THE UNIVERSITY OF CHICAGO

INTONATION SYSTEMS OF MANDARIN AND ENGLISH: A FUNCTIONAL APPROACH

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE HUMANITIES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF LINGUISTICS

BY

FANG LIU

CHICAGO, ILLINOIS

MARCH 2009

This dissertation is dedicated to the memory of my father,

Guangqi Liu (1936-2008),

who should have lived to see its completion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would first like to thank my committee, Alan Yu, Gina-Anne Levow, and John Goldsmith. I am very fortunate to have Alan as my advisor, especially when there is no phonetician in the department and I nonetheless major in phonetics. Alan has always been a source of encouragement and inspiration to me. Without his constant support, motivation, and guidance, I would still be enjoying my time as a graduate student and full-time mom and not consider finishing my Ph.D. anytime soon. Alan also frequently calls my attention to a theoretical rather than pure-phonetic approach to my research topics, from which I benefited greatly. Gina has enriched my understanding of speech prosody by including me in her reading group on the computational and linguistic perspectives of tone and intonation recognition and synthesis in the Department of Computer Science. Like Alan, Gina has worked with me on my qualifying papers and dissertation throughout the years and offered many valuable suggestions and comments on my manuscripts. Many thanks go to John Goldsmith for agreeing to be on my committee. He has been a role model for me as a teacher and scholar. His teaching in computational linguistics is inspiring, and his erudition and academic brilliance seem unreachable for me. It was he who suggested the topic of Chapter 3 to me. His generous and insightful comments were essential for the development of this dissertation.

I am very thankful to the Division of the Humanities at the University of Chicago for granting me Century Fellowship for the first five years of my graduate study so that I could receive an education for free, and for providing travel grants for me to attend phonetics conferences. Besides a Ph.D. in linguistics, I am privileged to have studied

statistics in the Department of Statistics at the University of Chicago, for which I would like to express my deep gratitude to Jerrold Sadock (former chair of the Department of Linguistics) and to Thomas Thuerer (Dean of Students, Division of the Humanities), who kindly gave me permission to pursue a second Master's degree in statistics. My sincere gratitude also goes to Joseph Chang, Sean Fulop, John Goldsmith, and Jiong Shen for being my references in support of my applications to statistics programs.

Yali Amit, my statistics advisor, guided me through the design and writing of Chapter 3 of this dissertation; I am very grateful to him for bringing something special to my dissertation. Thanks also go to Marc Coram, Paul Rathouz, Stephen Roberts, Mei Wang, and Kenneth Wilder for very helpful discussions on this project, and to Dinoj Surendran for writing the Matlab program used in F_0 normalization in Chapter 3.

Of course, none of this would have been possible without the education I received at Peking University, Beijing, China. I owe a great debt of gratitude to Jiong Shen, my B.A. and M.A. advisor, who not only introduced me to the beauties of experimental phonetics but also inspired me to study Mandarin intonation. Much gratitude also goes to Ziyu Liu, Futang Wang, Hongjun Wang, Lijia Wang, and Yan Zhou, who helped me in many ways during my time at Peking University.

I owe a great deal to the other members of Class of 2001 in linguistics, Nikki Adams, Jon Cihlar, Rod Edwards, Stefanie Kuzmack, and Thomas Wier. Without their friendship and help, I could not have survived the first two years of my graduate study at the University of Chicago. Other fellow students, including Greg Davidson, Irene Kimbara, Anne Pycha, Mayu Yoshihara, Keiko Yoshimura, and Ichiro Yuhara, also

provided me with very helpful advice on the fulfillment of graduate school requirements. Also in that regard, I am especially thankful to Vanessa Wright, who has answered my many questions regarding the maze of graduate school requirements in an efficient and timely manner even when I am away from Chicago.

The experiments in this dissertation were carried out in New Haven, CT and Chicago, IL. I am very grateful to Haskins Laboratories and to Barbara Need in the Language Laboratories and Archives at the University of Chicago for providing me with recording equipments and testing rooms. I would also like to thank all my subjects, and Yi Xu for providing subject fees from his NIH grants R01DC03902 and 1R01DC006243.

Furthermore, I owe special thanks to Lauren Stewart at Goldsmiths, University of London for hiring me as a part-time research assistant (as of January 2008) working on her amusia project, which not only pays me enough for me to afford my daughter's nursery fees, but also introduces a new research area to me. On a related note, special thanks should also go to Goldsmiths College Nursery and Jancett Group of Day Nurseries for taking good care of my daughter. Without their nursery places, I would not have been able to find time to work on my dissertation.

Finally, I would like to extend my utmost appreciation to my family. Yi Xu, my husband, has been by my side throughout the entire journey. He is not only a caring and thoughtful partner to me in everyday life, but also a professional and intelligent collaborator on most of my projects. He has always been my source of strength and comfort when our lives were either in order or in chaos. I thank him from the bottom of my heart for his unwavering support and love, and for his patience and faith in me.

Thanks must also go to Mia Xu, our little princess, for bringing so much fun and happiness to our lives, and for being such a wonderful reminder of the world outside academia.

It is my greatest regret that my father, Guangqi Liu, was unable to see me complete my Ph.D. He was always there for me whenever I needed him, but now I have to finish my dissertation without him. It is like I owe a debt that I can never repay, which saddens me greatly whenever I think of him. I would have been unable to come abroad to pursue my dreams without the support from my father and my mother, Yuzhen Yang. I am also indebted to my sisters(-in-law) Hui Liu, Baoju Liu, and Yitao Chen, and brothers(-in-law) Fuzhong Li, Guimin Liu, and Feng Liu. They have given me both financial and moral support throughout the years.

ABSTRACT

In the currently dominant autosegmental-metrical (AM) theory of intonational phonology, intonational forms are derived from observed intonational contours without reference to their associated functions. Consequently, not only the categorical status of the resulting intonational components needs subsequent proof, but also the meaning of the intonational contours requires explanations outside the definition of the components. In order to counteract these problems and to better understand speech intonation, this dissertation investigates intonation systems of Mandarin Chinese and General American English through a functional approach—surface forms being analyzed through underlying linguistic functions. Specifically, the following theoretical issues are explored on the intonation of the two languages: 1) the functional domains of lexical tone/stress, focus, and sentence type, 2) the role that focus plays in distinguishing sentence types, and 3) the interaction between lexical tone/stress, focus, and sentence type.

Five experiments were conducted to address these issues. Experiments 1 and 2 investigated whether focus and sentence type could be produced and perceived simultaneously in Mandarin, and if yes, how they would interfere with each other. Experiment 3 aimed to identify feature vectors that are most effective in characterizing statements and yes/no questions in Mandarin, where decision trees were implemented in the classification of intonational contours. Experiments 4 and 5 examined whether focus and sentence type are realized differently through lexical items (tone vs. word stress) in

Mandarin and English, and how the results are explained by the Parallel Encoding and Target Approximation (PENTA) model and the AM theory of English intonation.

The main findings include: (1) statement/question intonation is realized in parallel with focus and lexical items that also use pitch for their encoding in both languages, and (2) the similarities and differences between Mandarin and English intonation are essentially caused by the way sentence type interacts with focus and lexical tone/stress in the two languages. These findings are in support of the functional view of intonation, according to which components of intonation are defined and organized by individual communicative functions that are independent of each other but are encoded in parallel.

1 INTRODUCTION

1.1 Form and function

> It is the pervading law of all things organic and inorganic, of all things physical and metaphysical, of all things human and all things superhuman, of all true manifestations of the head, of the heart, of the soul, that the life is recognizable in its expression, that form ever follows function. *This is the law*.
>
> -- American architect Louis Sullivan, "father of modernism," 1896.

Whether or not form follows function is still an important and lively topic of debate in architecture and in many other disciplines. In intonational phonology, however, it seemed to be a non-issue or merely a personal preference, as explained in Pierrehumbert (1980: 59):

> In the literature, one can distinguish two approaches towards the problem of establishing which intonation patterns are linguistically distinct and which count as variants of the same pattern. One approach attacks the problem by attempting to deduce a system of phonological representation for intonation from observed features of F_0 contours. After constructing such a system, the next step is to compare the usage of F_0 patterns which are phonologically distinct. The contrasting approach is to begin by identifying intonation patterns which seem to convey the same or different nuances. The second step is to construct a phonology which gives the same underlying representation to contours with the same meaning, and different representations to contours with different meanings. … The work presented here takes the first approach, in fact, it stops at the first step in the first approach.

So in her far-reaching dissertation work Pierrehumbert (1980) took the "function follows form" approach and left the functional aspect of intonation virtually unaddressed after a full-fledged intonation theory was outlined. This theory has since been widely adopted in linguistic research on intonation, but its afunctional approach not only has left the question of intonational meaning widely open, but also has run into problems on a number of critical issues about its own robustness in accounting for various intonational forms.

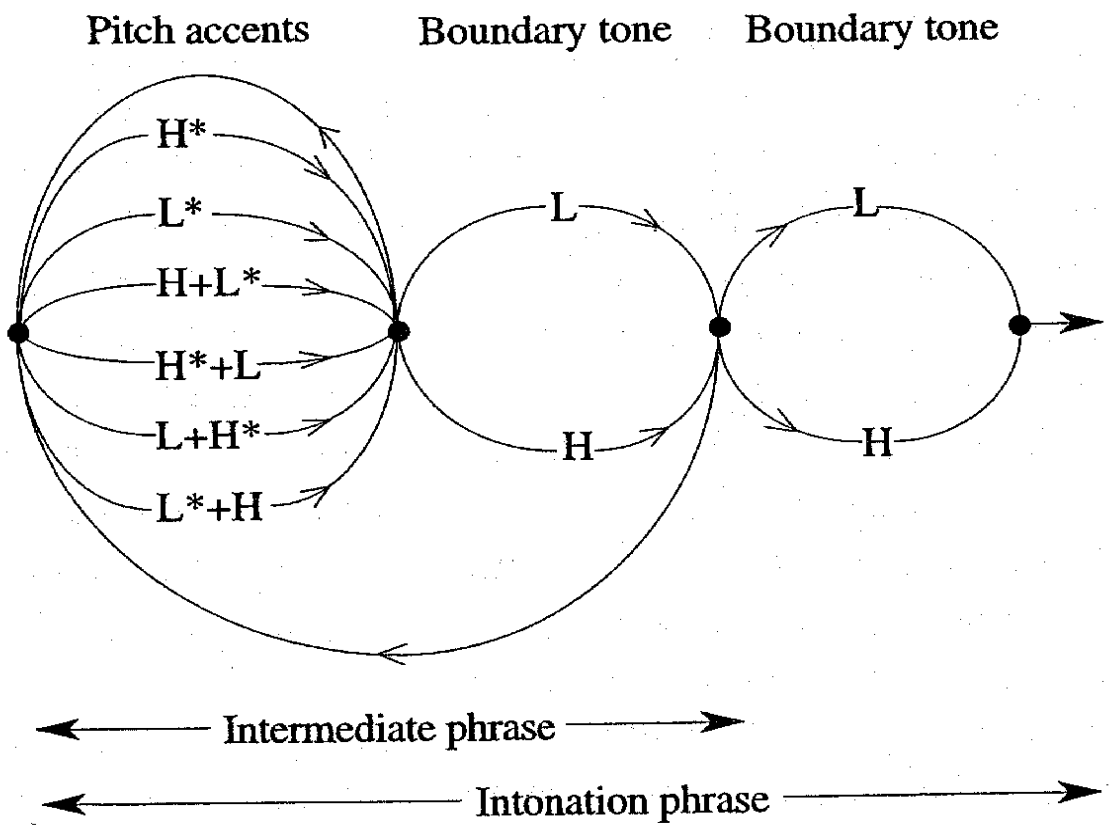

Figure 1.1. The autosegmental-metrical (AM) theory of English intonation. Adapted from Pierrehumbert (2000: 22).

Firstly, the finite-state grammar for the intonational system of American English (as shown in Figure 1.1), proposed in Pierrehumbert (1980), comprises a set of pitch accents (H*, L*, L+H*, L*+H, H+L*, H*+L, and H*+H, the last of which was later eliminated in Beckman & Pierrehumbert, 1986), phrase accents (H- and L-), and boundary tones (H% and L%). All the tunes of American English are supposed to be made up of at least one pitch accent, a phrase accent, and a boundary tone. Thus, intonation consists of linearly concatenated local tones, which are assumed to be independent from one another, and surface F_0 comes from interpolation between local

tones. More specifically, pitch accents are prominence-lending tones linked to a stressed syllable (denoted by an asterisk). They can be either single-toned (H* or L*) or bitonal with a leading or trailing tone (L+, H+, +L, or +H). Phrase accents span the tonal space between the last pitch accent and the end of an intermediate phrase. Boundary tones signify the edges of an intonational phrase. Since this grammar (which was later developed into the autosegmental-metrical (AM) theory of intonational phonology, cf. Ladd, 1996) was deduced from observed surface F_0 contours of some English utterances without rigorously comparing minimal pairs of tunes based on their meanings, the proposed intonational components appear more like phonetic transcriptions than phonological representations (see detailed discussion in Chapter 5). The resulting ambiguities have also lead to continuous dispute over even the existence of some of these components. For example, the categorical distinction between the two pitch accents L+H* and L*+H has never gained a consensus among intonational phonologists. In an effort to prove the phonological status of these two pitch accents, Pierrehumbert and Steele (1989) conducted an imitation experiment where five participants were asked to imitate 15 versions of the phrase *only a millionaire* with continuously manipulated F_0 peak timings on *millionaire*. Four of the participants produced bimodal distributions for the early peak and delayed peak variants of the rise-fall-rise contour (L+H* L H% vs. L*+H L H%). The authors then concluded that L+H* and L*+H are categorically rather than continuously distinct, even though they "function similarly in causing the word with the pitch accent to be implicitly compared to a scale of alternatives" (Pierrehumbert & Steele, 1989: 184; cf. Pierrehumbert & Hirschberg, 1990). Gussenhoven (1984, 2004)

and Ladd (1983), however, believe that L+H* should be analyzed as H*, because the leading tone L in L+H* belongs to the F_0 feature of the prenuclear accent.

Besides the dispute over the three-way distinction of L*+H, L+H*, and H*, the uniqueness of other pitch accents has also been questioned. For instance, as pointed out by Steedman (2000: 666), "the two accents H*+L and H+L*" are "much like particularly emphatic or theatrical versions of H* and L*." In fact, the identification of pitch accents and edge tones (either phrase accents or boundary tones) in Pierrehumbert's framework, which was later adapted into the ToBI (Tones and Break Indices) intonation labeling system (Silverman et al., 1992), is so difficult that pairwise agreement between ToBI transcribers did not reach 50% for six of the eight pitch accent label types (with L+H* and H* being most confusable) and for six of the nine edge tone label types (Syrdal & McGory, 2000). The confusability and similarity of the prosodic categories in Pierrehumbert's intonational system raise a fundamental question on the nature of her analysis: is it truly phonological or simply phonetic or something in between?

Secondly, as stated clearly by Ladd (1996: 98), intonational phonologists study "intonational *meaning*" only to verify the existence of "intonational *form*", which was defined *a priori* in Pierrehumbert (1980). Ladd (1987, 1996: 39) also advocated the so-called "Linguist's Theory of Intonational Meaning"—"*the elements of intonation have meaning*." This makes one wonder where the meaning came from if intonational elements were extracted without reference to the meaning-bearing aspects of the utterances. Pierrehumbert & Hirschberg (1990: 286) put forward "a compositional approach to tune meaning", namely, "the components of tune—pitch accents, phrase

accents, and boundary tones—are each interpreted with respect to their distinct phonological domains." They insisted on the independence of these intonational components, and assigned specific meanings to individual pitch accents, phrase accents, and boundary tones. However, statistical analysis of a subset of the ToBI-labeled Boston University Radio News Corpus indicates that there are significant interactions among pitch accents, phrase accents, and boundary tones such that the categories of the former two largely determine the identity of the following boundary tone (Dainora, 2001, 2002). Dainora thus argued against the compositional approach to intonational meaning in favor of a tunal approach, which assumes that meanings are directly associated with global tunes. Indeed, the meaning of a tune does not seem to be derivable from the individual meanings of its components, which is probably why people tend to analyze the meaning of a tune as a whole (e.g. the "uncertainty" reading for the rise-fall-rise nuclear contour L*+H L- H% in Ward & Hirschberg, 1985). Interestingly, Hirschberg (2004: 518) stated that "differences in accent type convey differences in meaning when interpreted in conjunction with differences in the discourse context and variation in other acoustic properties of the utterance," which seems to undermine the compositional approach (Pierrehumbert & Hirschberg, 1990) and the "Linguist's Theory of Intonational Meaning" (Ladd, 1987, 1996). Furthermore, despite her years of investigation into intonational meaning, Hirschberg (2004: 533) admitted that the "intrinsic meaning" of most intonational contours in Pierrehumbert's framework still remains "controversial" and "elusive".

As argued by Saussure (1998: 118), "a linguistic system is a series of phonetic differences matched with a series of conceptual differences", and "for the essential function of a language as an institution is precisely to maintain these series of differences in parallel." As a linguistic system, intonational phonology should not be an exception to this rule. It is against the basic phonemic principle to try to identify intonational components *a priori* without considering their meanings, because, as pointed out by Kohler (2005: 104), there is neither proper form nor proper function in the resulting "phonological" system:

> The dissection of global contours into elements and their symbolization, as in the dominant framework of autosegmental-metrical phonology and ToBI, can be no more than heuristic devices to get symbolic records of pitch data, and they should not be reified into underlying phonological entities in cognitive speech processing by speakers and listeners, which the phonetician fills with phonetic measurement in the laboratory. To arrive at these cognitive entities the opposite path is necessary, i.e. speech is investigated in an experimental framework to derive language categories that determine speech production and perception.

As reviewed above, Pierrehumbert's framework, despite its linguistic objective, has not established unambiguous links between intonational forms and intonational meanings. Meanwhile, a number of functional alternatives to the AM theory have been suggested in recent years. Kohler (2004, 2005) proposed that "function, time, and the listener" should be introduced into intonational phonology based on speech production and perception experiments. Hirst (2005: 338-339), on the other hand, suggested adapting ToBI into "Toneless ToBI or StarBI" by "dropping the tonal specification and keeping only the boundaries and prominences," since pairwise agreement between ToBI transcribers is rather low on the former but quite high on the latter (under 50% vs. over 90% in Syrdal & McGory, 2000). Steedman (2000: 653) emphasized that "there is no

single definitive characterization of the components of intonational contour, much less a definitive theory of their information-structural meanings," and he made it explicit that the information marked by pitch accents is essentially the "focus" (p. 656). He further distinguished "theme focus" from "rheme focus" so that they bear different types of pitch accents. However, the most comprehensive alternative to the AM theory to date has been the Parallel Encoding and Target Approximation (PENTA) model proposed by Xu (2005). This is a model of tone and intonation based directly on communicative functions and articulatory mechanisms (as shown in Figure 1.2). It assumes that different melodic components are defined in terms of meaningful communicative functions, which are encoded in parallel by controlling various articulatory parameters (pitch target, pitch range, duration, and strength) to generate surface F_0 contours through the target approximation process (Xu & Wang, 2001). More specifically, PENTA assumes that A) communicative functions are parallel to each other and are encoded by specifying individual encoding schemes, B) Encoding schemes specify distinctive values of Target Approximation (TA) parameters (which can be seen as within the domain of phonology), C) the TA parameters, including local pitch target, pitch range, articulatory strength and duration, are control parameters for the Target Approximation process, and D) the Target Approximation process successively approaches local pitch targets, each synchronized with a syllable, across specific pitch ranges, and with specific articulatory strengths (which can be seen as within the domain of phonetics). The PENTA model thus provides an extensive framework for systematic and cumulative examination of specific function-form relations in intonation.

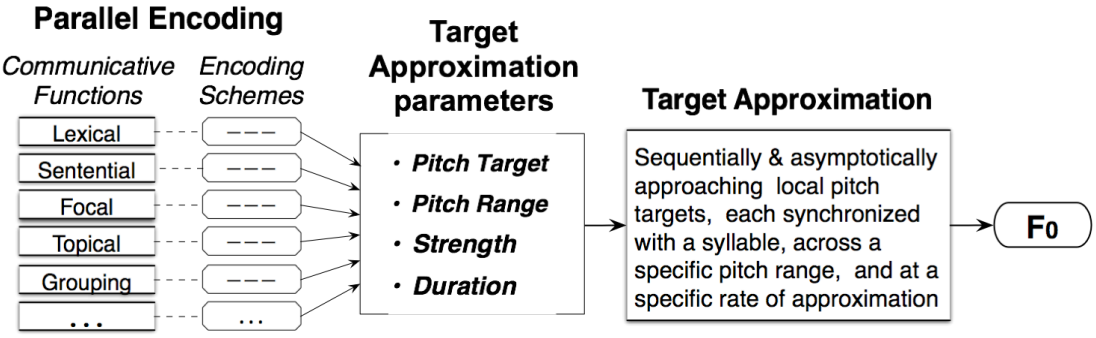

Figure 1.2. A schematic diagram of the Parallel Encoding and Target Approximation (PENTA) model. Modified from Xu (2005).

In summary, all the aforementioned alternative approaches point to the importance of communicative functions in determining intonational forms, rather than the other way around. Intonation studies based directly on explicitly functional approaches are still relatively few, however, and many of the intonation phenomena discussed in the AM theory have yet to be systematically examined under a functional framework. One difficulty in studying intonation in a language like English, however, is that most intonational categories, unlike segmental phonemes, are not orthographically represented. This means that the functional categories and their corresponding surface forms are equally unknown, which makes it hard to escape circularity in studying intonation. One solution is to use tone languages as a reference (Goldsmith, 1981; Xu and Xu, 2005), because lexical tones are phonologically explicit just like segments, which may help to reveal the underlying mechanisms of F_0 generation (Xu & Wang, 2001). The knowledge thus obtained can then be used in studying intonation in both tone and non-tonal languages (Xu, 1999; Xu & Xu, 2005). The use of this knowledge is crucial because the underlying forms of tonal and intonational components cannot be identified from the surface F_0 contours unless the articulatory mechanisms of F_0 production is taken into

consideration (Xu, 2004). This is a matter largely ignored by both intonational phonology and most of the functional approaches, and is taken seriously only in the PENTA model (Xu, 2005).

In this dissertation, I will take a functional approach to intonation in an investigation into statement and question intonation in Mandarin Chinese and General American English, two languages that have been extensively studied but are typologically very different, one being tonal and the other non-tonal. I will try to find out how multiple functions can be encoded in intonation in both languages by paying special attention to the interactive effects of lexical tone/stress, focus, and sentence type on surface F_0 contours. The typological difference between the two languages may help to highlight the commonalities shared by both. In particular, the surface variations of the Mandarin tones resulting from interactions with focus and sentence type may serve as a reference in my attempt to recognize the local pitch targets in English. As a general strategy, I will adopt the functional framework of the PENTA model proposed by Xu (2005). Specifically, the following general principles from the PENTA model will be assumed.

(1) Pitch target (either static or dynamic) of a lexical tone/stress is defined by two parameters of a linear function: height and slope; its approximation is synchronized with the syllable and determined by the articulatory process as a biomechanical system (Xu & Wang, 2001). Hence, there is no underlying specification of F_0 peaks or valleys, or their alignment. Both the turning points and their alignment are the natural result of articulatorily approaching the targets.

(2) Surface F_0 is the result of successively approaching a string of syllable-bound underlying pitch targets, whose specifications come both from lexically specified local targets (such as lexical tone/stress) and from supra-lexically specified target modifications (such as focus and sentence type). All communicative functions are encoded in parallel, and each of them is associated with a separate encoding scheme that specifies the parameters of the target approximation process (Xu, 2005).

1.2 Structure of the dissertation

While studying statement and question intonation in Mandarin Chinese and General American English, the following theoretical issues will be addressed in regard to intonation systems of the two languages. (1) What are the underlying targets of lexical tones/stresses in Mandarin and English, and how are they realized under different focus and sentence type conditions? (2) What role does focus play in highlighting certain lexical items and in differentiating sentence types? (3) How is sentence type manifested when focus is on different lexical items and when the degree of interrogative meaning varies? (4) What are the functional domains of lexical tone/stress, focus, and sentence type in the two languages?

Five experiments were conducted to investigate these issues. Experiments 1 and 2 (partial results of which were reported in Liu & Xu, 2005) explore whether focus and sentence type could be produced and perceived together in a sentence, and if yes, how they would interfere with each other. Experiment 3 (partial results of which were reported in Liu et al., 2006) compares the pros and cons of three sets of feature representations of

intonational contours in automatic classification of statements and yes/no questions in Mandarin. Experiments 4 and 5 (partial results of which were reported in Liu & Xu, 2007a, 2007b) investigate how focus and sentence type are realized through lexical tone and word stress in Mandarin and English, and how the AM theory and the PENTA model explain the interaction of the three communicative functions in English intonation. Below is a more detailed outline of the five experiments.

In Chapter 2, Experiment 1 examines the differences between Mandarin statements, yes/no questions, particle questions, wh-questions, rhetorical questions and confirmation questions under different focus conditions (initial, medial, final, or neutral), from which the following questions are to be answered. (1) How does focus interact with interrogative meaning in determining the F_0 contours of questions? (2) What is the functional domain of question intonation? (3) Are there F_0 differences among different types of questions in Mandarin? If so, is it possible to separate the phonetic manifestation of inquiry (or information-seeking) from additional connotations that may also be associated with interrogation, including incredulity, confirmation, and rhetoric?

Experiment 2 tests whether focus and sentence type (statement vs. yes/no question) could be perceived simultaneously by Mandarin listeners. The goal is to see whether acoustic variations related to the two communicative functions could be concurrently transmitted in speech perception, and to identify possible combinations of tone, focus, and sentence type that are more likely to cause perceptual confusions.

In Chapter 3, decision trees with three different sets of feature vectors are implemented in Experiment 3 to determine the most significant elements in an utterance

that signify its sentence type. Classification accuracy rates based on these decision trees are compared with human performance on the perception of sentence type obtained from Experiment 2. The results indicate that automatic classification of sentence type is much improved and simplified when the effects of speaker, tone, and focus are normalized on the final F_0 of each syllable in the sentence.

In Chapter 4, Experiment 4 investigates how question intonation in Mandarin is realized through the neutral tone, with the effects of the preceding full tones and different focus conditions also taken into consideration. The research questions are: (1) Which of the two has a stronger influence on the F_0 trajectory of the neutral tone: the preceding tone or sentence type? (2) Does the neutral tone have different targets in different sentence types? (3) What is the effect of focus on the neutral tone syllables in statements and questions? (4) Is a neutral tone more effective than a full tone in manifesting the statement/question distinction in Mandarin?

