Woo, N. (1969). *Prosody and phonology*. Ph.D. dissertation, MIT.

Wright, R. (1993). Trading relations and informational models. *University of California Working Papers in Phonetics, 83,* 75-95.

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

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics, 27*, 55-105.

Xu, Y. (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics, 5,* 757-97.

Yip, M. (1980). *The tonal phonology of Chinese*. Ph.D. dissertation, MIT.

Yip, M. (1982). Against a segmental analysis of Zahao and Thai: A laryngeal tier proposal. *Linguistic Analysis, 9*, 47-57.

Yip, M. (1989). Contour tones. *Phonology, 6(1),* 149-174.

Yip, M. (1993). Tonal register in East Asian languages. In H. van der Hulst, & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register*. Berlin: Mouton de Gruyter.

Yip, M. (1995). Tone in East Asian languages. In J. Goldsmith, (Ed.), *Handbook of Phonological Theory* (pp. 476-494). Oxford: Basil Blackwell

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Yuan, J.(2005).*Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling.* Ph.D. dissertation, Cornell University.

Zhang, J. (2002). *The effects of duration and sonority on contour tone distribution--A typological survey and formal analysis.* New York: Routledge.

Zoll, C. (1997). Conflicting directionality. *Phonology, 14*, 263-286.

BIOGRAPHICAL SKETCH

Mingzhen Bao was born and grew up in Hangzhou, China. She went to Zhejiang University in her hometown, where she received a Bachelor of Arts degree in English in 2001 and a Master of Arts in applied linguistics in 2004. During her M.A. study, she traveled to Beijing, China, for one year as a visiting student in Speech Group, Microsoft Research Asia. Mingzhen moved to the U.S. in the same year of her M.A. graduation to study linguistics at the University of Florida. In her four years at UF, she completed a Doctor of Philosophy in linguistics, with specialization in phonetics. During her PhD. training, She worked as a teaching assistant for the linguistics Program from 2005 to 2006 and as a research assistant for Professor Ratree Wayland from 2005 to 2008. She received a four-year Alumni Fellowship from the university, four annual awards of Outstanding Academic Achievement from the UF International Center as well as several travel grants from College of Liberal Arts and Sciences, and the Graduate Student Council. She will be working as an assistant professor in the Department of Modern and Classical Languages, Literatures, and Cultures at the University of Kentucky after graduation.