Experiment 5 studies the interaction of word stress, focus, and sentence type (statement vs. yes/no question) in General American English, and addresses the following questions. (1) Do focused stressed syllables have different pitch targets in statements and yes/no questions in English? (2) How does focus affect the global pitch contours of statements and yes/no questions in English? (3) Are there local pitch modifications for the pre- and post-focus content words that also signify the differences between statement and question intonation in English?

In Chapter 5, intonation systems of Mandarin Chinese and General American English are further compared using the PENTA model. Theoretical differences between PENTA and the AM theory are also discussed.

Chapter 6 draws the conclusions.

2 PRODUCTION AND PERCEPTION OF STATEMENTS AND QUESTIONS IN MANDARIN

This chapter consists of two experiments. The first investigates the phonetic manifestations of different sentence types (statement, yes/no question, particle question, wh-question, rhetorical question, and confirmation question) under different focus conditions (initial, medial, final, and neutral) in Mandarin. The aim is to find out whether different types of questions differ in their F_0 contours and how focus affects such differences. The second experiment explores whether Mandarin listeners can simultaneously identify the sentence type (statement vs. yes/no question) and focus information of a sentence, and if yes, how the two factors would interactively influence listeners' identification rates. Partial results of this chapter were reported in Liu and Xu (2005).

2.1 Introduction

Every utterance we say in a conversation or a monologue may be of one of several sentence types: statement, question, exclamation, command, request, etc. In addition to various, often optional, morphosyntactic manipulations, these sentence types are frequently conveyed through prosodic means, pitch contours in particular, or more broadly known as intonation. The difference between statement and question intonation, in particular, has been much researched in many languages. The general consensus is that sentences bearing the meaning of completion, termination, finality or assertion are associated with low or falling pitch, and those bearing the meaning of inquiry,

uncertainty, question and non-finality with high or rising pitch (Ladd, 1996). As summarized by Bolinger (1978), around 70% of the nearly 250 languages examined use a rising terminal to signal questions, whereas others use a higher overall pitch in questions than non-questions. There is much less agreement, however, over the details of such fall/rise dichotomy.

The first is in regard to the temporal scope of the rise/fall contrast in question versus statement. Many experimental studies have concluded that the relevant acoustic difference only occurs at the end of the sentence, e.g., Chang (1958) for Chengtu Chinese, Fok-Chan (1974), Vance (1976) and Lee (2004) for Cantonese, Rumjancev (1972) and Lin (2004) for Mandarin. Likewise, in the autosegmental-metrical (AM) theory of intonational phonology (Ladd, 1996; Pierrehumbert, 1980), the statement/question contrast is said to be linked only to boundary tones. A boundary tone, transcribed as H% or L% for a high- or low-pitched tone, is defined as a phonological tone located only at the right edge (i.e., the end) of an intonational phrase, although it may take the entire intonational phrase as its association domain.

Studies that have explored longer temporal domains in search of the acoustic correlates of the question/statement contrast have found evidence for non-local components. The patterns that have been reported are not highly consistent, however. Two general patterns have been described. The first is that in questions the F_0 of an entire sentence is raised (Haan, 2002; Ho, 1977; X.-N. S. Shen, 1990; Yuan et al., 2002). The other is that the question/statement contrast is time-dependent: the closer to the end of the sentence, the greater the difference between the two sentence types (Lindau, 1986

and Inkelas & Leben, 1990 for Hausa, Ma et al., 2004 for Hong Kong Cantonese, and Thorsen, 1978, 1979, 1980 for Danish).

To complicate things further, some researchers have suggested that global F_0 contours of variable shapes are associated with different types of questions. For example, X.-N. S. Shen (1990), as also supported by Ni and Kawai (2004) with the same sentence materials, proposes that the feature that distinguishes assertive intonation from interrogative intonation is a difference in register at the starting point: Interrogative intonation begins at a higher register than the assertive, but may end with either a high key (in unmarked questions and particle questions) or a low key (in A-not-A questions, alternative questions, and wh-questions).

The complicated temporal patterns reported by X.-N. S. Shen (1990) actually suggest that the question/statement contrast should not be investigated independently of other intonational functions. One such function is known as focus, namely, discourse/pragmatics motivated emphasis. It is now well established that focus plays a critical role in determining the global pitch shape of a declarative sentence in many languages, where a single (non-final) focus is manifested as a tri-zone pitch range adjustment: expanding the pitch range of the focused item, suppressing (lowering and narrowing) the pitch range of all post-focus items, and leaving the pitch range of pre-focus items the same as that in a sentence with no narrow focus (Botinis, Bannert & Tatham, 2000; Cooper, Eady & Mueller, 1985; Selkirk & T. Shen, 1990; J. Shen, 1985; Thorsen, 1979; Xu, 1999; Xu & Xu, 2005). In addition, focus has also been found to be accompanied by an increase in duration on the focused words (Cooper et al., 1985; Xu,

1999). Furthermore, perception patterns reflexive of such tri-zone pitch range adjustments have been reported (Mixdorff, 2004; Rump & Collier, 1996; Xu, Xu & Sun, 2004). More importantly, evidence for similar pitch range adjustment has been reported in question intonation as well. For Danish, Thorsen (1980:1021) noted that in sentences containing "emphasis for contrast", the difference between statement and question "lies partly in the level and movement of the emphatic syllable, but mainly in the course of the 'unstressed' ones after it, which perform less of a fall in questions than in statements." For English, Eady and Cooper (1986) found that in sentences with initial focus, there is no difference in peak F_0 between the focused word in statements and in questions, but the F_0 toplines depart radically after focus, with statements falling to a low F_0 and questions staying relatively high. For sentences with neutral or final focus, the peak F_0 of the final word in questions is significantly higher than that in statements. For Chinese, Wang (2003) observed that narrow focus is realized in three ways in both statements and questions: the abrupt decrease of the F_0 peak of the syllable following the focused word, the expansion of the pitch range of the focused word and the increase of the F_0 peak of the focused word.

Also as observed by Cooper et al. (1985) and Xu & Kim (1996, c.f. Xu, 1999), when not given any specific context or instructions, speakers in a recording session often spontaneously emphasize a particular part of a sentence in an unpredictable manner. This means that the occurrence of focus cannot be easily prevented, and thus its effect, if any, cannot be easily avoided. Hence it is possible that at least some of the discrepancies in the reported question intonation are due to uncontrolled spontaneous focus. In some other

cases, the syntactic structure of the sentence may favor a narrow focus on a particular part of the sentence. In X.-N. S. Shen's (1990) study, for example, focus can be anywhere in unmarked and particle questions, but in A-not-A questions, focus is likely to occur on the positive component, in disjunctive questions, on the alternative components, and in wh-questions, on the wh-words, especially when used as nouns (cf. Ishihara, 2002; Li & Thompson, 1979; Tsao, 1967). Consequently, the phenomena she observed are likely the combined effects of interrogative meaning and focus.

It has also been reported that certain additional factors may further affect question intonation, especially its pitch range. Bolinger (1986) has suggested that speaker involvement may affect pitch range: the greater the involvement, the larger the pitch range. Hirschberg and Ward (1992:250) showed that pitch range plays the largest role in interpreting the rise-fall-rise contour (L* + H L H% in ToBI's transcription), with larger pitch ranges indicating incredulity and smaller ones indicating uncertainty. Herman (1996) reported that in Kipare (a Bantu tone language), statements are signified by non-expanded pitch range with final lowering, yes/no questions by expanded pitch range with final lowering, and incredulous questions by expanded pitch range with final raising. Jun & Oh (1996) suggested that for some Korean speakers, incredulity questions (echo questions expressing incredulity) are distinguished from wh-questions by a larger pitch range, higher amplitude, and boundary tones.

The above discussion shows that several issues still need to be resolved about question intonation in Mandarin and other languages. First, it is not yet clear whether question intonation involves F_0 variations only at the sentence-final position, or rather in

a larger temporal domain. The final-only hypothesis is essential to the notion of boundary tone in the AM theory of intonation. Although studies such as Eady and Cooper (1986) have shown that there are F_0 differences non-local to the final word between statements and questions, it has been argued that the non-local patterns can be all accounted for in terms of phonetic implementation of sequential phonological units, involving L% boundary tone plus downstep for statement, but H% plus suspension of downstep for question (Ladd, 1996). Since downstep is assumed in the AM theory as a phonetic implementation rule triggered only by certain pitch accents such as H*L (Pierrehumbert and Beckman, 1988), it is possible to resolve this issue in a language like Mandarin where sentences can be found consisting of only H tones, thus preventing downstep from being triggered. Second, the role of focus in shaping question intonation is not yet fully clear: does focus involve the same tri-zone pitch range adjustment in question as in statement? Third, the conflicting findings about whether the F_0 of an entire question is raised need to be resolved. In this respect we note the frequent mention in the literature of the meanings nonessential to the interrogative meaning of questions, such as incredulity, surprise, etc. and the possible link between these nonessential meanings and global pitch raising. There is therefore a need to separate these meanings from the interrogative meaning when investigating question intonation.

This chapter was therefore designed to address three issues regarding question intonation in Mandarin: (1) Does question/statement contrast involve pitch differences only at the sentence-final position, or over a larger temporal domain? (2) Can focus and interrogative meaning be produced and perceived together in question intonation? If yes,

do they also interfere with each other? (3) Is it possible to separate the phonetic manifestation of interrogative meaning from those of non-interrogative meanings, e.g., incredulity, confirmation, and rhetoric? Two experiments were conducted to answer these questions. Experiment 1 examined the acoustic patterns related to focus and interrogative meaning in several sentence types. Experiment 2 tested whether focus and interrogative meaning could be perceived simultaneously by Mandarin listeners.

2.2 Experiment 1: Production of Statements and Questions in Mandarin

The goal of Experiment 1 is to investigate the acoustic manifestations of question intonation in Mandarin by addressing the following questions: (1) How does focus interact with interrogative meaning in determining the F_0 contours of questions? (2) What are the basic constituents of question intonation? (3) Are there F_0 differences among different types of questions (yes/no question, particle question, wh-question, rhetorical question, and confirmation question) in Mandarin?

2.2.1 Methods

2.2.1.1 Materials

Four basic sentence frames (each consisting of 10 syllables, all having identical tones: High, Rising, Low or Falling, corresponding to Tone 1, 2, 3 or 4 in the tonal phonology of Mandarin) were used, as shown in Table 2.1. These sentence frames were converted to six sentence types (statement, yes/no question, particle question, wh-question, rhetorical question, and confirmation question) by alternately adding an interrogative pronoun or verb phrase (*shúi* ('who/whom'), *gànmá* ('do what')), a negative particle (*bùshì* ('not')), a yes/no particle (*shìbùshì* ('yes/no')), an interrogative particle

(*ma*), a period, and/or a question mark. The sentences were to be said with focus at four possible locations (initial, medial, final, and none, i.e., neutral focus). 76 distinct sentences are thus constructed by varying tone component, sentence type, and focus location. Each sentence was to be repeated 5 times by each subject. Therefore, a total of 3040 sentences (76 sentences × 5 repetitions × 8 subjects) were investigated. The F_0 contours of three keywords in each sentence, shown as italicized in Table 2.1, was extracted and measured.

Table 2.1. Basic sentence frames used for constructing test materials.

Focus/Keyword	***Initial***		***Medial***		***Final***
Frame 1	*ZhāngWēi*	dānxīn	*XiāoYīng*	kāichē	*fāyūn*
(Tone 1)	*ZhangWei*	worry	*XiaoYing*	driving	*dizzy*
(High)	'*ZhangWei* worries that *XiaoYing* will get *dizzy* while driving'				
Frame 2	*WángMéi*	huáiyí	*LiúNíng*	huáchuán	*zháomí*
(Tone 2)	*WangMei*	suspect	*LiuNing*	canoeing	*obsessed*
(Rising)	'*WangMei* suspects that *LiuNing* will get *obsessed* with canoeing'				
Frame 3	*LǐMǐn*	fǎngǎn	*LiǔYǔ*	diǎnhuǒ	*qǔnuǎn*
(Tone 3)	*LiMin*	dislike	*LiuYu*	light a fire	*keep warm*
(Low)	'*LiMin* dislikes *LiuYu* to light a fire to *keep warm*'				
Frame 4	*YèLiàng*	hàipà	*ZhàoLì*	shùijiào	*zuòmèng*
(Tone 4)	*YeLiang*	afraid	*ZhaoLi*	sleep	*dream*
(Falling)	'*YeLiang* is afraid that *ZhaoLi* will *dream* while sleeping'				

The following is the set of sentence types composed from Frame 1 with all High tones.

Statement:
ZhāngWēi dānxīn *XiāoYīng* kāichē *fāyūn*.
'*ZhangWei* worries that *XiaoYing* will get *dizzy* while driving.'

Yes/no Question:
ZhāngWēi dānxīn *XiāoYīng* kāichē *fāyūn*?
'*ZhangWei* worries that *XiaoYing* will get *dizzy* while driving?'

Wh-Question:
Shúi dānxīn *XiāoYīng* kāichē *fāyūn*?
'*Who* worries that *XiaoYing* will get *dizzy* while driving?'

Particle Question:
ZhāngWēi dānxīn *XiāoYīng* kāichē *fāyūn* ma?
'Does *ZhangWei* worry that *XiaoYing* will get *dizzy* while driving?'

Rhetorical Question:
Bùshì *ZhāngWēi* dānxīn *XiāoYīng* kāichē *fāyūn* ma?
'Isn't *ZhangWei* who worries that *XiaoYing* will get *dizzy* while driving?'

Confirmation Question:
Shìbùshì *ZhāngWēi* dānxīn *XiāoYīng* kāichē *fāyūn*?
'Is it the case that *ZhangWei* worries that *XiaoYing* will get *dizzy* while driving?'

2.2.1.2 Subjects

Eight native speakers of Mandarin, 4 males and 4 females, served as subjects. They were either students at Yale University or residents in New Haven, Connecticut, who were born and raised in the city of Beijing where Mandarin is the vernacular. They had no self-reported speech or hearing disorders and their ages ranged from 22 to 34.

2.2.1.3 Recording

Recording was done in a sound-isolated booth at Haskins Laboratories, New Haven, Connecticut. A JavaScript program running under a web browser controlled the

flow of the recording. The subject was seated comfortably in front of a computer screen, wearing a headset microphone. The microphone was about 2 inches away from the left side of the subject's lips. The target sentences were displayed on a computer screen, one at a time, in random order. Subjects were instructed to say each sentence as a statement or question depending on whether it ended with a period or a question mark, and to emphasize any word that was surrounded by square brackets. The utterances were directly digitized onto a hard disk at 44.1 kHz sampling rate and 16-bit amplitude resolution. The digitized sound was later re-sampled at 22.05 kHz.

2.2.1.4 F_0 Extraction and Measurement

Using a custom-written script for the Praat program (www.praat.org), the waveform and spectrogram of each sentence and a label window were displayed automatically on a computer monitor. Onset and offset labels were manually inserted for the three keywords of the sentence whose F_0 trajectories and duration measurements were to be taken. Vocal pulse markings generated by Praat were displayed in another window. The pulses were inspected and any erroneous markings (such as missing pulses and double markings) were manually corrected. A custom-written Perl program then read in all the segment files and F_0 files saved by the Praat script. It applied a trimming algorithm to remove local spikes in the F_0 curves (Xu, 1999). The Perl program then extracted various measurements, including the mean, maximum, and minimum F_0 values of target words and their locations. Each mean F_0 value was computed by averaging over all the F_0 points in a word. For visual inspection and graphic analysis, the Perl program also computed time-normalized F_0 contours by getting the same number of evenly spaced F_0

points from each syllable. Durations of the key words were also taken by the program. This allowed the display of average F_0 contours against average time, assuring minimal information loss.

2.2.2 Results

For direct visual comparison, average F_0 curves (from 40 repetitions by 8 subjects) of the three keywords with different lexical tones, under different focus conditions, and in different sentence types are displayed in Figures 2.1-2.8. In computing these curves, the F_0 values were converted to a logarithmic scale before averaging, so as not to bias the means toward speakers with larger F_0 range. The mean values were converted back to Hz after averaging.

2.2.2.1 Effect of focus on the global F_0 curve

Figures 2.1-2.4 show the average F_0 contours of the three keywords in different sentence types (statement, yes/no question, particle question, rhetorical question, confirmation question, and wh-question) and focus conditions (neutral, initial, medial, and final) with the High, Rising, Low, and Falling tone sentence frames, respectively. What can be clearly seen is that regardless of sentence type and lexical tone, the pitch range of the focused words is raised and expanded (expansion most apparent in the Low tone in Figure 2.3), that of the post-focused words compressed and lowered, and that of the pre-focused words largely unaffected.

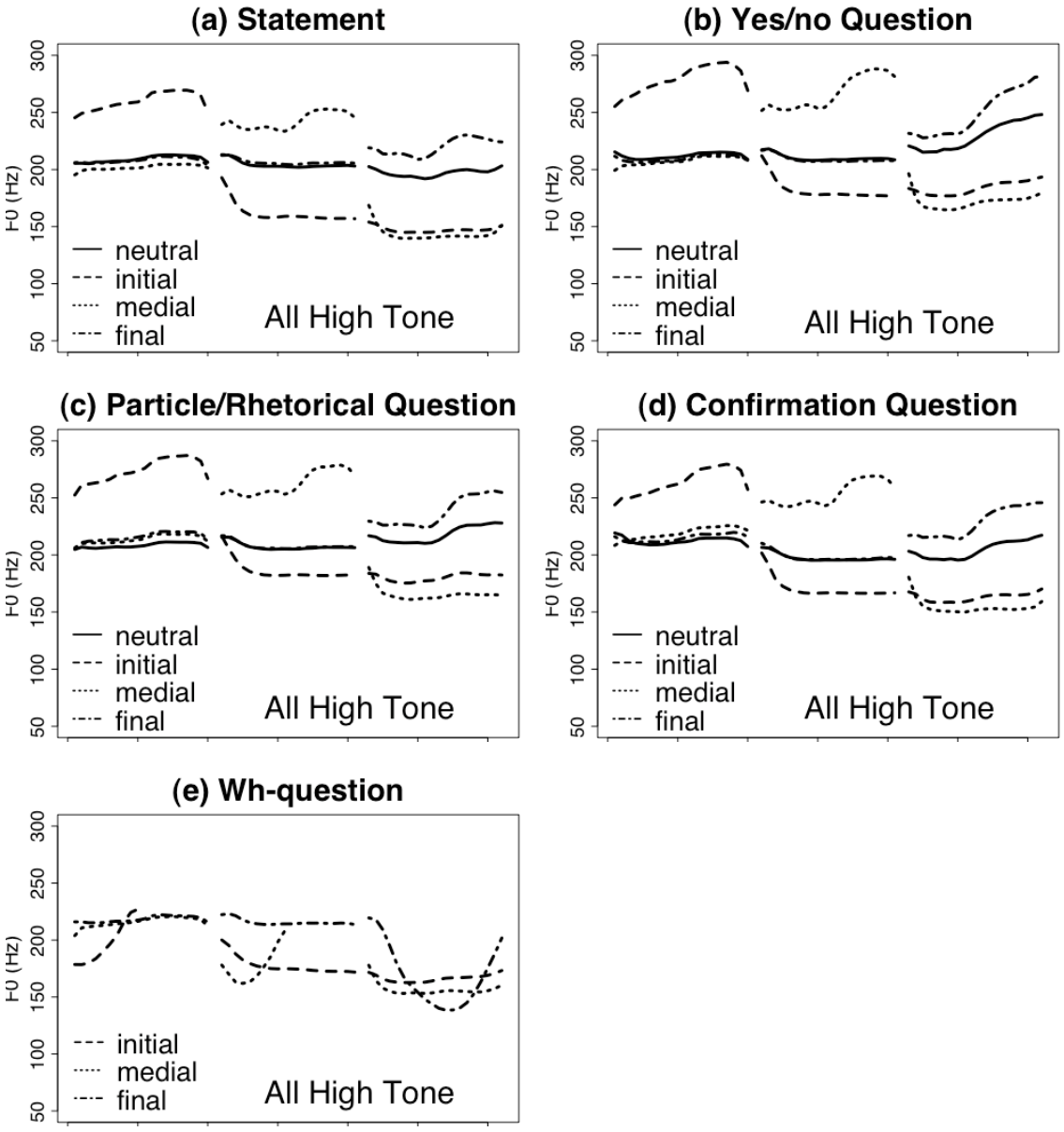

Figure 2.1. The effect of focus on global F_0 of different sentence types in the first sentence frame (containing all High tone). In each graph, the curves separated by the breaks are the F_0 contours of the initial, medial and final key words, averaged across all the repetitions and individual speakers. All the curves are time-normalized. The F_0 shapes of the wh-words (in (e) Wh-question) are very different from those of other words at the same position because they have different syllabic and tonal compositions (see explanations in section 2.2.1.1).

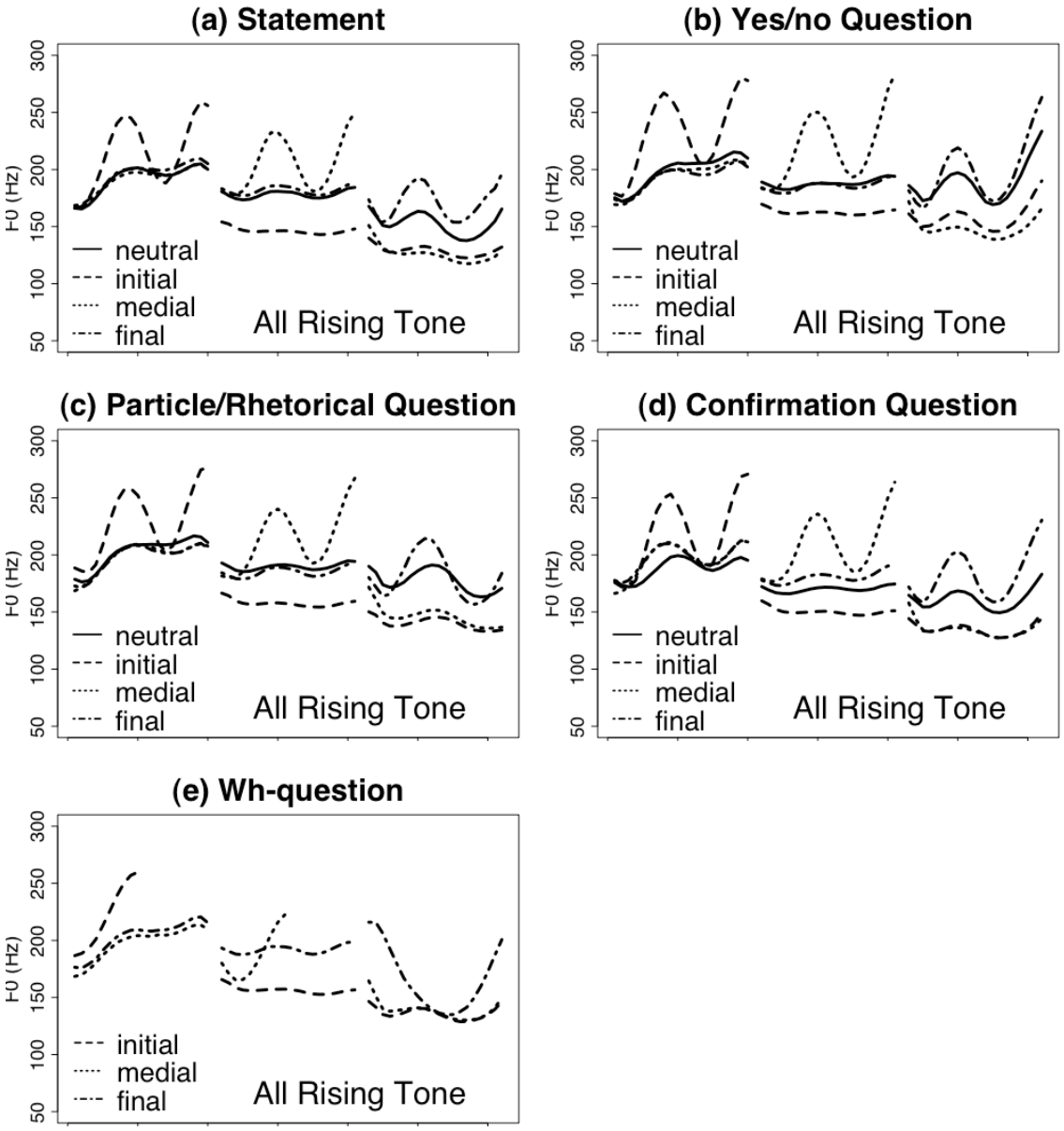

Figure 2.2. The effect of focus on global F_0 of different sentence types in the second sentence frame (containing all Rising tone).

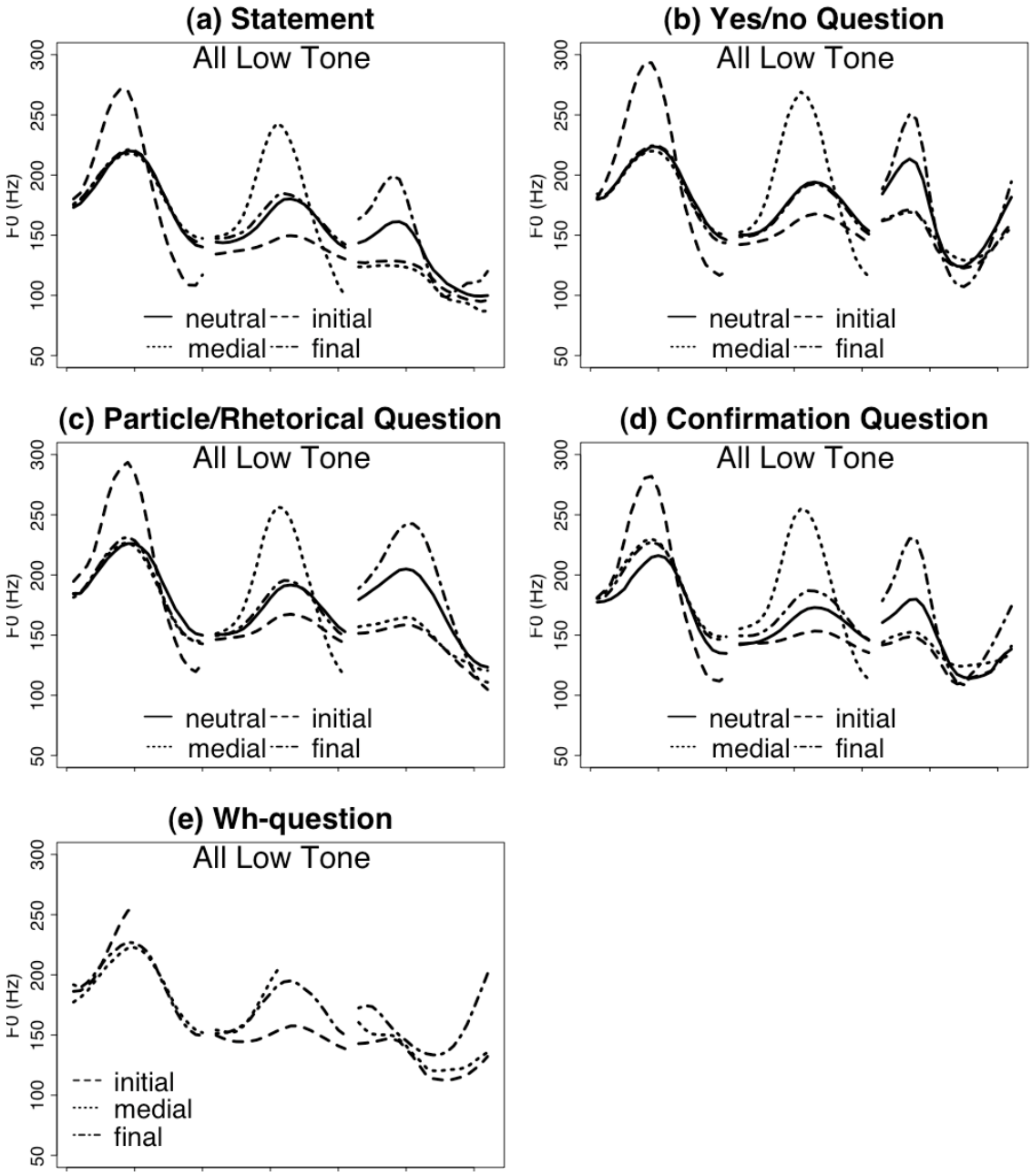

Figure 2.3. The effect of focus on global F_0 of different sentence types in the third sentence frame (containing all Low tone). Note that the first Low tone in each word is changed into the Rising tone due to the phonological rule Low + Low > Rising + Low (Chao, 1968).

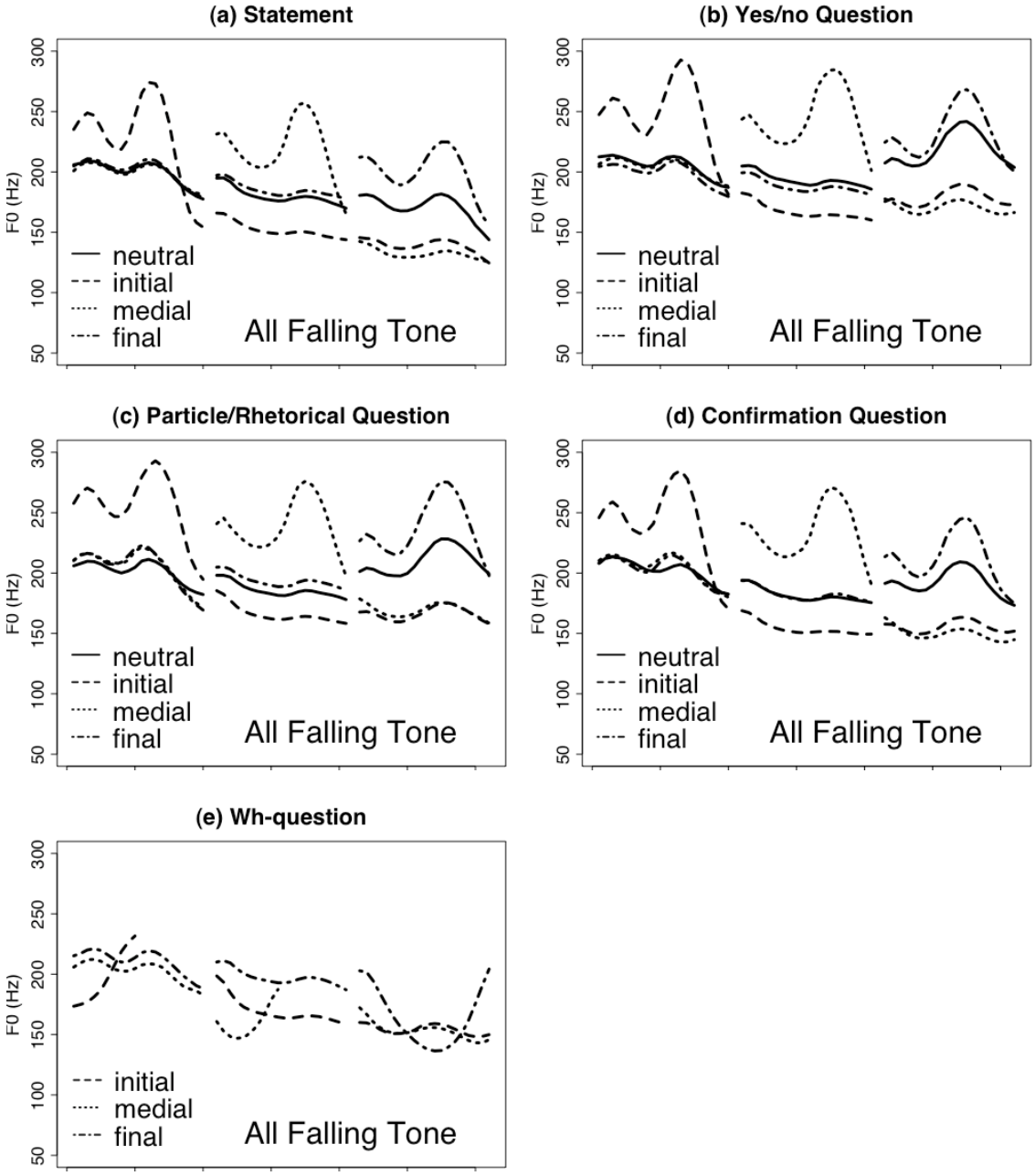

Figure 2.4. The effect of focus on global F_0 of different sentence types in the fourth sentence frame (containing all Falling tone).

Repeated-measures ANOVAs on duration and mean F_0 of initial, medial, and final key words were conducted, with lexical tone (High, Rising, Low, and Falling), focus (neutral, initial, medial, and final) and sentence type (statement, yes/no question, and

confirmation question) as fixed factors and subjects as replication factor (Table 2.2). Both duration and mean F_0 of focused words are found to be increased, regardless of sentence type and tone composition. Furthermore, mean F_0 of the final key words under neutral focus is only slightly lower than that under final focus (though marginally significant according to a linear mixed-effects regression model: t = -2.13, p = 0.0338). The suppression effect of focus on post-focused words is manifested by the significantly lowered mean F_0 of the medial key words under initial focus and by the significantly lowered mean F_0 of the final key words under initial and medial focus.

Table 2.2. Results of repeated-measures ANOVAs on the effect of focus on the global F_0 curve.

Key words	Duration (ms)	Mean F_0 (Hz)
Initial	$F(3, 21) = 35.501$, $p < 0.0001$ Initial (418.33) > final (283.99), medial (279.59), neutral (278.72)	$F(3, 21) = 41.579$, $p < 0.0001$ Initial (249.80) > neutral (203.40), final (201.59), medial (199.67)
Medial	$F(3, 21) = 49.14$, $p < 0.0001$ Medial (378.00) > final (241.09), neutral (230.87), initial (225.62)	$F(3, 21) = 73.284$, $p < 0.0001$ Medial (231.37) > final (188.68), neutral (188.01) > initial (159.76)
Final	$F(3, 21) = 26.828$, $p < 0.0001$ Final (408.97) > neutral (348.64), initial (348.62), medial (345.11)	$F(3, 21) = 50.282$, $p < 0.0001$ Final (211.69) > neutral (189.24) > initial (149.52), medial (148.10)

2.2.2.2 Effect of sentence type on the global F_0 curve

Figures 2.5-2.8 show mean F_0 contours of the three keywords in different sentence types and tone compositions under initial, medial, final, and neutral focus condition, respectively. As can be seen, in sentences with initial or medial focus (Figures 2.5-2.6), the difference between question and statement is manifested as a moderate raise in pitch range, starting from the focused word. Focus thus serves as a pivot at which statement and question contours start to diverge. In sentences with final or neutral focus (Figures 2.7-2.8), the difference between statements and questions is manifested mainly

in the final word, suggesting that the widely recognized question intonation with a final rise is that of a question with final or neutral focus.

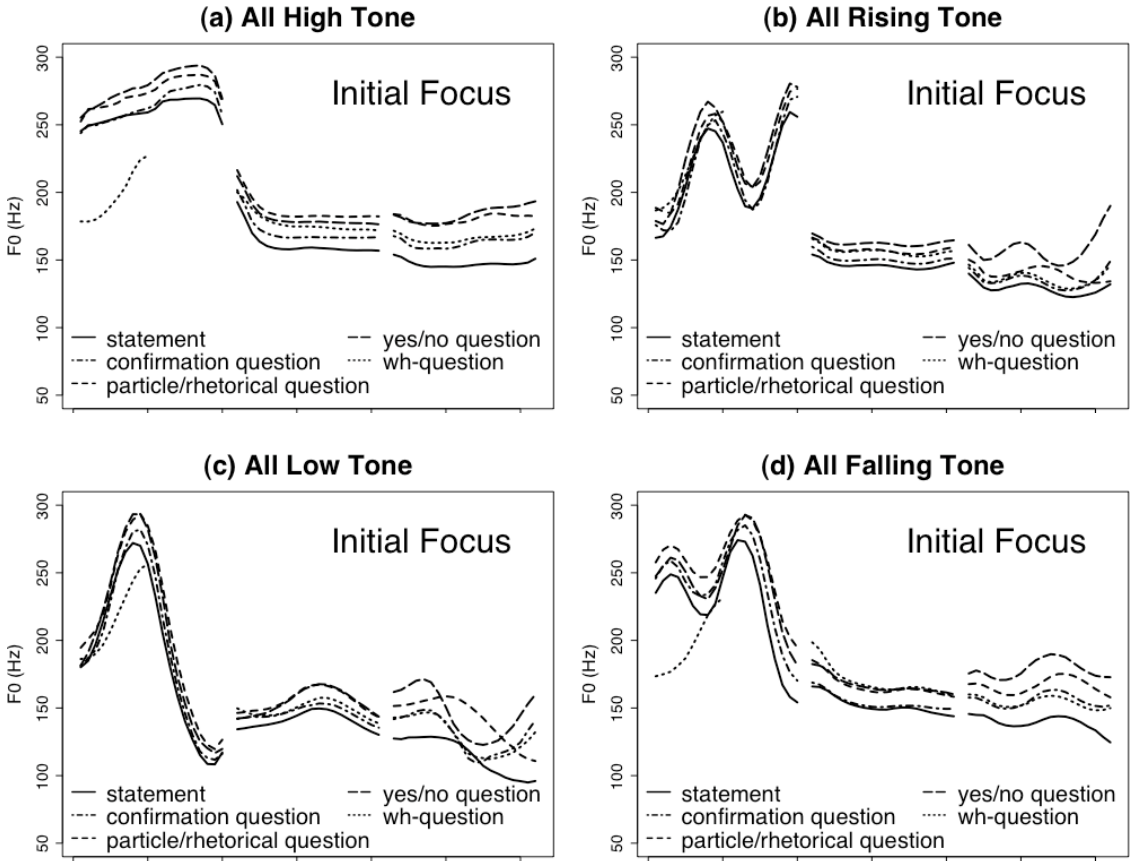

Figure 2.5. The effect of sentence type on global F_0 of four sentence frames in initial focus condition. See caption of Figure 2.1 for detailed explanations.

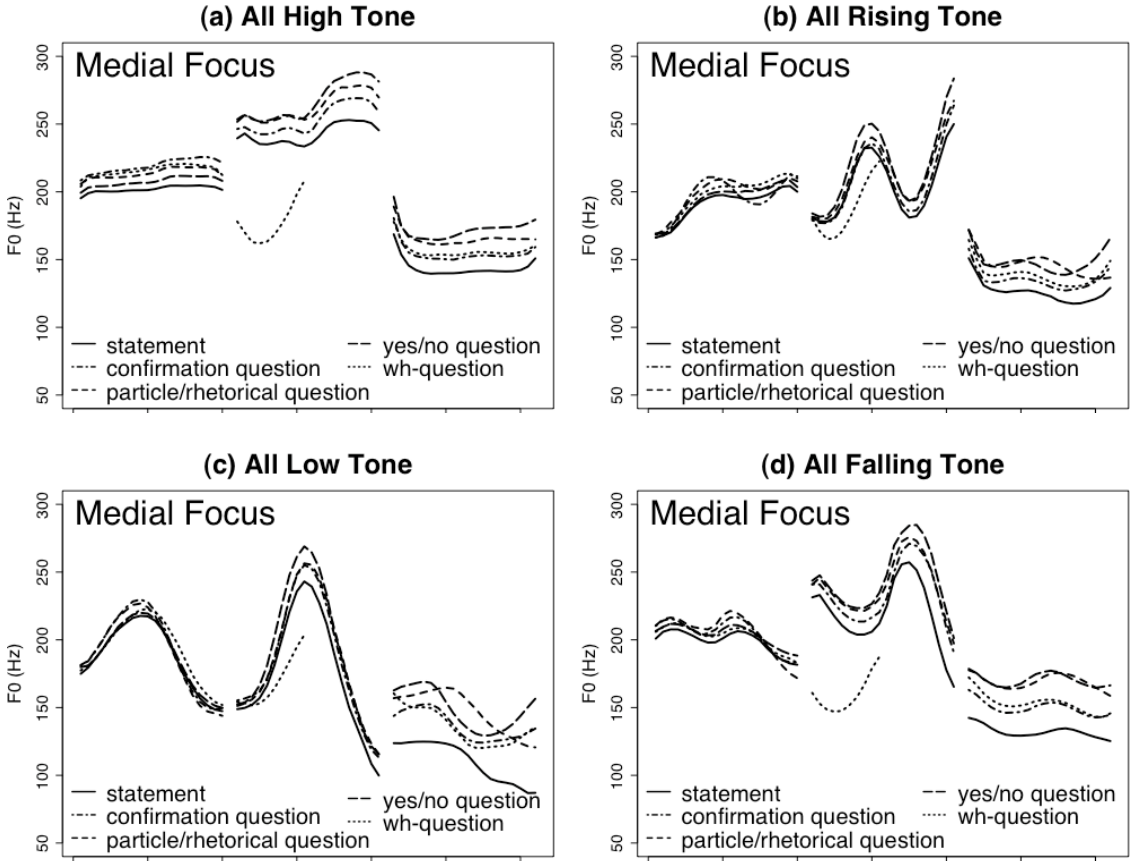

Figure 2.6. The effect of sentence type on global F_0 of four sentence frames in medial focus condition.

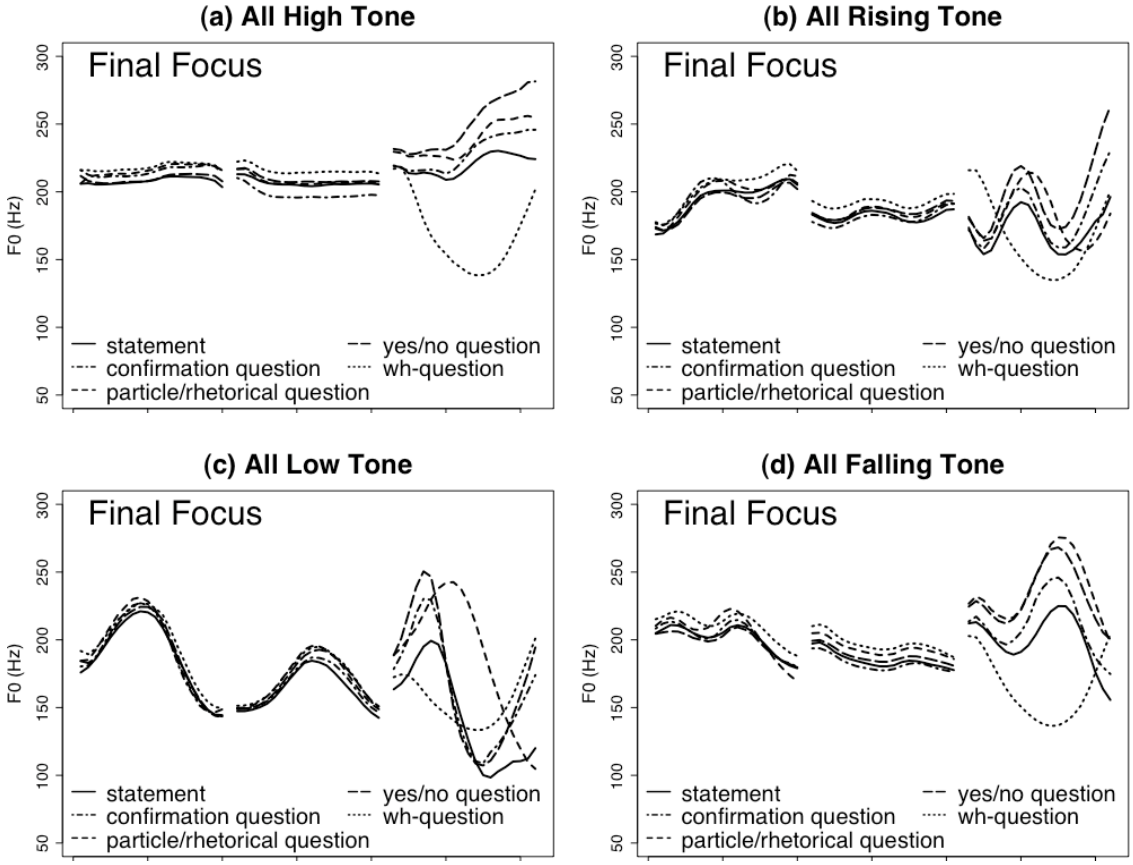

Figure 2.7. The effect of sentence type on global F_0 of four sentence frames in final focus condition.

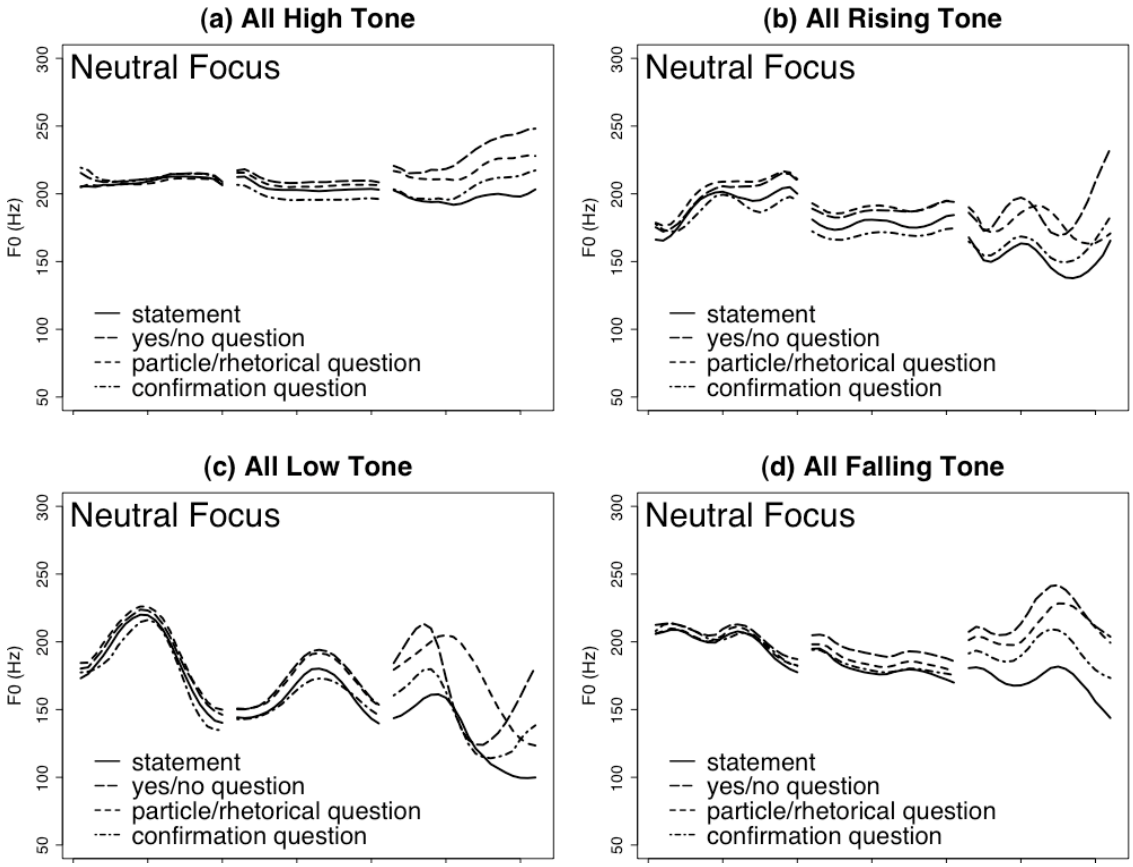

Figure 2.8. The effect of sentence type on global F_0 of four sentence frames in neutral focus condition.

To compare the effects of sentence type at different sentence locations, repeated-measures ANOVAs on mean F_0 of non-focused initial, medial, and final key words were conducted, with lexical tone (High, Rising, Low, and Falling), focus (initial, medial, and final, but no neutral focus because both wh- and rhetorical questions have implicit narrow focus) and sentence type (statement, yes/no question, wh-question, rhetorical question, and confirmation question) as fixed factors and subjects as replication factor. As can be seen from Table 2.3, the differences in mean F_0 among the sentence types increase as the sentence approaches the final position (see also Figure 2.9). Thus the greatest difference among sentence types is found at the sentence-final position. Pitch raising by question

intonations is greater in yes/no and rhetorical questions than in confirmation and wh-questions. Intonation of particle questions and wh-questions is different from that of statement, indicating that in addition to the use of particle and wh-word, pitch raising also occurs in these questions (see Figures 2.5-2.8).

Table 2.3. Results of repeated-measures ANOVAs on the effect of sentence type on the global F_0 curve. "Q" stands for question.

Key words	Mean F_0 (Hz)
Initial	$F(4,28) = 1.51$, $p = 0.2261$. Wh-Q (204.77), Confirmation-Q (203.09), Rhetorical-Q (202.98), Yes/no-Q (200.26), Statement (198.54)
Medial	$F(4,28) = 5.848$, $p = 0.0015$. Rhetorical-Q (181.52), Yes/no-Q (179.71), Wh-Q (179.41) > Confirmation-Q (172.57), Statement (170.38)
Final	$F(4, 28)=21.817$, $p < 0.0001$. Yes/no-Q (163.26), Rhetorical-Q (161.74) > Wh-Q (152.13), Confirmation-Q (148.51) > Statement (134.67)

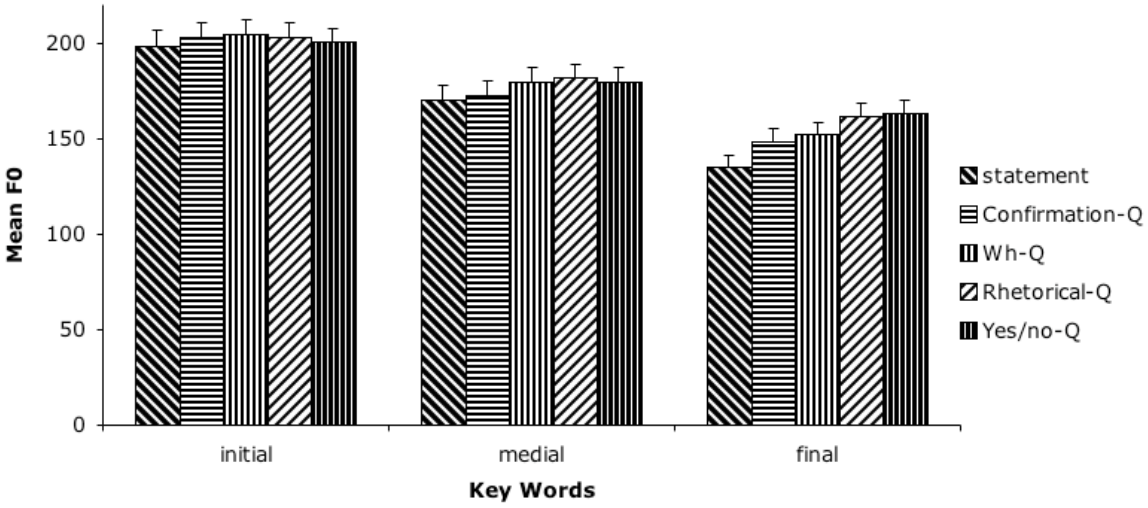

Figure 2.9. Mean F_0 (Hz) of the key words across the entire sentences under different sentence types. "Q" stands for question.

2.2.2.3 Interaction of focus and sentence type

Repeated measures ANOVA on post-focus pitch drop (= F_0 (st) of focused key word – F_0 (st) of post-focus key word) was conducted, with lexical tone (High, Rising, Low, and Falling), focus (initial and medial) and sentence type (statement, yes/no question, rhetorical question, and confirmation question) as fixed factors and subjects as replication factor. Post-focus pitch drop is not significantly different between initial and medial focus conditions ($F < 1$). Thus initial and medial focuses have similar amounts of suppression effects on the post-focus words. Questions with stronger incredulity connotations slightly reduce the degree of suppression of post-focus words ($F(3, 21) = 3.239$, $p = 0.0427$), as indicated by the following ordering of post-focus pitch drop: statement (8.55), confirmation question (8.22) > yes/no question (7.55), rhetorical question (7.19).

2.2.2.4 Linearity of question-statement difference

The analyses in section 2.2.2.2 show that the F_0 raising in a question accelerates toward the end of the sentence. It is therefore possible that the raising is non-linear. To examine this possibility, difference F_0 curves were obtained by subtracting the F_0 of statements from that of yes/no questions. In the difference curves, confounding factors common to both question and statement, new topic in particular (Xu, 2005), is largely left out, leaving only the "pure" contrast between the two sentence types. Figure 2.10 displays the difference curves averaged across repetitions, tones (excluding the Low tone) and subjects and grouped by focus. The columns group the fitted curves and equations obtained through three types of regressions: linear, exponential (both using mean time as

predictor) and double-exponential (using mean time squared as predictor). The dependent variables are corresponding mean F_0 values in semitones plus 1 **(**to make all values positive so that the exponential regressions are operable). As can be seen, except for medial focus, the curve fitting gets better from linear to exponential and to double-exponential. Table 2.4 shows the *t* values of the three types of regressions grouped by focus type (*p* values are all less than 0.0001 and thus not displayed). With the only exception of medial focus, the *t* values are always the largest in the double-exponential regression and the second largest in the exponential regression. Thus the 'pure' difference between question and statement does not appear to be linear, but at least exponential, or even double-exponential.

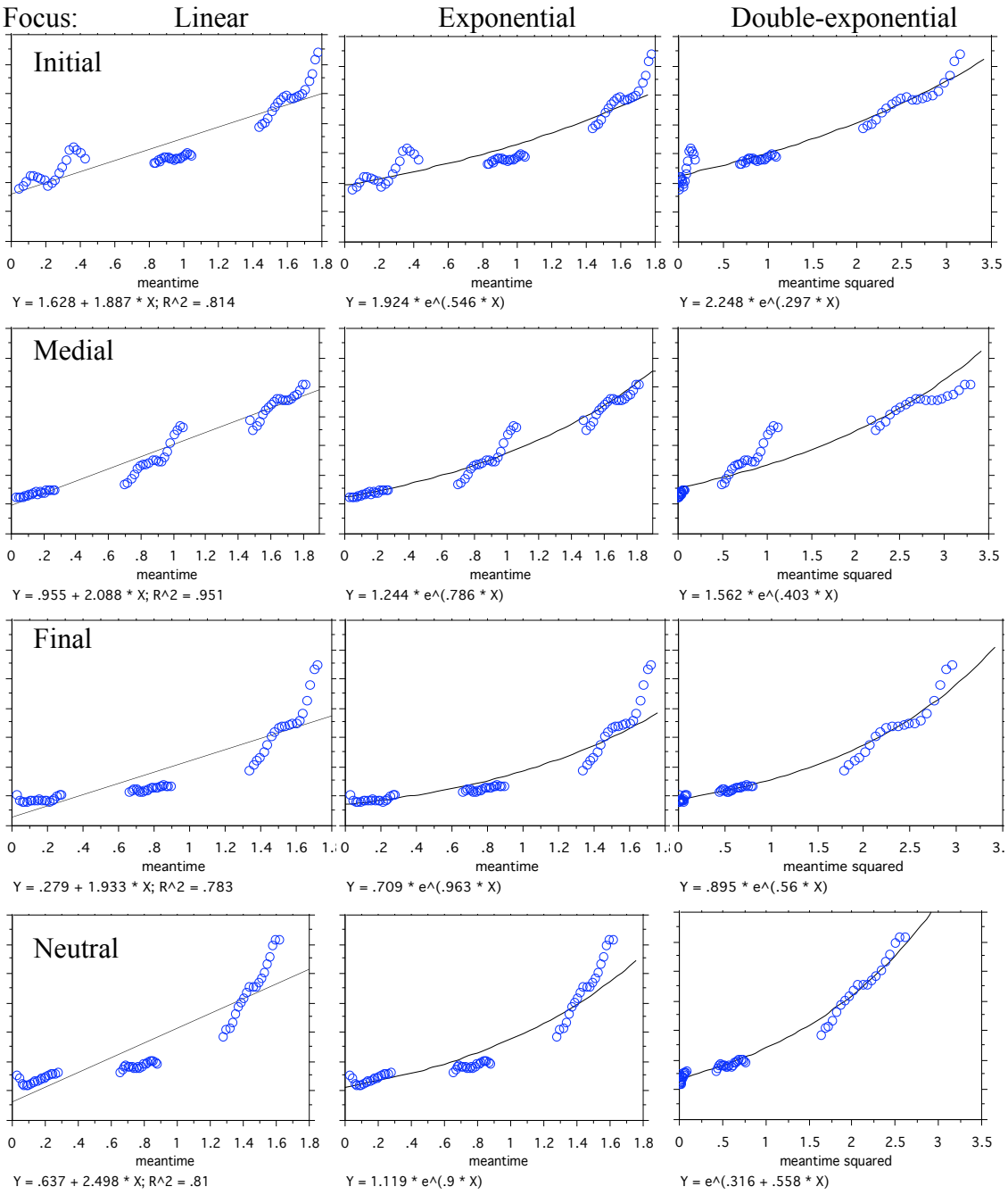

Figure 2.10. Circles: Mean difference F_0 in semitones (F_0 of yes/no question – statement + 1) averaged over the four basic sentence frames and grouped by focus conditions. Thin curves: fitted curves obtained through linear (left), exponential (middle) and equivalent of double-exponential (right) regressions. Equations of the curves are at the bottom of each graph.

Table 2.4. *t* values of linear, exponential and double exponential regressions for four focus types.

Regression type / Focus type	Linear	Exponential	Double-exponential
Neutral	15.57	24.12	54.79
Initial	15.79	18.75	23.42
Medial	33.36	41.37	21.75
Final	14.35	25.56	55.61

In Figure 2.10, one can also see the effect of focus on the statement/question difference. That is, the difference receives a boost at the location of focus. This boost probably has led to the much smaller *t* values of the double-exponential regressions for initial and medial focus than for neutral and final focus as seen in Table 2.4.

2.2.2.5 The neutral-tone question-particle ma

This section briefly shows how the pitch contour of the neutral-tone question-particle *ma* is adjusted by its preceding full tone under different focus conditions in question intonation in Mandarin. This issue will be further discussed in Chapter 4.

As explained in section 2.2.1.1, the neutral-tone question-particle *ma* can be added onto the four sentence frames in Table 2.1 to create a simple particle question with neutral focus or a rhetorical question under three different focus conditions (initial, medial, or final), depending on whether or where the negative particle *bùshì* (‘not’) is inserted in the sentence. The following is the set of particle/rhetorical questions with all High tones composed from Frame 1 in Table 2.1.

Particle question with neutral focus:
ZhāngWēi dānxīn XiāoYīng kāichē fāyūn ma?
'Does ZhangWei worry that XiaoYing will get dizzy while driving?'

Rhetorical question with initial focus:
Bùshì **ZhāngWēi** dānxīn XiāoYīng kāichē fāyūn ma?
'Isn't it **ZhangWei** who worries that XiaoYing will get dizzy while driving?'

Rhetorical question with medial focus:
ZhāngWēi bùshì dānxīn **XiāoYīng** kāichē fāyūn ma?
'Isn't it **XiaoYing** who ZhangWei worries will get dizzy while driving?'

Rhetorical question with final focus:
ZhāngWēi bùshì dānxīn XiāoYīng kāichē **fāyūn** ma?
'Isn't it getting **dizzy** that ZhangWei worries that XiaoYing will be while driving?'

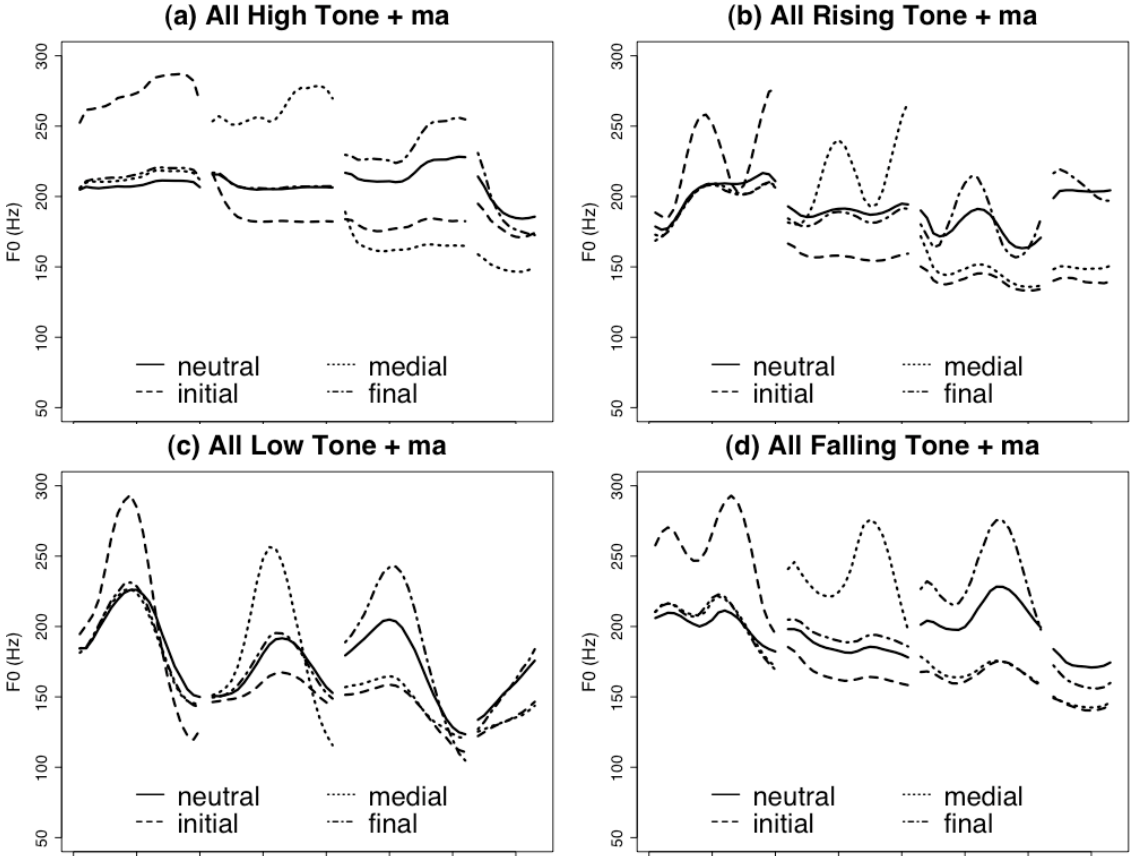

Figure 2.11. Pitch contours of the three keywords and *ma* in the four sentence frames under different focus conditions (neutral, initial, medial, and final).

Figure 2.11 displays time-normalized pitch contours of the three keywords and *ma* under different focus conditions (neutral, initial, medial, and final), each averaged

across 40 repetitions by eight subjects. As can be seen, the pitch contour of the question particle *ma* is affected both by the tone and by the focus condition of the preceding syllable. Firstly, the direction of the pitch contour of *ma* varies with the preceding full tone: having a falling contour after the High and Falling tones; having a rising contour following the Low tone; having a level contour in most cases after the Rising tone. Secondly, the slope of the pitch contour of *ma* is modified by the focus condition of the preceding syllable: steeper if the preceding syllable is on-focus (i.e., the sentence has a final focus), less steep if the preceding syllable is post-focus (i.e., the sentence has an initial or medial focus), and being moderate if the sentence has neutral focus.

Therefore, it appears that the pitch pattern of the neutral-tone question-particle *ma* is influenced both by the tone and by the focus condition of the preceding syllable. This agrees with the finding by Chen and Xu (2006:71) that "the direction and magnitude of the F_0 contour variations in the neutral tone" are largely determined by the final velocity of the preceding full tone. Furthermore, the specific F_0 contours of the neutral tone after different full tones in the present data are in support of the hypothesized static mid target for the neutral tone in Chen and Xu (2006). As seen in Figure 2.11, the neutral-tone question-particle *ma* falls after the High tone because its mid target is lower than the high target. The F_0 contour of *ma* is flat after the Rising tone, which is presumably due to the combined effect of the final positive velocity of the Rising tone and its own movement toward the mid target. F_0 of *ma* falls moderately after the Falling tone due to both the final negative velocity of the Falling tone and its movement to the mid target. Finally, *ma* rises considerably after the Low tone due to its upward movement to the mid target.

However, what is not clear in the present data is the contribution of question intonation to the F_0 pattern of the particle *ma*, as the current experiment does not include examples of *ma* (modal particle, indicating that something is very obvious) in statements such that comparisons between minimal pairs cannot be made regarding the effect of intonation on the F_0 variations in the neutral tone. Experiment 4 in Chapter 4 was thus designed to address this issue.

2.2.3 Discussion

The results of the above analyses suggest that the effects of focus are present in both statements and questions: the pitch range of the focused words is expanded, that of the post-focus words compressed and lowered, and that of the pre-focus words largely unaffected. The whole pitch level is shifted upward in questions with initial focus (in comparison with corresponding statements). In sentences with medial focus, the difference between questions and statements is manifested as a moderate raise in pitch range starting from the focused words. Focus thus serves as a pivot at which statement and question contours start to diverge. F_0 of both statements and questions with no narrow focus (neutral focus) is similar to those with final focus, i.e., showing the greatest sentence type difference in the final syllable. This seems to be evidence that the widely recognized question intonation with an extensive final rise is that of a question with final or neutral focus. The gradual pitch range raise caused by interrogative meaning is greater in yes/no questions and rhetorical questions than in wh-questions and confirmation questions, suggesting a possible separation of incredulity and interrogation as independent functions.

After establishing the significant effect of focus on F_0 contours of statements and questions in speech production in Mandarin (Experiment 1), it is important to examine whether Mandarin listeners can detect both focus and sentence type information from a spoken sentence through speech perception (Experiment 2).

2.3 Experiment 2: Perception of Statements and Questions in Mandarin

Experiment 2 was designed to address the second question raised in section 2.1, namely, whether native listeners are able to simultaneously perceive the statement/question contrast and the presence and location of focus.

2.3.1 Methods

2.3.1.1 Materials

320 statements and yes/no questions by two of the speakers (one male, one female; 2 speakers × 4 tones × 4 focus × 2 sentence types × 5 repetitions = 320) in Experiment 1 were used as stimuli.

2.3.1.2 Subjects

Eleven native speakers of Mandarin, 5 males and 6 females, served as subjects in this experiment. They were either students at Yale University or residents in New Haven, Connecticut, who were born and raised in the city of Beijing. They had no self-reported speech or hearing disorders and their ages ranged from 22 to 34. Five of them also served as speakers in Experiment 1.

2.3.1.3 Procedure

The task of the subject was to identify the sentence type (statement or question) and focus (initial, medial, final, or none, explained to them as emphasis [1]) of each of the 320 sentences. During the test, the subject was seated comfortably in front of a computer screen in a quiet room, wearing a headphone set. In each trial, eight response categories ("Question/Initial, Question/Medial, Question/Final, Question/None, Statement/Initial, Statement/Medial, Statement/Final, Statement/None") were displayed as selection boxes on the computer screen, and the subject clicked on the one that matched his/her impression after hearing each sentence. A new sentence was played 1.0 second after a choice was made. The whole process took about 45 minutes on average.

2.3.2 Results

There were 3520 trials (11 subjects × 320 trials) in total, of which 89.12% and 88.72% were perceived with correct sentence type and focus, respectively. Among the 713 misperceived trials (20.26%), 67 were perceived with both wrong sentence type and wrong focus (1.90%), 316 with only wrong sentence type (8.98%), and 330 with only wrong focus (9.38%).

[1] It is critical to instruct the subjects to listen for emphasis as opposed to listening for prominence as done in some studies. The identification of the former is quite robust as demonstrated by a number of studies (Mixdorff, 2004; Rump & Collier, 1996; Xu, Xu & Sun, 2004).

Table 2.5. Matrix of classification percentage (%) for each combination of sentence type and focus. "S" stands for statement, and "Q" for question.

To \ From	S.Initial	S.Medial	S.Final	S.Neutral	Q.Initial	Q.Medial	Q.Final	Q.Neutral
S.Initial	**89.32**	0.23	0	3.18	6.82	0	0	0.45
S.Medial	0.23	**86.82**	0.23	2.05	0	10.45	0.23	0
S.Final	0.68	0.45	**63.86**	12.50	0.68	0.23	13.86	7.73
S.Neutral	5.00	1.14	4.32	**81.14**	0.45	0.23	0.23	7.50
Q.Initial	12.27	0	0.23	0.91	**82.27**	0	2.72	1.59
Q.Medial	0	12.50	0.23	0.23	0.45	**84.55**	0.91	1.14
Q.Final	0	0	2.27	0.91	0.23	0.45	**85.23**	10.91
Q.Neutral	0.45	0.23	1.82	6.14	2.95	0.91	22.73	**64.77**

Table 2.5 is a matrix showing percentage of classification of each of the categories (row) to all eight categories (column). The following mismatching patterns can be observed: (1) Statement and question with initial or medial focus are most likely confused with each other. This is attributable to the effect of post-focus suppression: initial and medial focus compress and lower the pitch range of the post-focused words, making the pitch range of the final word in a question close to that of the final word in a statement. (2) Statements with final focus are the least easy to recognize (12.50% were identified as statements with neutral focus, 13.86% as questions with final focus, and 7.73% as questions with neutral focus). This indicates the similarity between neutral and final focus for both statement and question, and highlights the competing effects of sentence type and final focus on the F_0 contour of the final word: when the pitch range of the final word in a statement is raised by focus and thus somewhat resembling the F_0 pattern of the final word in a question, it is likely for the two sentence types to be confused. (3) The second most difficult to recognize are questions with neutral focus, of which 22.73% were heard as questions with final focus, and 6.14% as statements with neutral focus. Again, this is due to the similarity in F_0 between neutral and final focus. (4)

10.91% of the questions with final focus are recognized as with neutral focus. Once more, this is due to the similarity in F_0 between neutral and final focus.

Table 2.6. Mean accuracy rate of focus perception for each lexical tone grouped by sentence type and focus (collapsed across gender).

Focus perception		Lexical Tone				Mean
Sentence type	Focus	High	Rising	Low	Falling	
Statement	Initial	95.5	97.3	98.2	93.6	96.2
	Medial	98.2	97.3	97.3	96.4	97.3
	Final	85.5	70.9	70	84.5	77.7
	Neutral	90	90.9	87.3	86.4	88.7
Question	Initial	93.6	97.3	94.5	92.7	94.5
	Medial	95.5	97.3	96.4	99.1	97.1
	Final	81.8	86.4	95.5	86.4	87.5
	Neutral	74.5	68.2	74.5	66.4	70.9
Mean		89.3	88.2	89.2	88.2	88.7

Table 2.6 displays the mean accuracy rate of focus perception for each lexical tone grouped by sentence type and focus, collapsed across gender and focus location. Repeated measures ANOVAs on focus perception were conducted, with speaker (male and female), focus (initial, medial, final, and neutral), lexical tone (High, Rising, Low, and Falling), and sentence type (statement and yes/no question) as fixed factors and listeners as replication factor. The perception of the female speaker has significantly higher accuracy rate (90.6%) than that of the male speaker (86.9%) ($F(1, 10) = 5.343$, $p = 0.0434$). The accuracy rates of perception for different focus locations are significantly different ($F(3, 30) = 12.37$, $p < 0.0001$). A Student-Newman-Keuls post-hoc test indicates that accuracy rates of medial (97.2%) and initial (95.3%) focus are significantly higher than those of final (82.6%) and neutral (79.8%) focus. Lexical tone has no significant effects on the perception of focus. However, the interactions of sentence type × focus

(F(3, 30) = 31.28, $p < 0.0001$), sentence type × tone (F(3, 30) = 3.37, $p = 0.0313$), and sentence type × focus × tone (F(9, 90) = 3.60, $p = 0.0007$) are all significant. These are possibly due to the following perception results: (1) Final focus in statements and neutral focus in questions are the hardest to identify. (2) Focus perception in questions with High and Falling tones is worse than that in questions with Rising and Low tones, which is exactly opposite of the trend of focus perception in statements.

Table 2.7. Mean accuracy rates of sentence type perception for each lexical tone grouped by sentence type and focus (collapsed across gender).

Sentence type perception		Lexical tone				Mean
Sentence type	Focus	High	Rising	Low	Falling	
Statement	Initial	92.7	91.8	92.7	93.6	92.7
	Medial	85.5	84.5	92.7	94.5	89.3
	Final	62.7	73.6	78.2	95.5	77.5
	Neutral	89.1	92.7	87.3	97.3	91.6
Question	Initial	68.2	87.3	96.4	94.5	86.6
	Medial	71.8	85.5	99.1	91.8	87.1
	Final	95.5	95.5	98.2	98.2	96.9
	Neutral	84.5	90	95.5	95.5	91.4
Mean		81.3	87.6	92.5	95.1	89.1

Table 2.7 displays the mean accuracy rates of sentence type perception, collapsed across gender and focus location. Repeated measures ANOVAs on sentence type perception were conducted, with speaker (male/female), focus (initial, medial, final, and neutral), lexical tone (High, Rising, Low, and Falling), and sentence type (statement and yes/no question) as fixed factors and listeners as replication factor. The perception of the female speaker (93.4%) is easier than that of the male speaker (84.8%) (F(1, 10) = 80.91, $p < 0.0001$). The accuracy rates for statements and questions are not significantly different. Both focus (F(3, 30) = 3.02, $p = 0.045$; neutral (91.5%) > initial (89.7%) >

medial (88.2%) > final (87.2%)) and lexical tone ($F(3, 30) = 48.06$, $p < 0.0001$; Falling (95.1%) > Low (92.5%) > Rising (87.6%) > High (81.2%)) have significant effects on the perception of sentence type. Student-Newman-Keuls post-hoc tests indicate that the perception of sentence type under neutral focus is significantly better than under final focus, and that the pairwise differences between lexical tones for the perception of sentence type are all statistically significant. The interactions of sentence type × focus ($F(3, 30) = 25.48$, $p < 0.0001$), sentence type × tone ($F(3, 30) = 4.36$, $p = 0.0116$), focus × tone ($F(9, 90) = 3.91$, $p = 0.0003$), and sentence type × focus × tone ($F(9, 90) = 6.74$, $p < 0.0001$) are all significant, which is reflexive of the following facts. (1) The identification of statement under final focus has the lowest accuracy rate among the four focus locations, and initial and medial focuses cause trouble for the identification of question; (2) Questions with the High tone are hard to identify. (3) Sentences with the High tone have the lowest accuracy rates across all four focus conditions.

2.3.3 Discussion

Overall, the results of Experiment 2 show that listeners could identify both sentence type (statement versus yes/no question) and focus most of the time (89.1% and 88.7%, respectively). Nevertheless, low accuracy rates were found for neutral focus in questions (71%) and statements with final focus (78%). These confusions seemed to arise from the competing F_0 adjustments by sentence type and focus at the sentence-final position (final F_0 was raised for both question and final focus, but not for statement).

2.4 Summary of Experiments 1 and 2

The results of Experiments 1 and 2 largely answered two of the three questions raised in section 2.1. First, a clearer picture than before has emerged regarding the general pattern of question intonation in Mandarin (and potentially in other languages as well). Previously, question intonation has been described as involving (a) boundary tone only (Pierrehumbert, 1980; Ladd, 1996; Lin, 2004), (b) raising of F_0 of the entire sentence (Haan, 2002; Ho, 1977; X.-N. S. Shen, 1990; Yuan et al., 2002), or (c) superposition of a linear baseline onto the sentence starting from the first accented word (Thorsen, 1980). The current results are most consistent with account (c), but the detailed acoustic data also reveal that the global function is not linear, but more likely exponential or even double-exponential. Such exponential raising, unlike the linear raising proposed by Thorsen, can explain the much larger F_0 raising at the end of a question which has been the major motivation for the boundary tone account. But the exponential shape also demonstrates that the accelerated final rise is only part of the global question function, albeit the most prominent phase of the function.

Second, both the production and perception data show that focus and interrogative meaning can be transmitted concurrently. In production, focus exerts similar effects on the overall F_0 contours of questions as on those of statements, i.e., expanding the on-focus pitch range, suppressing the post-focus pitch range, and leaving the pre-focus pitch range largely neutral. Interrogative meaning raises pitch over the course of the sentence in an accelerated manner resembling an exponential or double-exponential function. In perception, listeners can simultaneously perceive both focus and interrogative meaning with high accuracy in most cases.

The current data have also provided preliminary evidence for answering the third question raised in section 2.1. That is, it is possible to separate the phonetic manifestation of interrogative meaning from that of non-interrogative meanings. As seen in Table 2.3 and Figure 2.9, the pitch range raise by question intonation is greater in yes/no and rhetorical questions than in confirmation and wh-questions. As shown by Hu (2002), speakers tend to raise the pitch register of the entire sentence to express surprise. Ho (1977) has found that F_0 is even higher in an exclamatory than in an interrogative sentence. The order of pitch raising found in section 2.2.2.2 seems to agree with the amount of incredulity/surprise in the different types of questions. However, because the sentences in Experiment 1 were produced without context or realistic discourse interaction, only subtle differences were observed across the question types. Clearer separations may be made in future studies using paradigms that involve realistic contexts or dialogue interactions to control the element of incredulity/surprise.

Regarding the complex global F_0 patterns related to different types of questions as reported by X.-N. S. Shen (1990), the current data also provide possible explanations. As seen in the difference curves in Figure 2.10, in addition to the exponential raising, F_0 in a question receives an extra boost at the location of focus. Thus if focus consistently occurs in certain sentence structures, such as in A-not-A questions, alternative questions, and wh-questions, the extra boost would interact with the global F_0 raising by question, increasing the on-focus pitch as well as decreasing the post-focus pitch. As for why F_0 is boosted at focus in a question, again it is possibly related to the element of incredulity/surprise that seems to be closely related to focus. That is, while focus in a

statement serves to highlight the importance of the focused item, in a question it seems to naturally carry a connotation of surprise/incredulity: is it really *that thing/person* that you mean?

The current data do not imply, however, that the encoding scheme of interrogative meaning in a language can only involve exponential pitch raising found in Experiment 1. It is highly likely that there exist language-specific components related to question intonation. In English, for example, in addition to the final pitch raising, a nonfinal focused word in a question probably has low, rather than high pitch, as in a statement (Eady & Cooper, 1986). This is apparently not the case in the preset data, presumably because in a tone language the local underlying pitch targets are not easily changed, for they encode lexical information (statement/question intonation of Mandarin and English will be compared in Chapters 4-5). In Greek, Hungarian, Romanian and Neapolitan Italian, it is known that F_0 always drops at the end of a yes/no question rather than rises as in English (D'Imperio, 2002; Grice et al., 2000). It is possible that in these languages the local pitch targets are also not free to change, just as in the case of the Mandarin Falling tone, in which F_0 drops even in a question. But also like Mandarin, questions in these languages may involve non-local pitch raising, as is already reported for Neapolitan Italian (D'Imperio, 2001).

Last but not the least, the new data appear to have provided a key to solving the long-standing puzzle as to why final focus is manifested much less effectively than an earlier focus (Botinis & Bannert, 1997; Botinis et al. 1999; Botinis et al. 2000; Cooper et al., 1985; Jin, 1996; Rump & Collier, 1996; Xu, 1999). Because the sentence-final

position is where the exponential pitch raising generates the greatest distinction between statements and questions, pitch range modification by any other function at the end of a sentence would directly compete with the question intonation. As shown by the results of Experiment 2, final focus in a statement, which already raises F_0 much less than does an earlier focus, still led listeners to often hear the sentence as a question rather than a statement with final focus. It thus seems that it is this competition that has been preventing final focus from significantly raising pitch range at the end of a sentence.

Overall, the findings of Experiments 1 and 2 seem to support the functional view of intonation — the Parallel Encoding and Target Approximation (PENTA) model (Xu, 2005), according to which components of intonation are defined and organized by individual communicative functions that are independent of each other. These functions are encoded in parallel, each with an encoding scheme that is distinct from those of all other functions. There are nevertheless frequent interactions among the encoding schemes because of limited availability of acoustic/articulatory dimensions and space, which has resulted in a delicate balance between functions that share the same articulatory/acoustic parameters. But each communicative function has to have at least one dominant encoding characteristic for it to be functional; the dominance of that encoding characteristic would lead to compromises by other functions that sometimes also have a need to use a similar encoding characteristic, as is the case with final focus versus question. The application of the PENTA model to intonation systems of Mandarin and English will be further explored in Chapters 4 and 5.

3 EXPERIMENT 3: AUTOMATIC CLASSIFICATION OF STATEMENT AND QUESTION INTONATION IN MANDARIN

This chapter is an initial attempt to apply phonetic findings in Experiment 1 to automatic classification of statement and question intonation in Mandarin using decision trees. The purpose is two-fold: (1) to identify the most important acoustic elements that signify the distinction between statement and question intonation in Mandarin, and (2) to search for the most effective feature vectors in the representation of F_0 contours of statements and yes/no questions in Mandarin. Partial results of this chapter were reported in Liu, Surendran, and Xu (2006).

3.1 Introduction

As discussed in section 2.1, there has been much controversy over the difference between statement and yes/no question intonation in the studies of Chinese prosody. One of the prevailing theories is that the whole F_0 level is shifted upward in questions as compared to statements (Ho, 1977; X.-N. S. Shen, 1990; Yuan et al., 2002), whereas an opposing view asserts that the essential difference between the two sentence types resides only in the last word or boundary tone (Rumjancev, 1972; Lin, 2004). However, results of Experiment 1 indicate that the pitch contour difference between statements and questions in Mandarin varies according to different focus conditions: (1) with initial focus, questions show an overall higher pitch contour than statements (Figures 3.1a-3.4a), (2) with medial focus, the difference is manifested as a moderate raise in pitch range starting from the focused words in questions (Figures 3.1b-3.4b), (3) F_0 contours of statements and questions with final focus are similar to those with neutral focus, i.e., showing the greatest difference in the final syllable (Figures 3.1c-3.4c and 3.1d-3.4d),

and (4) across all four focus conditions, the difference in F_0 between statements and questions increases nonlinearly toward the end of the sentence.

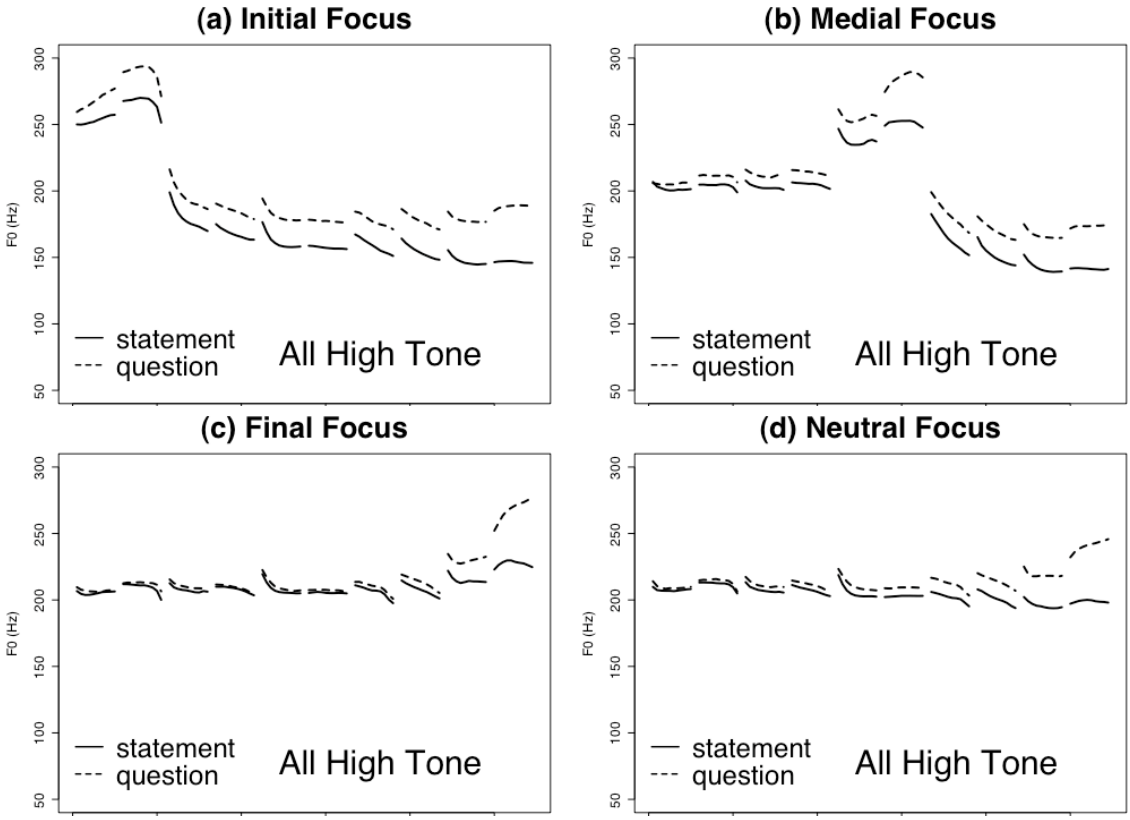

Figure 3.1. Time normalized F_0 contours of statements and yes/no questions (*ZhāngWēi dānxīn XiāoYīng kāichē fāyūn* ('ZhangWei worries that XiaoYing will get dizzy while driving')) with all High tones under initial, medial, final and neutral focus. F_0 contours in each plot were averaged across 40 repetitions by 8 subjects. Data were extracted from Experiment 1, but with all the syllables (not only the three key words) in the sentences shown.

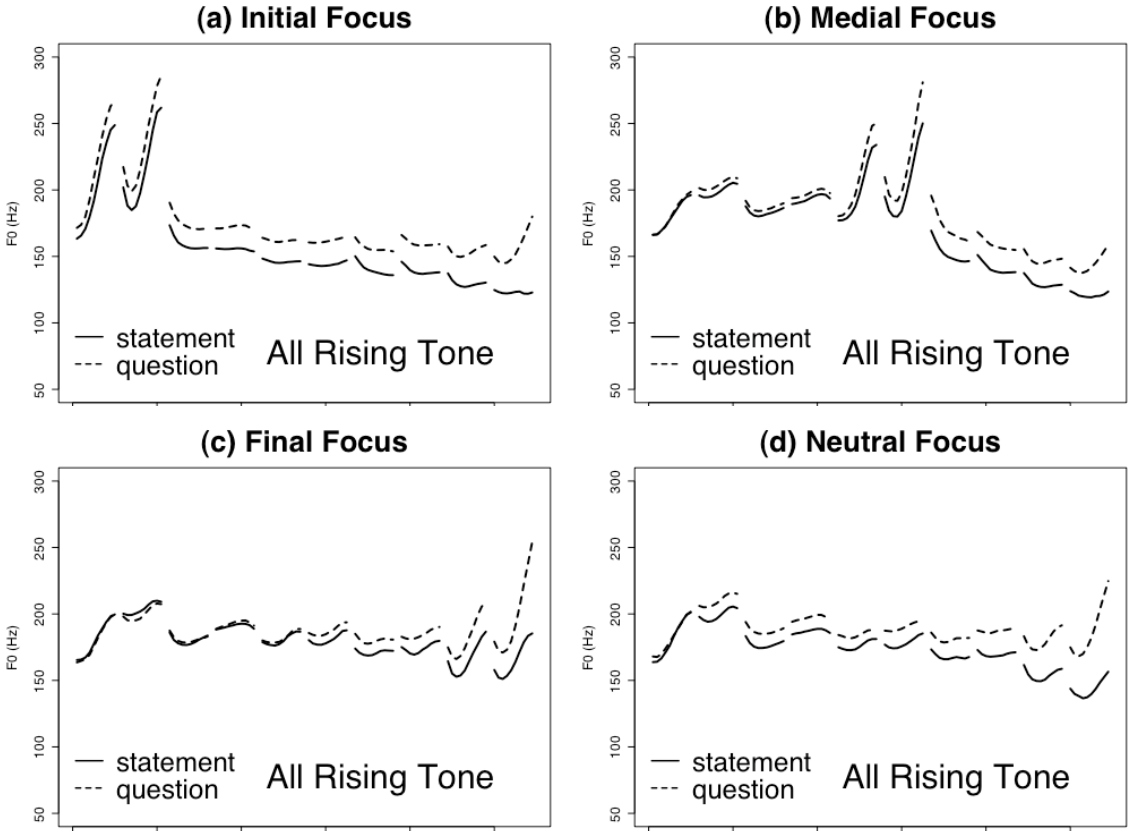

Figure 3.2. Time normalized F_0 contours of statements and yes/no questions (*WángMéi huáiyí LiúNíng huáchuán zháomí* ('WangMei suspects that LiuNing will get obsessed with canoeing')) with all Rising tones under initial, medial, final and neutral focus. See caption of Figure 3.1 for detailed explanations.

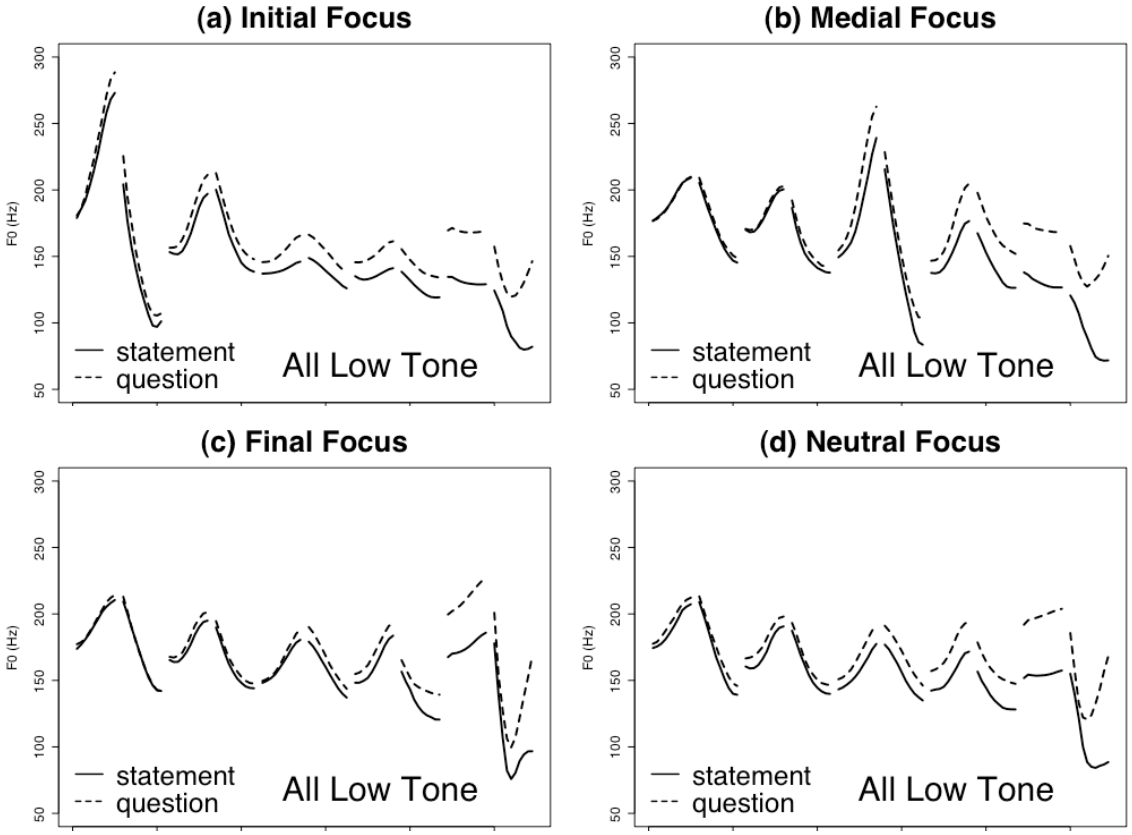

Figure 3.3. Time normalized F_0 contours of statements and yes/no questions (*LǐMǐn fǎngǎn LiǔYǔ diǎnhuǒ qǔnuǎn* ('LiMin dislikes LiuYu to light a fire to keep warm')) with all Low tones under initial, medial, final and neutral focus. See caption of Figure 3.1 for detailed explanations.

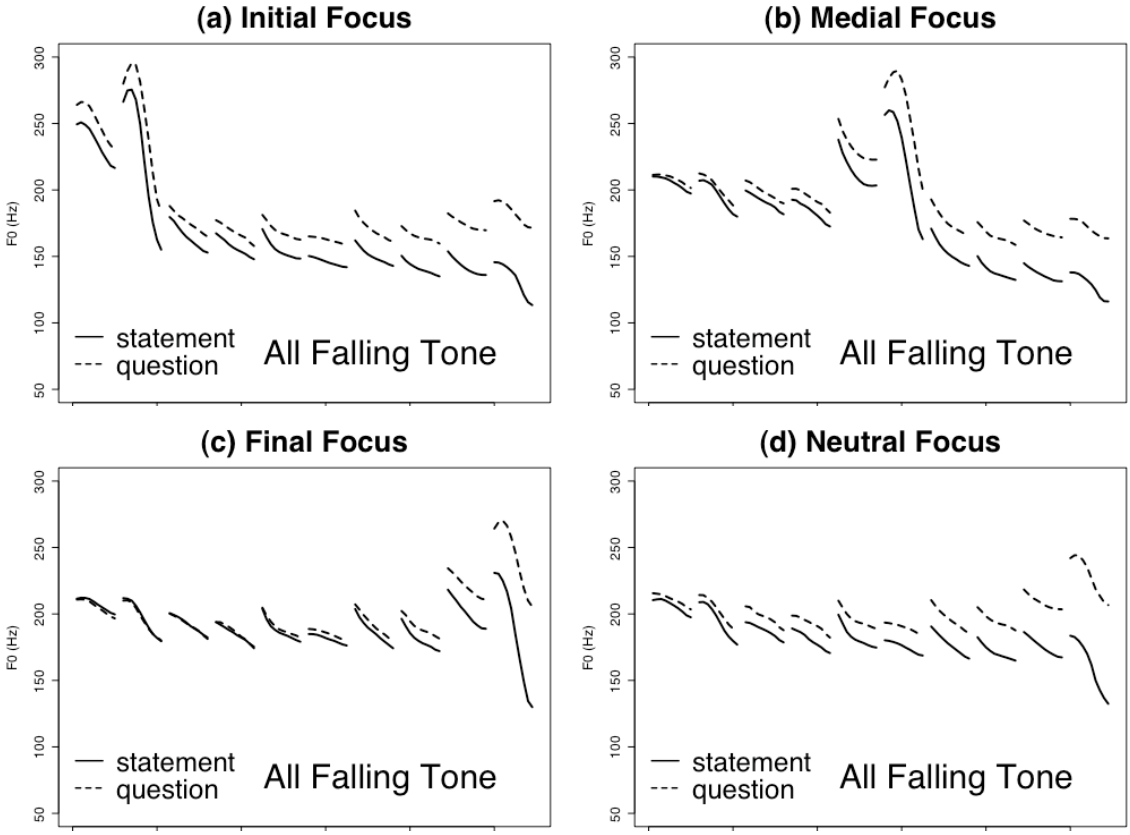

Figure 3.4. Time normalized F_0 contours of statements and yes/no questions (*YèLiàng hàipà ZhàoLì shùijiào zuòmèng* ('YeLiang is afraid that ZhaoLi will dream while sleeping')) with all Falling tones under initial, medial, final and neutral focus. See caption of Figure 3.1 for detailed explanations.

Furthermore, despite much research on intonation in different languages, speech engineers rarely use any of the proposed intonation models in detecting sentence types or dialog acts (e.g., statement, question, incomplete utterance, backchannel, etc.) in speech recognition, because it leads to little improvement (Mixdorff, 2002; Taylor et al., 1998; Wightman, 2002; Yuan & Jurafsky, 2005; Zue, 2007). Most often they employ as many prosodic features as possible in their implementation of decision trees to differentiate one dialog act from another. Disturbingly, removal of one set of features (e.g., F_0) can be compensated for by another functionally different set of features (e.g., pause) to achieve

roughly the same overall accuracy (Shriberg et al., 1998). Therefore, new approaches need to be explored to both improve the understanding of speech intonation and to apply intonation theories to the practice of speech recognition.

As shown in Experiment 1, pitch contours of Mandarin sentences are greatly affected by their tonal composition and focus condition, it is therefore desirable to first extract the most representative information of the syllables in each sentence and then use this information to characterize the entire sentence in order to model the difference between statement and question intonation in Mandarin. To test this hypothesis, in this chapter I will compare the efficacy of three types of features in distinguishing statements from yes/no questions, in each case using decision trees as the classification algorithm (Clark & Pregibon, 1992), with particular interests in the comparison between the B-spline coefficients which are obtained from a smoother representation of the local properties of F_0 and the syllable-final F_0's that are most characteristic of the tonal target of the syllable in Mandarin.

3.2 Methods

The dataset used here (as illustrated in Figures 3.1-3.4) is a subset from Experiment 1, with only 1280 statements and yes/no questions (= 8 subjects × 4 tone components × 2 sentence types × 4 focus locations × 5 repetitions) included. In implementing decision trees in the following sections (3.3.1-3.3.3), the dataset was further divided into a training and a testing set, with the former containing 960 sentences (480 statements and 480 yes/no questions) by 6 subjects (3 males and 3 females), and the

latter 320 sentences (160 statements and 160 yes/no questions, the same as the material used in Experiment 2 for speech perception) by 2 subjects (1 male and 1 female).

3.3 Analysis

3.3.1 Coefficients from cubic B-spline regressions

A fixed-knot cubic B-spline regression creates a piecewise cubic polynomial within certain knot spans that behaves well at the peaks, since each data point affects the global fit (Hastie et al., 2001). Figure 3.5 displays four examples of such regression, fitted using the R (R Development Core Team, 2005) command *lm(F_0 ~ bs(time, df = 13))*, where *df = 13* is to have the *bs()* function place 10 (= 13 - 3 because of cubic spline) interior knots uniformly along the range of *time* (Venables & Ripley, 2002).

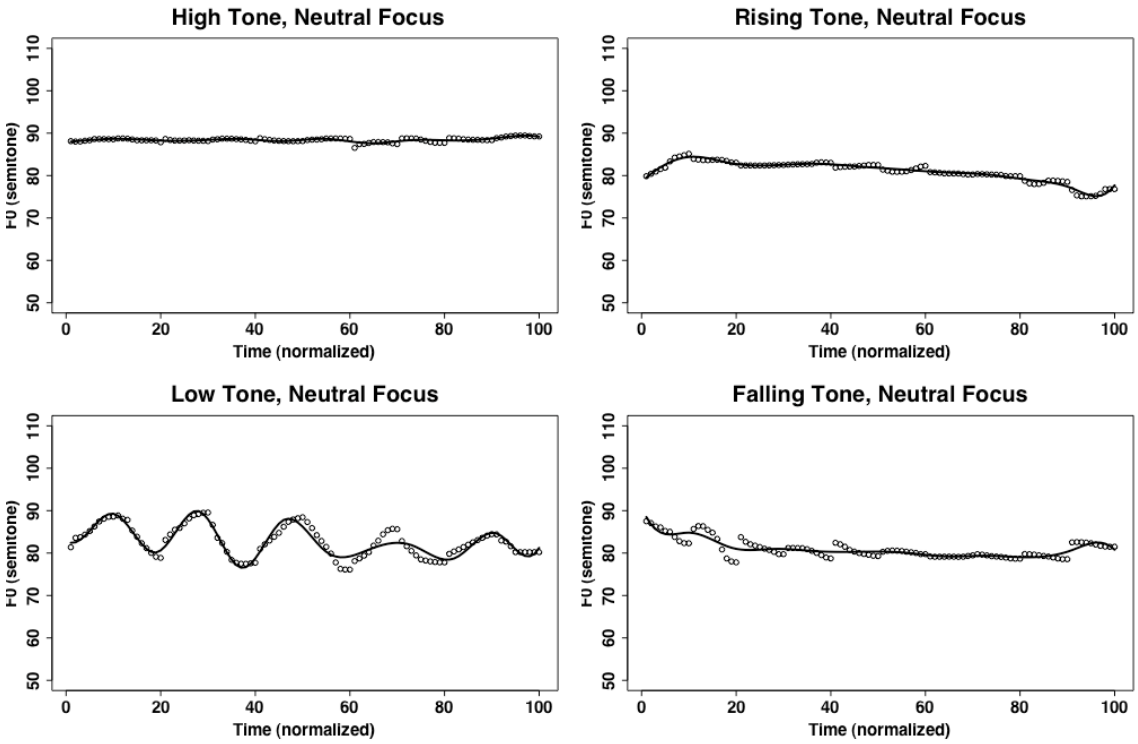

Figure 3.5. Examples of cubic B-spline regressions of F_0 (in semitones) on normalized time (1-100), where circles represent original data points and solid lines denote fitted curves.

Since the B-spline fit of the data captures the F_0 trend of a sentence reasonably well, it may be feasible to use the 14 corresponding coefficients (denoted by *intercept* and *bs1* – *bs13*) together with *sex* (the speaker's gender), *tone*, and *focus* as the input feature vector for each sentence in constructing decision trees. But first we need to explore the nature of B-spline regressions and the meaning of the obtained B-spline coefficients.

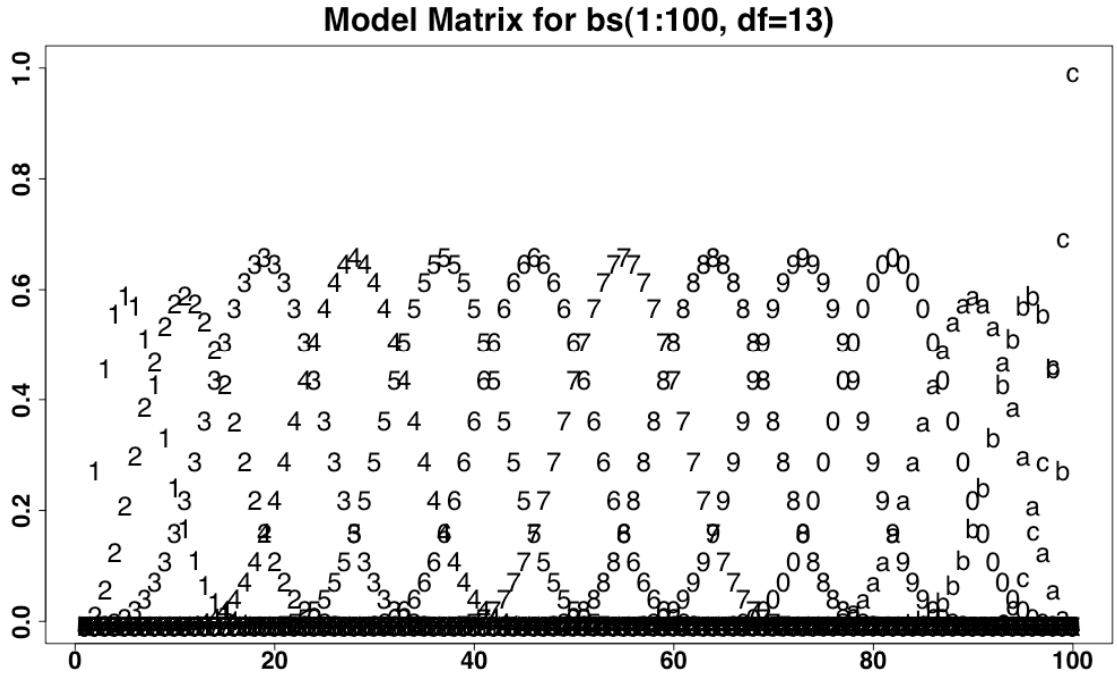

Figure 3.6. Plot of the model matrix for the cubic B-spline regression of F_0 on time (1 - 100) with 10 interior knots at time points 10, 19, 28, 37, 46, 55, 64, 73, 82, and 91.

A basis can be generated by the *bs()* function in R for the space (here, 14-dimensional) of spline functions given a polynomial degree (here, 3) and a set of knots (here, 10) with specific placements (here, uniformly placed), and the *bs* coefficients reported are for such a basis (Hastie, 1992). Figure 3.6 displays the basis (or model matrix) for the B-spline expression *bs(time, df = 13)*, with columns *1 – 9* corresponding to coefficients *bs1 – bs9* and columns *0, a, b, c* corresponding to coefficients *bs10 – bs13*. The supports of the thirteen columns (thus the *bs* coefficients) in terms of *time* (1 – 100) are summarized in Table 3.1.

Table 3.1. The supports of the thirteen columns (and thus *bs1 – bs13*) in the model matrix (see Figure 3.6).

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*0*	*a*	*b*	*c*
bs1	*bs2*	*bs3*	*bs4*	*bs5*	*bs6*	*bs7*	*bs8*	*bs9*	*bs10*	*bs11*	*bs12*	*bs13*
2-18	2-27	2-36	11-46	20-54	28-63	38-72	47-81	55-91	65-99	74-99	83-99	92-100

As can be seen in Table 3.1, the supports of the thirteen *bs* coefficients are overlapped across adjacent regions. Nonetheless, *bs1 – bs5* mainly cover the time period of syllables 1 – 5 (as each syllable has 10 time points), *bs6 – bs7* mainly cover that of syllables 4 – 7, and *bs8 – bs13* mainly cover that of syllables 6 – 10.

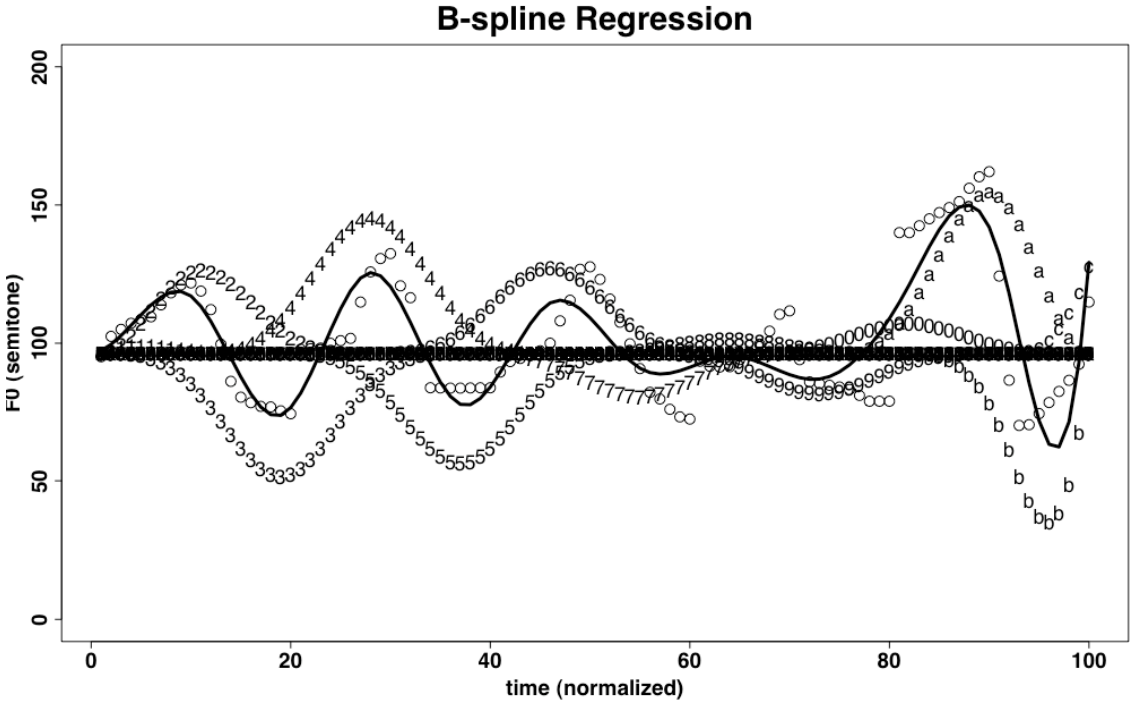

Figure 3.7. An example of the fitted curve (solid line) under the cubic B-spline regression, where black circles denote original data points, and the curves connected by *1 – 9*, *0*, *a*, *b*, and *c* are computed from '*intercept + bs coefficient × model matrix column*'.

Figure 3.7 gives an example of the fitted curve under the cubic B-spline regression of F_0 on *time*, from which we can see that the *intercept* approximates the average F_0 level of a sentence, and *bs1 – bs13* measure something essentially local about

the original data. The exact B-spline fitting is obtained by adding up the values of *'intercept + bs coefficient × model matrix column'* that are overlapped with one another. Therefore, positive *bs coefficients* tend to give rise to local peaks, and negative *bs coefficients* usually lead to local valleys in the trajectory of the data points.

With the background information of B-spline regressions in mind, I now proceed to construct decision trees using the 14 *bs coefficients* (*intercept* and *bs1 – bs13*), *sex*, *tone*, and *focus* as the input feature vector for the automatic classification of statement and question intonation in Mandarin.

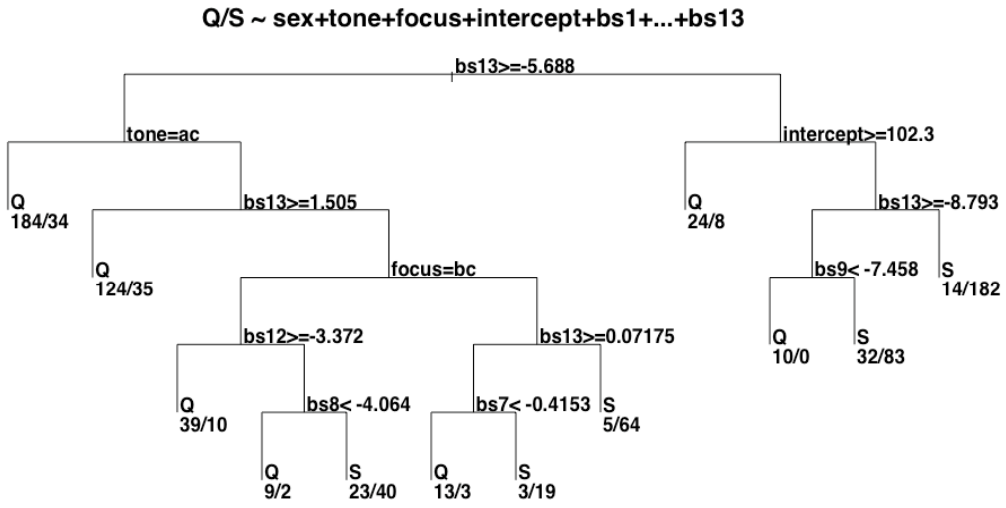

Figure 3.8. Classification tree of sentence type (Q: question vs. S: statement) on sex, tone, focus, and the 14 coefficients from B-spline regressions for individual sentences.

The decision tree (grown on the training set) that utilized *sex*, *tone*, *focus*, *intercept*, and *bs1 – bs13* as the input vector is shown in Figure 3.8. As can be seen, only *tone*, *focus*, *intercept*, *bs7*, *bs8*, *bs9*, *bs12*, and *bs13* are actually used in the tree construction. Sentences are first split depending on whether *bs13* is greater than or equal to -5.688. If so, they are split according to *tone* being Falling/Low or High/Rising; if not,

they are again split according to *intercept* being greater than or equal to 102.3. On the left branch, sentences with Falling/Low tones are classified as questions, of which 184 are indeed questions but 34 are actually statements; for sentences with High/Rising tones, *bs13 >= 1.505* becomes another criterion for further splitting (details omitted). On the right branch, sentences having *intercept >= 102.3* are grouped into questions with probability 0.75 (=24/32); those having *intercept < 102.3* are further split according to *bs13 >= -8.793* or not.

The relationship between sentence type and the B-spline coefficients selected by the above tree seems compatible with the findings in Experiment 1, since the sequential importance of *bs13, bs12, bs9, bs8* and *bs7* agrees with the fact that the difference between statements and questions becomes more and more pronounced as the sentences approach their end. Prediction of sentence type based on this tree for the testing set gives correct classification rate of 82.19% (= 263/320).

The relative differences between *bs13* and *bs1 – bs12* (denoted by *bs13_1*, *bs13_2*, … , *bs13_12*) also perform well in distinguishing statements from questions (with classification rate 83.44% (= 267/320) on the testing set), as demonstrated in Figure 3.9. The importance of *bs13_7,* which alone successfully singles out 233 questions, indicates that many questions begin to differ from statements from the middle of the sentence. 169 statements having *bs13_10 < -3.098* shows that the pitch contour difference in the later part of a sentence is also significant in determining its sentence type. Furthermore, for sentences under initial and medial focus, there are 90 questions having *bs13_3 >= -4.147* and *bs13_1 >= -10.58*. For those sentences under initial and

medial focus but having *bs13_3 < -4.147* and *bs13_1 >= -10.58*, *intercept >= 102.4* determines further partition. This seems to agree with the findings in Experiment 1 about the interaction between sentence type and focus: in initial- and medial-focused sentences, the pitch contours of statements and questions start to diverge from the focused word.

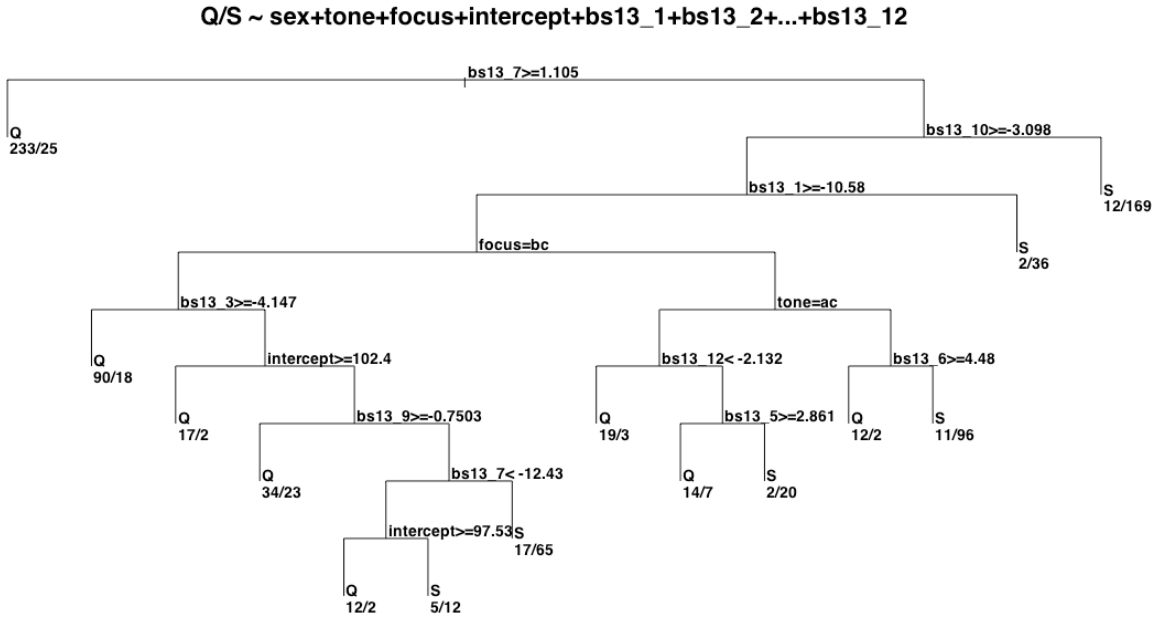

Figure 3.9. Classification tree of sentence type (Q: question vs. S: statement) on sex, tone, focus, intercept, and the differences between *bs13* and *bs1 – bs12* from B-spline regressions for individual sentences.

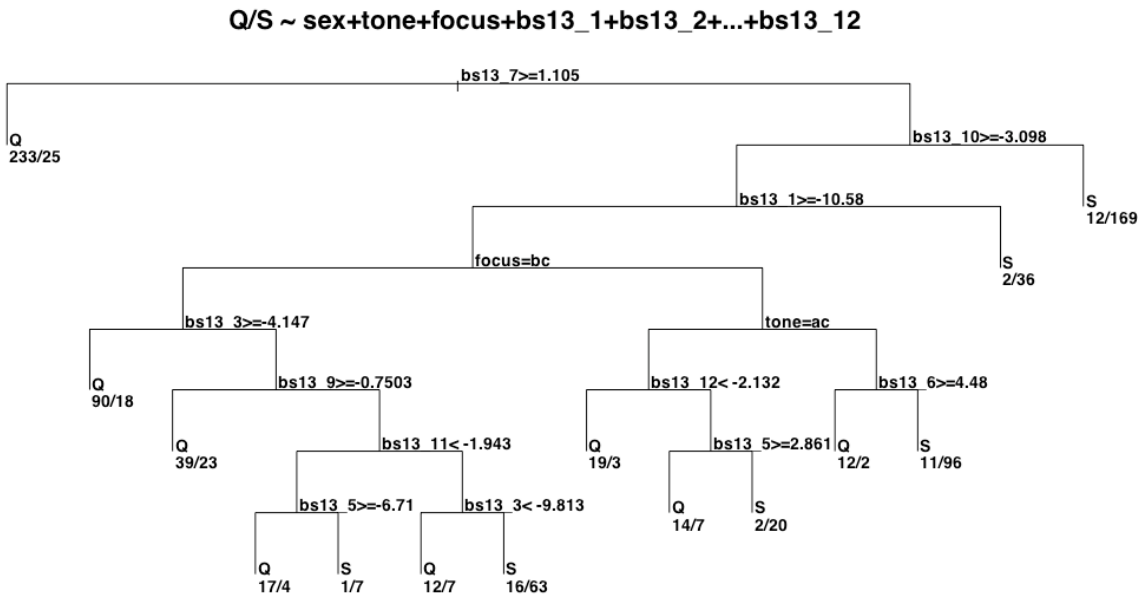

Figure 3.10. Classification tree of sentence type (Q: question vs. S: statement) on sex, tone, focus, and the differences between *bs13* and *bs1 – bs12* from B-spline regressions for individual sentences.

To test how much the overall F_0 level of a sentence (approximated by the intercept in B-spline regressions) contributes to the differentiation of statements from questions when the relative differences between the later and earlier portion of the sentence (*bs13_1, bs13_2, ... , bs13_12*) are taken into consideration, another tree is drawn in Figure 3.10, with the intercept excluded from the input vector. Figures 3.9 and 3.10 differ not only in their lower left branch (the presence/absence of the intercept), but also in their prediction rate of sentence type on the testing set (83.44% = 267/320 for the former vs. 80.63% = 258/320 for the latter). The important role that the intercept plays here suggests that B-spline coefficients are meaningful and comparable only when they are combined together for each sentence, considering the fact that the fitted curve is the result of adding up overlapped '*intercept* + *bs coefficient* × *model matrix column*' in each

knot span of the sentence (see Figure 3.7). In addition, the value of the intercept also depends on the sex, tone, focus, and sentence type condition of a sentence, which probably makes it an indispensable predictor of sentence type. I thus include the intercept in later tree constructions when B-spline coefficients are involved.

3.3.2 Original final F_0's of the 10 syllables

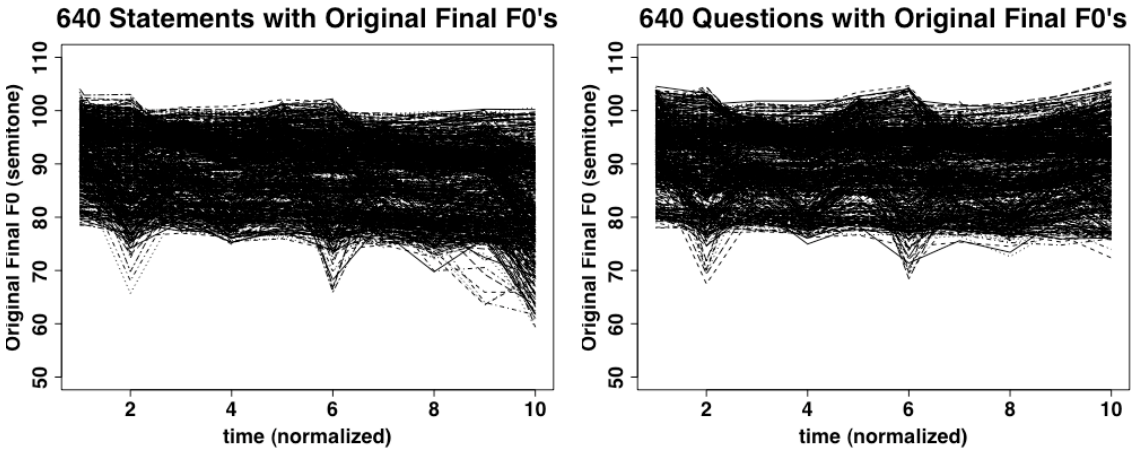

Figure 3.11. Pitch trajectories of individual sentences represented by the original final F_0 (in semitones) of each syllable.

Due to articulatory constraints, the underlying pitch target of a tone is most fully realized in its final region (Xu & Wang, 2001). Therefore, final F_0's of the syllables may largely represent the global pitch trend of a sentence. Pitch trajectories of individual statements and questions represented by the original final F_0 of each syllable are displayed in Figure 3.11. The large variability seen in the figure is due to this feature set not being normalized for speaker, tone, or focus as in section 3.3.3.

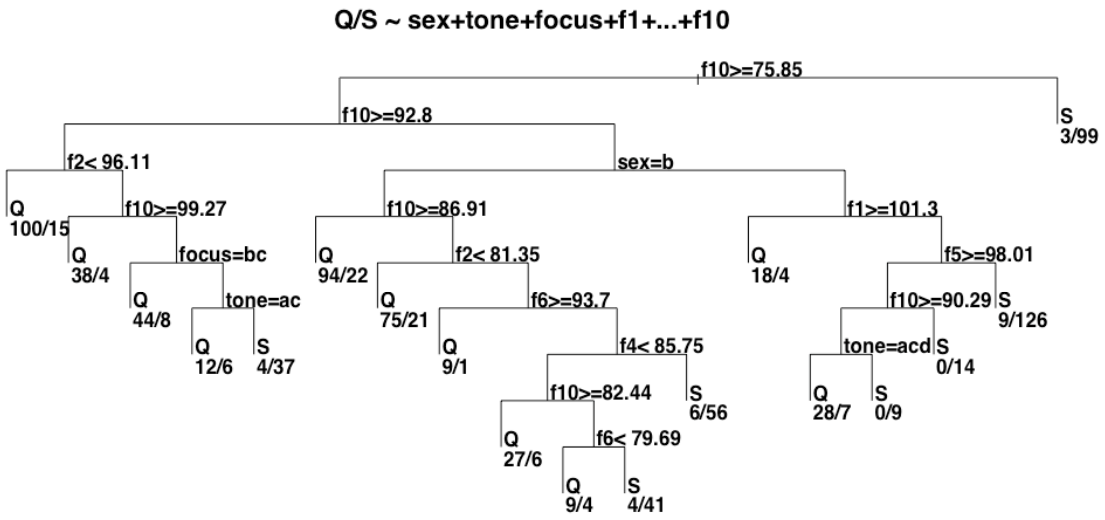

Figure 3.12. Classification tree of sentence type (Q: question vs. S: statement) on sex, tone, focus, and original final F_0's (denoted by *f1 – f10*, in semitones).

The classification tree using *sex*, *tone*, *focus*, and original syllable-final F_0's (*f1 – f10* for the 10 syllables) as the input feature vector was again grown on the same training set as in section 3.3.1. As shown in Figure 3.12, all predictors except *f3*, *f7*, *f8* and *f9* are included in the tree. Sentences are first split depending on whether *f10* (final F_0 of the last syllable) is greater than or equal to 75.85. If not, they are classified as statements (with 3 of them misclassified); if so, they are again split according to *f10* being greater than or equal to 92.8. For sentences with *f10* less than 92.8, either *f10 >= 86.91* or *f1 >= 101.3* determines further splitting of the tree depending on whether they were produced by male or female speakers. For sentences with *f10* greater than or equal to 92.8, *f2 < 96.11* becomes another criterion for differentiating questions from statements.

Prediction of sentence type based on the above tree for the testing set gives correct classification rate of 80.00% (= 256/320), which is slightly worse than the trees in

section 3.3.1 for which *sex*, *tone*, *focus*, and B-spline coefficients (*the intercept*, *bs1 – bs13,* or their relative differences *bs13_1, ..., bs13_12*) served as predictors.

The relative differences (denoted by *f10_1*, *f10_2*, … , *f10_9*) between *f10* and *f1 – f9* perform better than the 10 individual final F_0 values in differentiating statements and questions (with classification rate 82.19% (= 263/320) on the testing set), as shown in Figure 3.13.

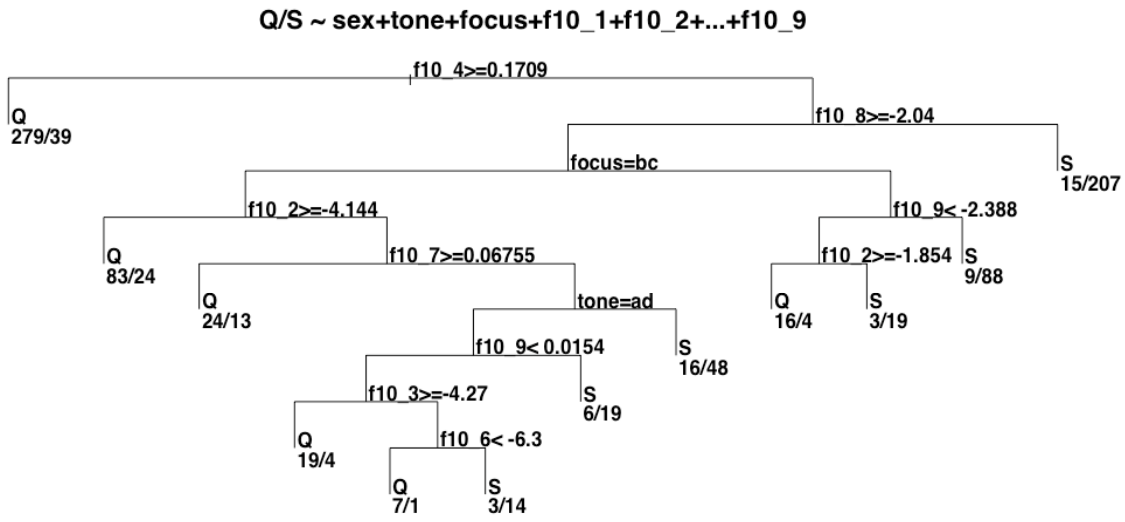

Figure 3.13. Classification tree of sentence type (Q: question vs. S: statement) on sex, tone, focus, and the differences between original *f10* and *f1 – f9* (in semitones).

Figure 3.13 resembles Figure 3.9 in the sense that sentences are first categorized according to F_0 differences in their later part: as questions (with 279 being correct and 39 wrong) if *f10_4 >= 0.1709*, and as statements (207 correct and 15 wrong) if *f10_4 < 0.1709* and *f10_8 < -2.04*. In addition, sentences under initial/medial focus are treated differently from those under final/neutral focus, with the former being partitioned by *f10_2 >= -4.144* and the latter by *f10_9 < -2.388*.

In summary, sections 3.3.1 and 3.3.2 indicate that information conveyed by the more detailed pitch contour representation (B-spline coefficients) can be approximately captured by the syllable level representation (syllable-final F_0's), and that the differences between the sentence-final element (*bs13* or *f10*) and the earlier elements (*bs1 – bs12* or *f1 – f9*) work better than the individual elements in classifying statements and questions. However, the drawback of the above four classification trees is that there exists collinearity between the factors (*sex*, *tone*, and *focus*) and the numeric predictors (*the intercept*, *bs1 – bs13*, *b13_1 – bs13_12*, *f1 – f10*, and *f10_1 – f10_9*) in the model formula. For example, as shown in Experiment 1, the pitch contour of a sentence is modulated severely by the effects of *focus*: the pitch range of the focused words is expanded, that of the post-focus words compressed and lowered, and that of the pre-focus words largely unaffected. Likewise, F_0 of a syllable varies differently depending on its tone. Therefore, it is not entirely clear which mechanism (*sentence type* vs. *tone/focus*) should be considered as having taken effect when interpreting the partition rules based on the values of the numeric predictors.

3.3.3 Normalized final F_0's of the 10 syllables

As seen in Figure 3.1d, F_0 contours of the neutral-focused sentences with all High tones reflect most directly the nonlinear increase in the difference between statements and questions along the time axis. In a sense, the difference pattern there is largely free of tone and focus effects. Therefore, it may help to remove these potentially confounding effects by transforming the F_0 contours under other conditions toward those under High-tone and neutral-focus condition through the following normalization method. Note that

the *speaker*, *tone*, and *focus* conditions of each sentence are assumed to be known (via preprocessing of the data) before applying the following normalization technique on each syllable.

Suppose that μ_{stf} and σ_{stf} are the mean and standard deviation of final F_0's of the syllables by speaker s (1, 2, …, 8), with tone t (1: High, 2: Rising, 3: Low, 4: Falling), and under focus condition f (1: pre-focus, 2: post-focus, 3: initial/medial focus, 4: final focus), final F_0 (x_{stf}) of each syllable under such speaker/tone/focus condition (*stf*) is standardized to z_{stf} in equation (1). Based on the findings in Experiment 1, all syllables in neutral-focused sentences are treated as pre-focus, and syllables under final focus are treated as different from those under initial/medial focus.

$$z_{stf} = (x_{stf} - \mu_{stf}) / \sigma_{stf} \quad (1)$$

Then, to remove the effects of *speaker*, *tone*, and *focus*, z_{stf} with $s \neq 1$, $t \neq 1$, and $f \neq 1$ is normalized to become y_{stf}, where

$$y_{stf} = z_{stf} \bullet \sigma_{111} + \mu_{111} = ((x_{stf} - \mu_{stf}) / \sigma_{stf}) \bullet \sigma_{111} + \mu_{111} \quad (2)$$

Or, equivalently,

$$z_{stf} = (y_{stf} - \mu_{111}) / \sigma_{111} \quad (3)$$

That is, for each speaker $s \neq 1$, assuming that the distribution of his/her syllables' final F_0's for any fixed tone (t) and focus (f) condition is Gaussian, then normalization can be viewed as mapping all his/her syllables' final F_0's to another Gaussian distribution (in this case $s = 1$, $t = 1$, and $f = 1$, i.e., the distribution of speaker 1's High-tone pre-focus syllables).

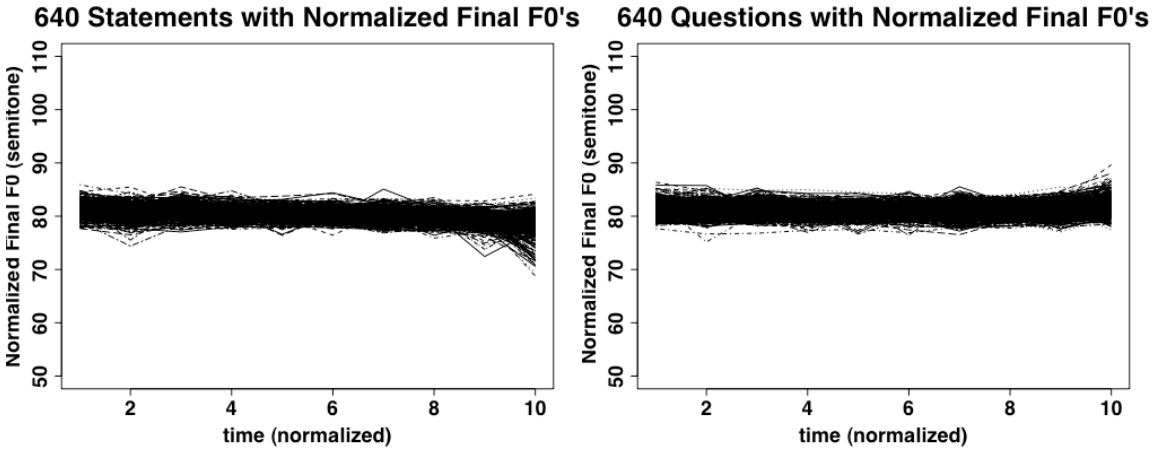

Figure 3.14. Pitch trajectories of individual sentences represented by the normalized final F_0 (in semitones) of each syllable.

The effects of *speaker*, *tone*, and *focus* on F_0 trajectories of the sentences in the training set are largely removed in Figure 3.14, where statements can be seen to have a gradually falling contour, and questions a rising contour.

As shown in the following classification tree (Figure 3.15) based on normalized final F_0's (denoted by *f1 – f10*) in the training set, among the 10 syllables in each sentence, only *f10* and *f7* are employed in the tree construction. The split on *f10 >= 80.47* partitions the 960 observations into groups of 447 and 513 individuals, with probability of question equal to 0.9105 (= 407/447) and 0.1423 (= 73/513), respectively. This second group is then partitioned into groups of 117 and 396 individuals, depending on whether *f10* is greater than 79.7 (inclusive) or not. The former group is subdivided into groups of 52 and 65 individuals, depending on whether or not *f7* is greater than or equal to 80.51, with probabilities of question equal to 0.7115 (= 37/52) and 0.2308 (= 15/65), respectively. The latter group is classified as statement with probability 0.9470 (= 375/396).

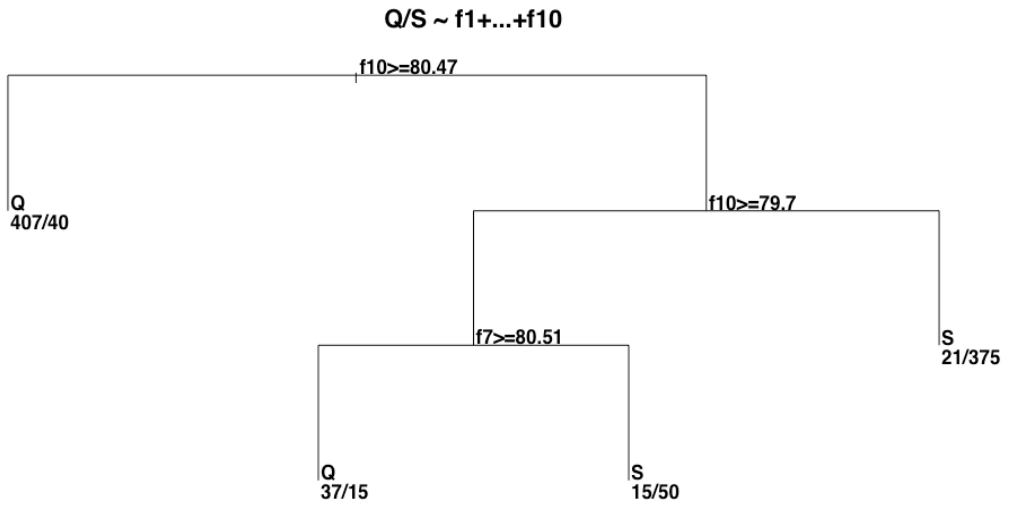

Figure 3.15. Classification tree of sentence type (Q: question vs. S: statement) on normalized final F_0's (denoted by *f1 – f10*, in semitones).

Prediction of sentence type based on the above tree for the testing set gives correct classification rate of 85.31% (= 273/320), which is better than those obtained from the tree models in sections 3.3.1 and 3.3.2.

To test if the normalization is flawless, the factors *sex*, *tone*, and *focus* are added to the list of potential predictors and the resulting tree is shown in Figure 3.16. As can be seen, *focus*, *f6*, *f9* and *f10* are incorporated in the tree construction, which achieves a classification rate of 87.19% (= 279/320) on the testing set. The reason that sentences under initial/medial focus are further split according to the rules *f9 >= 80.02* and *f6 >= 81.98* is because post-focus suppression (pitch lowering for the post-focus words) is greater in statements than in questions, as shown in Experiment 1. On the other hand, sentences under final/neutral focus and having *79.7 =< f10 < 80.47* are classified as statements because the F_0 value of the last word in a final/neutral-focused statement is usually high (Experiment 1).

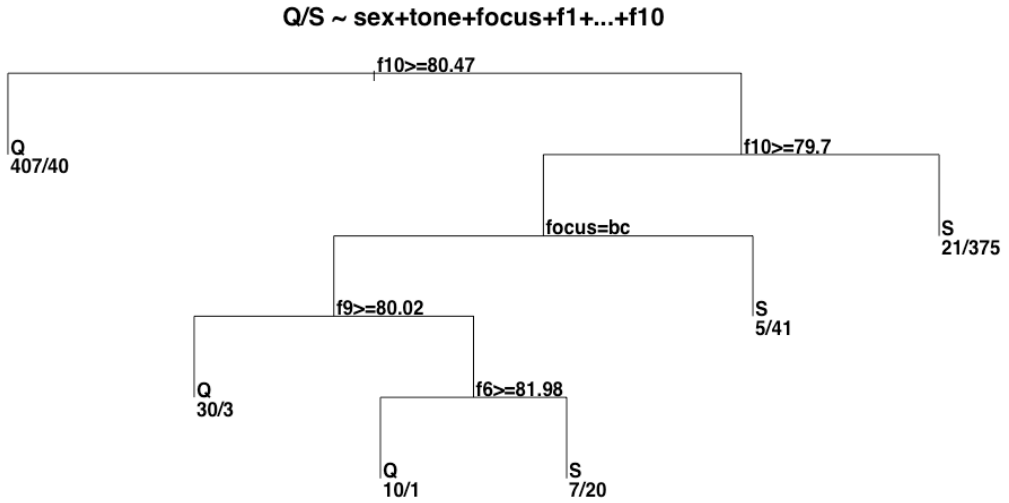

Figure 3.16. Classification tree of sentence type (Q: question vs. S: statement) on sex, tone, focus, and normalized final F_0's (*f1 – f10*, in semitones).

Figure 3.16 indicates that the effect of *focus* is not entirely eliminated through normalization, while those of *speaker* (thus *sex*) and *tone* are largely removed. One possible reason for the difficulty of removing the focus effect is that pre-focused syllables do not behave exactly the same as syllables under neutral focus condition (pitch range suppression is found in some speakers' pre-focused syllables, and sentence-final neutral-focused syllables usually have expanded pitch ranges). Nonetheless, all syllables in neutral-focused sentences are treated as pre-focused in normalization.

Figure 3.17. Classification tree of sentence type (Q: question vs. S: statement) on the differences between normalized *f10* and *f1 – f9* (in semitones).

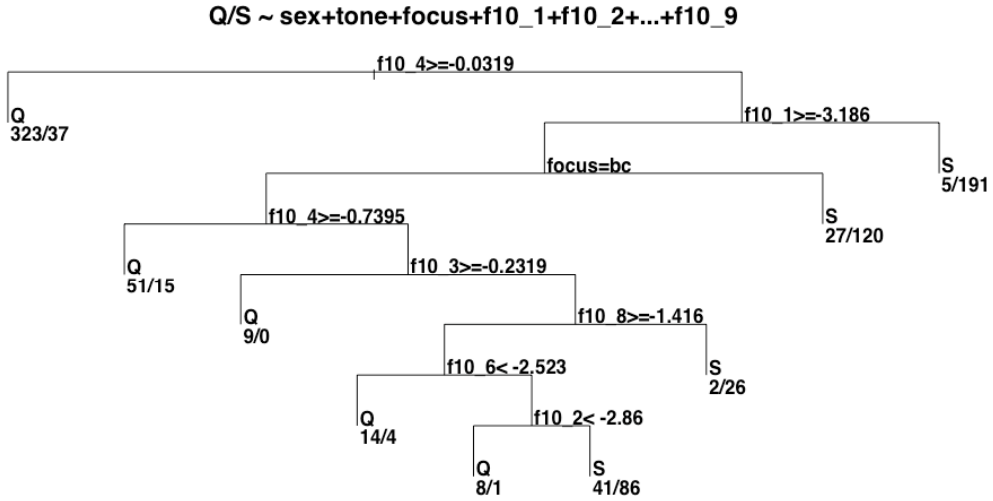

Figure 3.18. Classification tree of sentence type (Q: question vs. S: statement) on sex, tone, focus, and the differences between normalized *f10* and *f1 – f9* (in semitones).

Figures 3.17 and 3.18 display the trees grown on the differences (denoted by *f10_1, f10_2, ..., f10_9*) between the normalized *f10* and *f1 – f9*, with the latter having *sex*, *tone* and *focus* as possible predictors as well. Both trees give the same classification rate of 82.81% (= 265/320) when applied to the testing set. The inclusion of the

predictors in the tree constructions (*f10_1, f10_2, f10_3, f10_4, f10_8, f10_9* in Figure 3.17, and *focus, f10_1, f10_2, f10_3, f10_4, f10_6, f10_8* in Figure 3.18) again verifies the findings about the interaction between focus and sentence type in Experiment 1. That is, the difference between statements and questions becomes increasingly salient over time (thus the sequential importance of *f10_4* and *f10_1*) especially when the focus is final or neutral, and the post-focus suppression has greater magnitude in statements than in questions under initial/medial focus. However, unlike in sections 3.3.1 and 3.3.2, the differences between the last element and the earlier elements in this section did not yield better performance than individual elements themselves in predicting sentence type. This is probably because the normalization technique employed here has reduced the actual differences between the last element and the earlier elements in a sentence.

3.3.4 Summary of the classification trees in sections 3.3.1 – 3.3.3

Table 3.2. Summary of the classification trees in sections 3.3.1 – 3.3.3.

Decision trees		Variables used	Correct classification rate in testing set
3.3.1. B-spline coefficients	individual	*focus, tone, intercept, bs7, bs8, bs9, bs12, bs13*	82.19% (=263/320)
	difference	*focus, tone, intercept, bs13_1, bs13_3, bs13_5, bs13_6, bs13_7, bs13_9, bs13_10, bs13_12*	83.44% (=267/320)
3.3.2. Original final F_0's	individual	*focus, tone, sex, f1, f2, f4, f5, f6, f10*	80.00% (=256/320)
	difference	*focus, tone, f10_2, f10_3, f10_4, f10_6, f10_7, f10_8, f10_9*	82.19% (=263/320)
3.3.3. Normalized final F_0's	individual	*f7, f10*	85.31% (=273/320)
		focus, f6, f9, f10	87.19% (=279/320)
	difference	*f10_1, f10_2, f10_3, f10_4, f10_8, f10_9*	82.81% (=265/320)
		focus, f10_1, f10_2, f10_3, f10_4, f10_6, f10_8	82.81% (=265/320)

Table 3.2 summarizes the results from the eight classification trees (for consistency, the tree in Figure 3.10 is excluded in the summary) implemented in the above three sections. Section 3.3.1 to 3.3.2 shows that B-spline coefficients obtained from detailed pitch contour fitting are not much better than original syllable-final F_0's in characterizing and differentiating statements and questions. Section 3.3.2 to 3.3.3 indicates that F_0 normalization in preprocessing (rather than adding *sex*, *tone* and *focus* as additional factors in tree modeling) works well. The parsimony and interpretability of the tree in Figure 3.15 of section 3.3.3, where only normalized *f7* and *f10* are in effect for the classification of statements and questions, is worth special mention. Furthermore, the use of relative differences between the last element and the earlier elements helps to improve the performance of the classification trees when B-spline coefficients and original final F_0's are employed as input variables.

An important aspect of the testing set in this section is that perception results of statement/question classification by human subjects are available in Experiment 2 based on the same set of examples. There are very few tasks in speech recognition where human and machine performance can be compared, therefore, it is worth noticing that the machine in this study achieved 87% classification accuracy (the highest among all) when humans attained 89%.

3.4 Cross-validation

To compare the above three different approaches in a reliable way, cross-validation is done to rotate the speakers used for training and for testing. Three more trees were fitted in each of the eight cases (again, excluding the case in Figure 3.10) so that different sets of testing data (each by 1 male and 1 female speaker) were used in calculating classification rates. Table 3.3 lists the results obtained through cross-validation, with those from the previous section also included.

Table 3.3. Summary of the cross-validation results.

Decision trees		Variables used	Classification rate	Mean (std)
B-spline coefficients	Individual	*focus, tone, intercept, bs7, bs8, bs9, bs12, bs13*	82.19% (=263/320)	77.27% (4.87)
		focus, intercept, bs4, bs7, bs8, bs11, bs12, bs13	70.94% (=227/320)	
		focus, tone, intercept, bs3, bs6, bs10, bs12, bs13	76.25% (=244/320)	
		focus, tone, intercept, bs1, bs6, bs7, bs9, bs12, bs13	79.69% (=255/320)	
	Difference	*focus, tone, intercept, bs13_1, bs13_3, bs13_5, bs13_6, bs13_7, bs13_9, bs13_10, bs13_12*	83.44% (=267/320)	79.77 % (3.67)
		focus, bs13_1, bs13_2, bs13_3, bs13_5, bs13_7, bs13_9, bs13_10, bs13_11, bs13_12	74.69% (=239/320)	
		focus, tone, intercept, bs13_1, bs13_4, bs13_5, bs13_8, bs13_9	80.31% (=257/320)	
		focus, tone, intercept, bs13_1, bs13_2, bs13_5, bs13_6, bs13_7, bs13_8, bs13_11	80.63% (=258/320)	
Original final F_0's	Individual	*focus, tone, sex, f1, f2, f4, f5, f6, f10*	80.00% (=256/320)	73.67% (7.53)
		focus, tone, sex, f3, f4, f5, f8, f10	62.81% (=201/320)	
		focus, tone, sex, f4, f5, f6, f7, f10	76.88% (=246/320)	
		focus, f1, f2, f4, f5, f6, f7, f9, f10	75.00% (=240/320)	
	Difference	*focus, tone, f10_2, f10_3, f10_4, f10_6, f10_7, f10_8, f10_9*	82.19% (=263/320)	80.31% (1.89)
		focus, f10_2, f10_4, f10_5, f10_6, f10_8	77.81% (=249/320)	
		focus, tone, f10_2, f10_4, f10_6, f10_8, f10_9	81.25% (=260/320)	
		focus, tone, f10_2, f10_3, f10_4, f10_6, f10_8, f10_9	80.00% (=256/320)	

Table 3.3. (*continued*)

Decision trees		Variables used	Classification rate	Mean (std)
Normalized final F_0's	Individual	*f7, f10*	85.31% (=273/320)	86.02% (1.83)
		f1, f9, f10	83.75% (=268/320)	
		f5, f9, f10	87.50% (=280/320)	
		f3, f6, f10	87.50% (=280/320)	
		focus, f6, f9, f10	87.19% (=279/320)	87.50% (2.81)
		f1, f9, f10	83.75% (=268/320)	
		focus, f5, f9, f10	90.31% (=289/320)	
		focus, f10	88.75% (=284/320)	
	Difference	*f10_1, f10_2, f10_3, f10_4, f10_8, f10_9*	82.81% (=265/320)	79.77 % (2.17)
		f10_1, f10_2, f10_4, f10_7, f10_8,	77.81% (=249/320)	
		f10_4	78.75% (=252/320)	
		f10_4	79.69% (=255/320)	
		focus, f10_1, f10_2, f10_3, f10_4, f10_6, f10_8	82.81% (=265/320)	81.10% (1.74)
		focus, f10_1, f10_4, f10_6, f10_8	81.88% (=262/320)	
		f10_4	78.75% (=252/320)	
		focus, f10_1, f10_3, f10_4, f10_6	80.94% (=259/320)	

The above table further confirms that normalized individual final F_0's give the best performance in the classification of statements and questions in Mandarin using decision trees.

3.5 Discussion

As a screening method for selecting predictors, classification trees were used in this chapter to extract the most useful information in an utterance that characterizes its

sentence type. Three different sets of feature vectors were input into the trees and the corresponding results are summarized as follows.

The coefficients from B-spline regressions are reasonably good in classifying statements and yes/no questions in Mandarin, as demonstrated by the mean classification rates through cross-validation (77.27% for individual elements and 79.77% for the differences between the last element and the earlier elements). However, the fact that they performed no better than non-normalized original final F_0's of the syllables shows that parameters obtained from direct representation of detailed surface F_0 have limited benefit for sentence type classification. In contrast, the much larger improvement brought about by the normalization method (85.31%, with normalized *f7* and *f10* as predictors) demonstrates the importance of directly taking into account the effects of *speaker*, *tone* and *focus*. This performance is only slightly worse than the human performance (89.12%) reported in Experiment 2 for the speech of the same two subjects in the testing set.

In all the classifications, the parameters corresponding to the sentence-final F_0 are found to be the dominant factor for determining sentence type. Nevertheless, F_0 before the final syllable are also consistently found to be relevant. This agrees with the finding in Experiment 1 that, across all tone and focus conditions, the difference in F_0 between statement and question intonation is characterized by a nonlinear increase toward the end of the sentence.

The decision trees in this chapter are grown on a dataset in which statements and questions are elicited under laboratory conditions with tone and focus systematically controlled. The performances therefore are not equivalent to those on natural speech

databases. Importantly, in natural speech, many questions do not have rising intonation while many statements do (Hirschberg, 2000; Shattuck-Hufnagel & Turk, 1996), which presents an issue that is beyond the scope of the current chapter. In those cases, however, it is also an open question whether human listeners can identify the questions and statements that are taken out of context. What the current results show is that for those cases in which human listeners can make the identification, normalized syllable-final F_0's can achieve similar performance.

In this chapter, the normalization of syllable-final F_0's is made possible via data preprocessing. In cases where no manual processing (segmentation of syllables, categorization of tones, labeling of focus, etc.) of the data is available, B-spline regressions might be used as one of the automatic ways to represent the pitch contour of a sentence. The performance of such technique (compared to other possible techniques) on the classification of statement and question intonation needs to be investigated in future research.

Finally, the effectiveness of syllable-final F_0's in representing simplified F_0 contours of statements and questions has important implication for their role in signifying pitch targets of the syllables. In the following two chapters, velocities of syllable-final F_0's will be used as indicators of the pitch targets of the syllables in both Mandarin and English.

4 THE INTERACTION OF LEXICAL TONE/STRESS, FOCUS, AND SENTENCE TYPE IN MANDARIN AND ENGLISH

As mentioned in section 2.4, the encoding scheme of question intonation in English may differ from that in Mandarin, presumably because the former is a non-tonal language and the latter a tone language. This chapter explores the similarities and differences between the two languages in the way that lexical item, focus, and sentence type interact with one another. Partial results of this chapter were reported in Liu & Xu (2007a, 2007b).

4.1 Introduction

Languages of the world are often grouped into two types: those that use pitch to distinguish words at the level of the syllable, which are known as tone languages, and those that use pitch mainly to distinguish non-lexical meanings, which are known as intonational languages (Chao, 1968; Ladd, 1996; Yip, 2002). But there is increasing evidence that the distinction between the two types is not as clear as this simple divide may suggest. Some tone languages have been found to use pitch to also convey meaningful contrasts at a sentential level (Chang, 1958; Herman, 1996; Inkelas & Leben, 1990; Lindau, 1986; Liu & Xu, 2005; Ma et al., 2004; Selkirk & T. Shen, 1990; J. Shen, 1985; X.-N. S. Shen, 1990; Xu, 1999). And some intonational languages have been found to either use pitch to also make lexical contrasts (Fry, 1958; Xu & Xu, 2005), or exhibit intonational pitch events that are more tightly aligned to the syllable than previously assumed (Arvaniti, Ladd & Mennen, 1998; Atterer & Ladd, 2004; Ladd, Mennen & Schepman, 2000; Prieto & Torreira, 2007; Xu & Xu, 2005). As pointed out by Goldsmith (1981: 289), “our ideas about the less exotic languages which we call accentual can be

enriched by viewing them in the context of tone languages." Pierrehumbert (1980) developed this line of thinking into a full blown framework, known as the autosegmental-metrical (AM) theory of intonation, in which the intonation of English is composed of tones as the basic building blocks for the contrastive melodic events such as pitch accents and boundary tones. Xu and Xu (2005) went a step further to show that the parallel between the two types of languages can be seen even at the level of the syllable. The current chapter is a continuation of this line of research by comparing Mandarin and English, two languages that are widely recognized as clearly falling on two sides of the tonal divide, to examine in fine detail how they each use pitch patterns to convey the question/statement contrast together with several other communicative functions, including lexical tone (Mandarin), lexical stress (English) and focus (both languages).

4.1.1 The interaction of sentence type with tone and focus in Mandarin

When studying question intonation in Mandarin, one cannot escape the fact that every sentence is composed of words whose lexical meanings are distinguished not only by segmental but also by tonal components. There are four full lexical tones and a neutral tone in Mandarin. On a five–point scale (Chao, 1930), the High tone in citation form is transcribed as /55/, the Rising tone /35/, the Low tone /214/, and the Falling tone /51/. The pitch of the neutral tone is described as determined by its preceding tone: high after the Low tone, and relatively low after the other tones (Chao, 1933, 1968). Chen and Xu (2006), however, argued that the neutral tone has a static mid target, and that it is implemented with weak articulatory strength.

Although the basic characteristics of the four full tones in connected speech are largely retained, intonation and sentence environment do modify their register and contour characteristics: 1) The overall pitch values of the full tones in sentence-final positions are higher in questions than in statements (Ho, 1977; Liu & Xu, 2005; Ni & Kawai, 2004; Rumjancev, 1972; X.-N. S. Shen, 1990; Yuan, 2004; Zeng, et al., 2004). 2) The High tone falls moderately at the final position of a declarative sentence (Ho, 1976a, 1976b), but rises in an interrogative sentence (Liu & Xu, 2005; X.-N. S. Shen, 1990; Zeng, et al., 2004). 3) The steepness of the slope of the Rising and Falling tone of the final syllable differs in questions and statements (Rumjancev, 1972), with the Rising tone steepened (Yuan, 2004) and widened (Ni & Kawai, 2004; X.-N. S. Shen, 1990; Zeng, et al., 2004) and the Falling tone flattened (Yuan, 2004; Zeng, et al., 2004) and narrowed (Ni & Kawai, 2004) in questions. 4) According to Lin (2004), the sentence-final Low tone is realized as falling in statements but falling-rising in questions. However, in Ni and Kawai (2004) and Zeng, et al. (2004), the pitch realization of the final Low tone is falling-rising in both sentence types, except that the rising ending is higher in questions than in statements. The data in Experiment 1 (Liu & Xu, 2005) show that in a statement-final position only focused Low is realized as falling-rising, while non-focused Low is simply falling. In contrast, the question-final Low tone is realized as falling-rising no matter whether it is focused or not.

The issue of tone-intonation interaction becomes even trickier when the neutral tone is involved. If indeed targetless as traditionally viewed (Chao, 1933, 1968), the F_0 contour of the neutral tone at the final position of a question should fully or at least

largely reflect the interrogative intonation. Previous research, however, has produced mixed results. In Chao's impressionistic account (1933), in the sentence "*nǐ* /21/ *shuō* /55/ *shén* /35/ *me* /2/ *lai* /1/ *zhe* /1/? (*What did you say?*)", the last three neutral tones are realized with a low pitch if the question is newly-uttered, but with a high pitch ("*nǐ* /21/ *shuō* /55/ *shén* /35/ *me* /5/ *lai* /5/ *zhe* /5/?") if the same question is asked again. According to Qi (1956), however, the neutral-tone question-particle *ma* has a high pitch /5/ no matter what the preceding full tone is; and with a sequence of three neutral tones plus *ma* at the end of a question, their pitch can be transcribed as /4 4 4 5/ regardless of what the preceding full tone is. Upon observing that the question-final neutral-tone particle *ma* has a falling contour after the High and Falling tone, a rising contour after the Low tone, and a falling/level/rising contour after the Rising tone (similar results were presented in section 2.2.2.5 of this dissertation), X.-N. S. Shen (1990:48) claimed that the tonal value of the sentence-final interrogative particles is "the algebraic sum of the F_0 value of the preceding tone and the sentence intonation", and that they "always end on a high key". In the Pan-Mandarin ToBI system (Peng, et al., 2005), the sentence-final particle *ma* can carry two possible boundary tones: H% in questions and L% in statements. However, Lee (2005:122) proposed that "the *ma*-particle does not stand as an independent prosodic unit that could bear the terminal rise" in marked questions.

In summary, previous research on Mandarin has not reached a consensus on the pitch pattern of the neutral tone in interrogative sentences, and the picture presented is far from complete. There was no explicit indication of how the preceding tone and the sentence intonation interact to generate the F_0 contour of a question-final neutral tone in

the algebraic sum hypothesis in X.-N. S. Shen (1990). The limited examples in Peng, et al. (2005) and Lee (2005) also failed to present a full picture of the influence of different preceding tones on the neutral tone in different intonations. In particular, the following issues need to be resolved regarding the neutral tone in question intonation in Mandarin: (1) How is the neutral tone realized under the influence of both the preceding tone and sentence type? (2) Does the neutral tone have different targets in questions and statements? (3) What is the effect of focus on the neutral tone in statements and questions? (4) Is the neutral tone more effective than a full tone in manifesting the statement/question distinction?

4.1.2 The interaction of sentence type with lexical stress and focus in English

In English, there are certain classes of word (e.g., *SUBject* (noun) vs. *subJECT* (verb)) whose grammatical function is determined by its stress pattern. Through perception experiments, Fry (1958) showed that in citation form, such stress contrast is signaled by the differences in F_0, duration and intensity between the two syllables in disyllabic words. However, it has been argued that sentence intonation (e.g., question intonation) and intonational contexts (e.g., after pitch accent) may over-ride and cancel out the F_0 distinction between the two words (Ladd, 1996). Eady and Cooper (1986) reported that words in sentence-final positions have a falling F_0 contour in statements but a rising contour in questions, while F_0 of the initial word in both statements and questions is rising, with the exception of the initial focused word in a statement, which exhibits a rising-falling (or generally falling) contour. Xu and Xu (2005) found similar F_0 patterns for English statements, but treated them as evidence for the existence of syllable-sized

underlying pitch targets in the language. In particular, a continuous F_0 rise in a non-final stressed syllable (focused or non-focused) or a non-focused word-final stressed syllable was viewed as due to a [high] pitch target, whereas the final F_0 fall in a focused word-final stressed syllable or a sentence-final stressed syllable (focused or non-focused) was viewed as due to a [fall] target. But little is yet known about the underlying pitch targets in English questions, as not enough detail was reported by Eady and Cooper (1986), and only statements were investigated in Xu and Xu (2005).

When investigating focus in different sentence types, Eady and Cooper (1986) noted that statements and questions with neutral or final focus differ only in the F_0 topline of the last key word, with peak F_0 of questions being higher than that of statements. Nevertheless, there is no significant difference in peak F_0 on the focused word for statements and questions with initial focus. Rather, the two sentence types differ in the F_0 topline following the initial focused word, with that of statements dropping to a low value and that of questions remaining rather high. Pell (2001) confirmed the above observation, and further noticed that peak F_0 of the on-focus (initial or final) vowels is increased in statements but not in questions. Xu and Xu (2005:177) provided a more detailed account of the focus effect in statements: a narrow focus would "increase the size of the F_0 peak in the stressed syllable under focus, lower all the postfocus F_0, including that of the poststressed syllables in the focused word, and leave prefocus F_0 largely intact". This suggests that a more detailed account of the focus effect in questions is also needed, because there could be effects that pertain to syllable – a smaller temporal domain than word.

Regarding statement and question intonation in English, different studies have defined different temporal domains for their acoustic contrast. According to Gårding and Abramson (1965), the difference is mainly in the pitch movement of the last syllable, with statements showing a mid-falling movement and questions a rising movement. However, O'Shaughnessy (1979:145) reported that the cue for statements vs. yes/no questions is a terminal fall vs. rise, which starts from "the last F_0 stress feature". Furthermore, as opposed to statements, there are "no stress falls and other falling F_0" in yes/no questions. In the AM theory of English intonation (Hirschberg, 2004; Ladd, 1996; Pierrehumbert & Hirschberg, 1990; Pierrehumbert, 1980), a declarative intonation is transcribed as having a high pitch accent, a low phrase accent and a low boundary tone (H* L-L% in the ToBI convention), and a yes/no question a low pitch accent, a high phrase accent and a high boundary tone (L* H-H%). Therefore, at least three different temporal scopes have been proposed for the distinction between statement and question intonation in English: the last syllable, the terminal F_0 as well as the F_0 feature of the entire sentence, or on and beyond the nuclear accent. In addition, it is believed that such distinction is realized through either a falling/rising or a high/low F_0 contrast.

In summary, despite much research, the exact manner with which statements and questions are differentiated in English intonation remains unclear. In particular, three critical issues are unresolved: 1) the temporal scope of the acoustic contrast between the two sentence types, 2) the role of focus in their differentiation, and 3) whether there are any differences in terms of the underlying pitch targets of individual syllables.

4.1.3 Question Intonation in Mandarin and English: Why compare them and how?

The discussion so far shows that despite much research, there remain unresolved issues about the phonetic manifestation of sentence type in both languages. Although these issues can be addressed by further investigating the two languages separately, it may benefit us even more if we study them in parallel. This may allow us to explore questions that are common to both languages. To answer those questions, a seemingly obvious strategy is to simply make a number of acoustic measurements and report whatever patterns emerge from the data. The danger of doing so is that we will simply add to the already highly diverse picture one more set of descriptions. An alternative is to explore the possibility of developing a theoretical account that can capture the underlying mechanisms of encoding these functions in both languages. Because of the apparent typological dissimilarity between the two languages, a unified account, if shown possible, may have the potential of being applicable to other languages that are typologically just as diverse.

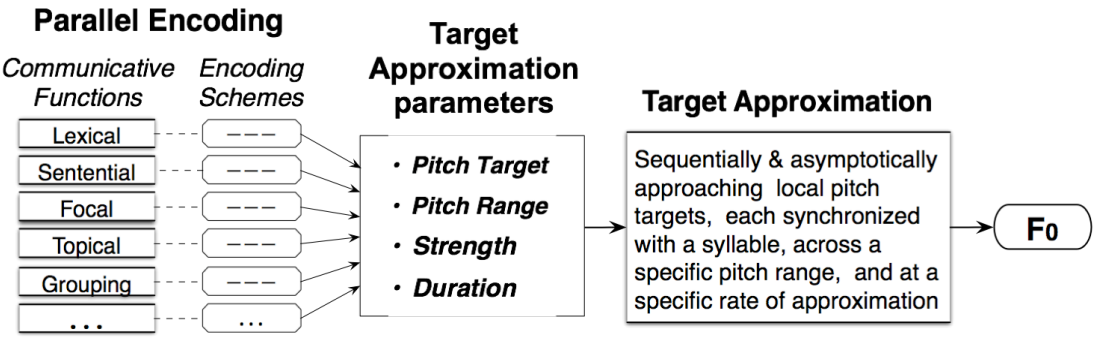

Figure 4.1. A schematic diagram of PENTA. Modified from Xu (2005).

The theoretical framework I would like to explore is the Parallel Encoding and Target Approximation model, or PENTA for short (Xu, 2005). As shown in Figure 4.1,

PENTA assumes that communicative functions (represented by the stacked boxes on the far left) are the driving force of prosody, and that they are phonetically encoded by controlling the Target Approximation (TA) parameters (open box) through specific encoding schemes. The TA parameters in turn control the TA model which simulates an articulatory process sketched in Figure 4.2a (Xu & Wang, 2001). In TA each syllable is associated with an underlying pitch target which is either static (e.g., syllable 2 in Figure 4.2a) or dynamic (e.g., syllable 1 in Figure 4.2a), and the production of F_0 in each syllable is a process of approaching its pitch target starting from an initial F_0 state either inherited from the previous syllable or from a neutral state at the utterance onset. The approximation of the pitch target ends at the syllable offset, but the final F_0 and F_0 velocity become the initial state of the next syllable. Thus the surface F_0 is a smooth and continuous contour resulting from sequentially approaching successive pitch targets each associated with a syllable. More importantly, the control parameters of the TA model, namely, pitch target, pitch range, strength and duration can be specified by the encoding schemes of various communicative functions as shown in Figure 4.1, and modifying each parameter results in changes in the output F_0. Figure 4.2b illustrates what may happen to surface F_0 when a four-syllable sequence is assigned different pitch targets and pitch ranges. As can be seen the F_0 contour of each syllable shows strong influence from the preceding syllable, yet it approaches the underlying target by the syllable offset. The effect of pitch range is to change the height and slope of the pitch targets, which in turn affect the global F_0 trend of the sequence. In Figure 4.2c, the strength of the first and last syllable is **strong**, which allows the pitch targets to be more closely approached, while

the strength of the middle two syllables is **weak**, which results in apparent undershoot of the targets even by the end of the syllable.

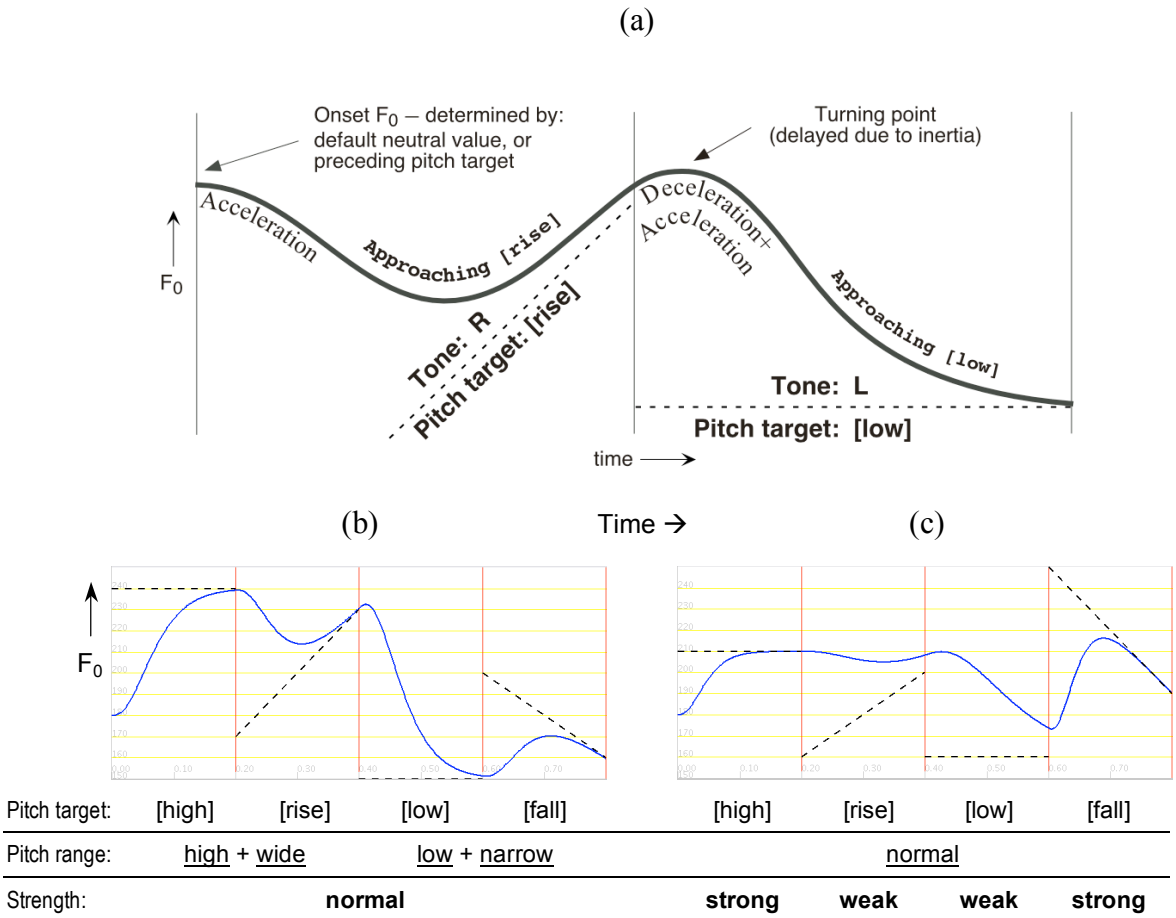

Figure 4.2. (a) A sketch of the Target Approximation (TA) model (Xu & Wang, 2001). The thick curve represents surface F_0; the dashed lines represent the underlying pitch targets, and the vertical lines represent syllable boundaries. (b) & (c) Illustrations of the effects of modifying the TA model. The labeling convention is based on Xu (2005).

In PENTA, the target approximation process as well as the overall framework are assumed to be universal (Xu, 2005). But the encoding schemes of specific communicative functions, and even whether they actually get encoded, are assumed to be language or even dialect dependent. Thus individual encoding schemes have to be discovered through empirical research. PENTA nevertheless provides a framework that

guides the discovery of the function-specified parameters. First, although every syllable is assumed to have a pitch target, the functional source of the target is language dependent and thus has to be established empirically. For the current chapter, I am particularly concerned with whether there are changes in the underlying pitch targets of individual syllables in both Mandarin and English. Second, the nature of the TA process entails that the underlying pitch target of a syllable is best approached by the end of the associated syllable. Thus the property of a pitch target is best estimated by examining the final F_0 contour of the syllable (as demonstrated in Chapter 3). In the current chapter, this will be done mainly by examining the velocity of F_0 at syllable-final positions. Third, the TA mechanism also entails that pitch range specifications cannot be applied directly to surface F_0 contours, but rather the control of pitch range has to be done by modifying the parameters of the local pitch targets. Thus for each syllable in a language, the pitch target and its pitch range need to be established separately, even if it is possible that a particular communicative function may involve both. In regard to the present study, past research has indicated that focus in both Mandarin and English manifests on-focus pitch range expansion and post-focus pitch range lowering and compression in statements. It is not clear, however, whether post-focus lowering and compression also both apply in questions in the two languages. Previously reported data seem to suggest that in English questions only compression remains (Eady & Cooper, 1986; Pierrehumbert, 1980). For Mandarin, Experiment 1 has shown that both compression and lowering still apply in questions when the post-focus syllables carry full lexical tones (Liu & Xu, 2005). But it is not known whether this is still true for post-focus neutral tone syllables.

Two experiments were designed to address the unresolved issues presented in 4.1.1 and 4.1.2. While they may appear to be independent, the results will become comparable in terms of target realization of lexical tones/stresses under different focus conditions and in different sentence types in Mandarin and English.

4.2 Experiment 4: The Neutral Tone in Statements and Questions in Mandarin

Experiment 4 investigates whether the neutral tone, when away from the sentential focus and at the sentence-final position, would reflect question intonation more effectively than the full tones in Mandarin. That is, would the neutral tone change its underlying target (if any) in order to realize question intonation in Mandarin?

4.2.1 Materials

Speech materials were composed such that they either end with a sequence of neutral tone syllables or with a disyllabic word with the High tone on both syllables. As shown below, in each groups of sentences, the tone of the second syllable is always Low, while the tone of the third syllable varies across four full tones.

1) Neutral-tone Ending: *Tā **mǎi mā**ma/**yé**ye/**nǎi**nai/**mèi**mei men de le ma*

(He bought mothers'/grandpas'/grandmas'/sisters')

2) High-tone Ending: *Tā **mǎi mā**ma/**yé**ye/**nǎi**nai/**mèi**mei men de māomī*

(He bought mothers'/grandpas'/grandmas'/sisters' kitten)

The following three conditions were imposed onto the above two sentence frames so that a total of 32 combinations of the factors were created.

Focus: nonadjacent vs. adjacent. The focus is either (on the second syllable *mǎi*) separated from the sequence of 5 neutral tones (in sentence frame 1, or 3 neutral

tones + 2 High tones in sentence frame 2) by a full tone, or (on the third syllable *mā/yé/nǎi/mèi*) immediately adjacent to them.

Preceding tone: High, Rising, Low, or Falling. The sequence of 5 neutral tones (or 3 neutral tones + 2 High tones in sentence frame 2) is preceded by each of the four full tones.

Sentence type: statement vs. question. Each utterance with the same components was produced with two alternate sentence types.

Each utterance was repeated five times by each subject, resulting in a total of 1280 sentences. The intended focus and sentence type were elicited by the following leading sentences, in which the bolded words are focused:

For High-tone ending sentences with focus on the second syllable *mǎi*:

Lǎowáng bú ***mài*** *māma/yéye/nǎinai/mèimei men de māomī./?*

(Laowang didn't ***sell*** *mothers'/grandpas'/grandmas'/sisters' kittens./?)*

For Neutral-tone ending sentences with focus on the second syllable *mǎi*:

Lǎowáng bú ***mài*** *māma/yéye/nǎinai/mèimei men de dōngxi./?*

(Laowang didn't ***sell*** *mothers'/grandpas'/grandmas'/sisters' goodies./?)*

For High-tone ending sentences with focus on the third syllable *mā/yé/nǎi/mèi*:

Lǎowáng bù mǎi ***jiějie*** *men de māomī./?*

(Laowang didn't buy ***older sisters'*** *kittens./?)*

For Neutral-tone ending sentences with focus on the third syllable *mā/yé/nǎi/mèi*:

Lǎowáng bù mǎi ***jiějie*** *men de dōngxi./?)*

(Laowang didn't buy ***older sisters'*** *goodies./?)*

4.2.2 Subjects

Eight native speakers of Mandarin, 4 females and 4 males, served as subjects. They were either students at Yale University or residents in New Haven, Connecticut, who were born and raised in the city of Beijing where Mandarin is the vernacular. They had no self-reported speech or hearing disorders and their ages ranged from 23 to 34.

4.2.3 Procedure

Recording was done in a sound-isolated booth at Haskins Laboratories, New Haven, Connecticut. A JavaScript program displayed the target sentences (and the leading sentences) one at a time on a computer screen in random order. The subject sat in front of the computer and read aloud both the leading and target sentences into a high quality omni-directional monaural microphone one meter away. The utterances were directly digitized at 44.1 kHz sampling rate and 16-bit amplitude resolution, and were later re-sampled at 22.05 kHz. For visual inspection and graphic analysis, a custom-written script for the Praat program (Boersma & Weenink, 2008) computed time-normalized F_0 contours of the sentences by getting the same number of evenly spaced F_0 points from each syllable (see Xu, 2005-2008 for a general-purpose version of the script). For statistical analyses, from each syllable mean F_0 (indicating pitch height, in st), max F_0 – min F_0 (indicating pitch span, in st), duration (in ms), and final velocity — velocity of F_0 (instantaneous rate of change of F_0, = $(F_{0i+1} – F_{0i-1}) / (t_{i+1} – t_{i-1})$, in st/s) at 30 ms before syllable offset (as evidence of the underlying pitch target, cf. Gauthier, et al. 2007; Xu & Liu, 2006, and will be further discussed later) were calculated by the script.

4.2.4 Results

4.2.4.1 Graphical comparison

To see how sentence type is affected by lexical tone and focus, mean time-normalized F_0 contours of the statements and questions with focus on *măi* and on *mā/yé/năi/mèi* are displayed in Figures 4.3 and 4.4, respectively, each averaged across 40 repetitions by eight speakers.

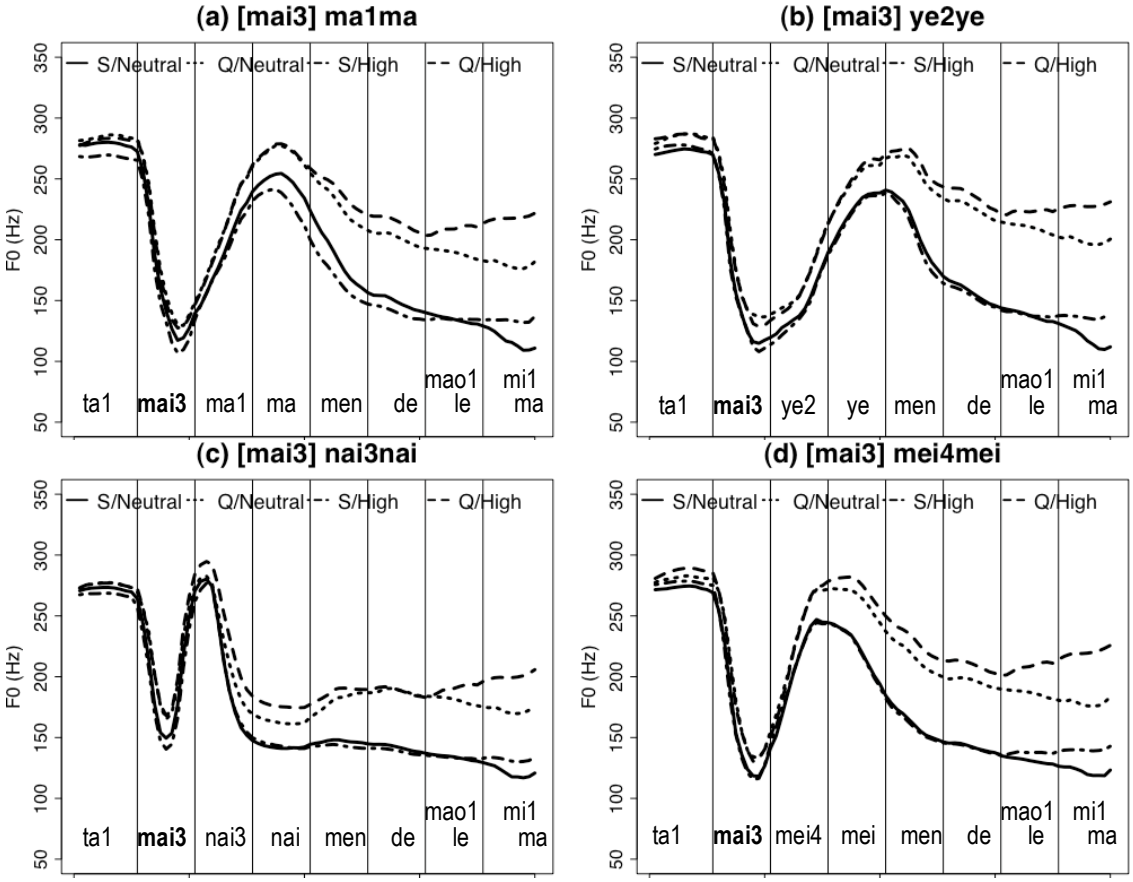

Figure 4.3. Time-normalized F_0 contours (averaged across 40 repetitions by 8 subjects) of Mandarin statements/questions with focus on "măi". In the legend, "S/Neutral" refers to a statement ending with 5 neutral tones, "Q/High" a question ending with 2 High tones, and so on. The title "[mai3] ma1ma" indicates that focus is on the second syllable "măi" and the 3rd and 4th syllables of the sentence are "māma".

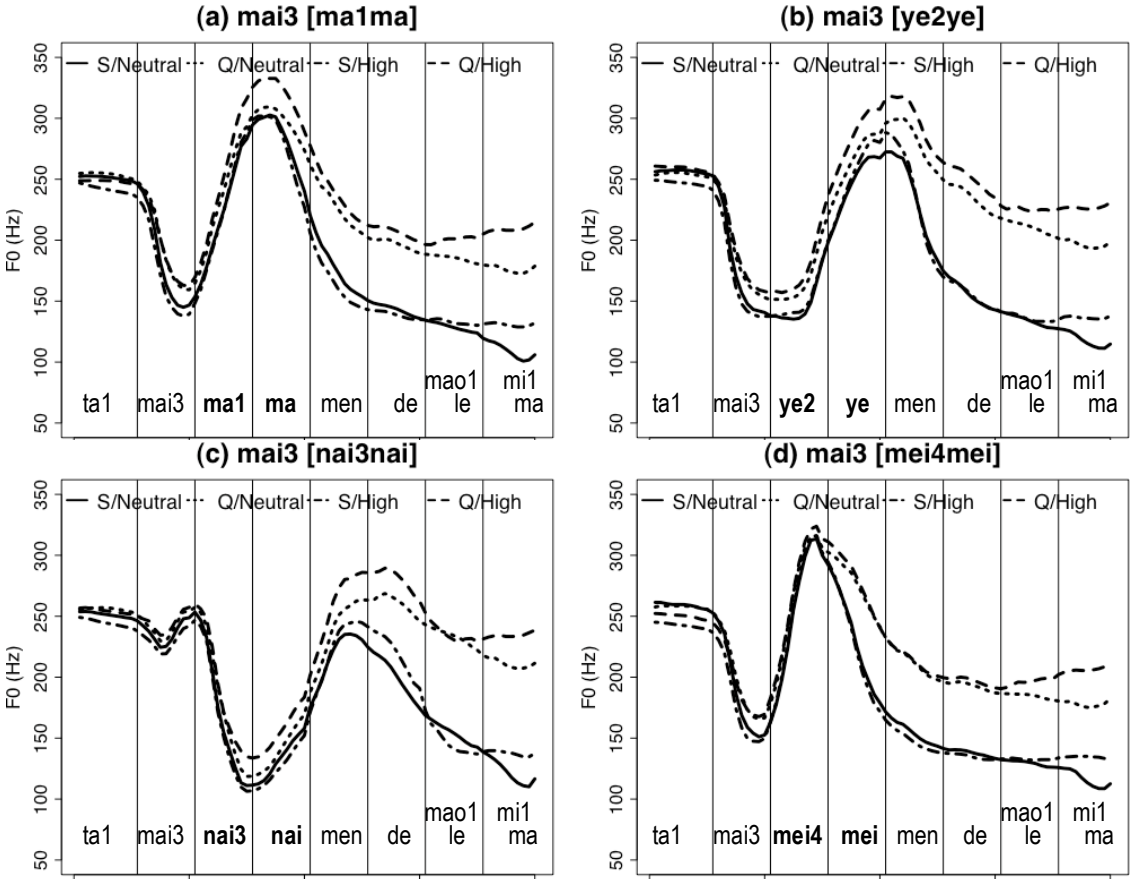

Figure 4.4. Time-normalized F_0 contours of Mandarin statements/questions with focus on the 3rd syllable "mā/yé/nǎi/mèi". The title "mai3 [ma1ma]" indicates that focus is on "māma", the 3rd and 4th syllables of the sentence.

As can be seen, regardless of the tone of the third and the last two syllables, in each graph, questions differ from statements with an increasingly higher F_0 starting from the focused syllable (*mǎi* in Fig. 4.3 and *mā/yé/nǎi/mèi* in Fig. 4.4) till the end of the sentence. Furthermore, this difference is larger in the final portion of the sentence when the last two tones are High than when they are neutral. The effect of focus can be seen by comparing Fig. 4.3 with Fig. 4.4. The most obvious is the shift in the size of the pitch range between the two figures: it is wider in syllable 2 in Fig. 4.3 but in syllable 3 in Fig. 4.4. In addition, we can see the effect of pitch range expansion on the immediately pre-

focus and post-focus syllables. In Fig. 4.3 focus on syllable 2 not only lowers its F_0 but also raises the F_0 of its preceding syllable *tā* (by 1.66 st). This is similar to the anticipatory dissimilation effect reported by previous studies (Gandour, Potisuk & Dechongkit, 1994; Laniran & Clements, 2003; Xu, 1997).[1] In Fig. 4.4 focus on the third syllable expands not only its pitch range, but also that of the following neutral-tone syllable, which is likely due to the same carryover effect as reported before (Xu, 1999). Furthermore, the F_0 trajectories of the rest of the syllables differ dramatically in the two figures because of the different focus locations. In Fig. 4.3c, F_0 of the neutral-tone syllable *nai* remains low when preceded by the post-focus Low-tone syllable *nǎi*, but that of the next neutral-tone syllable *men* increases slightly, which is in turn followed by a slow decrease in the rest of the neutral-tone syllables *de le ma*. In Fig. 4.4c, however, following the focused Low-tone syllable *nǎi*, F_0 of the next two neutral-tone syllables *nai men* rises sharply, and the rise does not stop until either syllable 5 (*men*) or syllable 6 (*de*). This phenomenon is described as "post-L raising" in Chen and Xu (2006). Fig. 4.3c suggests that this effect is quite weak when the Low tone itself is post-focus, but very strong when the Low tone is focused and the neutral tone is post-focus as in Fig. 4.4c.

To see the effect of four full tones on the following neutral tone sequences (and the 3-neutral plus 2-High sequences) under different focus and sentence type conditions, Figures 4.5 and 4.6 display mean time-normalized F_0 contours of the statements and

[1] In Fig. 4.3c and 4.4c, the anticipatory effect of the Low tone in syllable 3 on the preceding Low tone, which changes it into a Rising tone, however, is due to the well known tone sandhi rule (Low + Low > Rising + Low) in Mandarin (Chao, 1968).

questions alternating across *māma/yéye/năinai/mèimei* on the third and fourth syllable with the neutral-tone and High-tone ending, respectively.

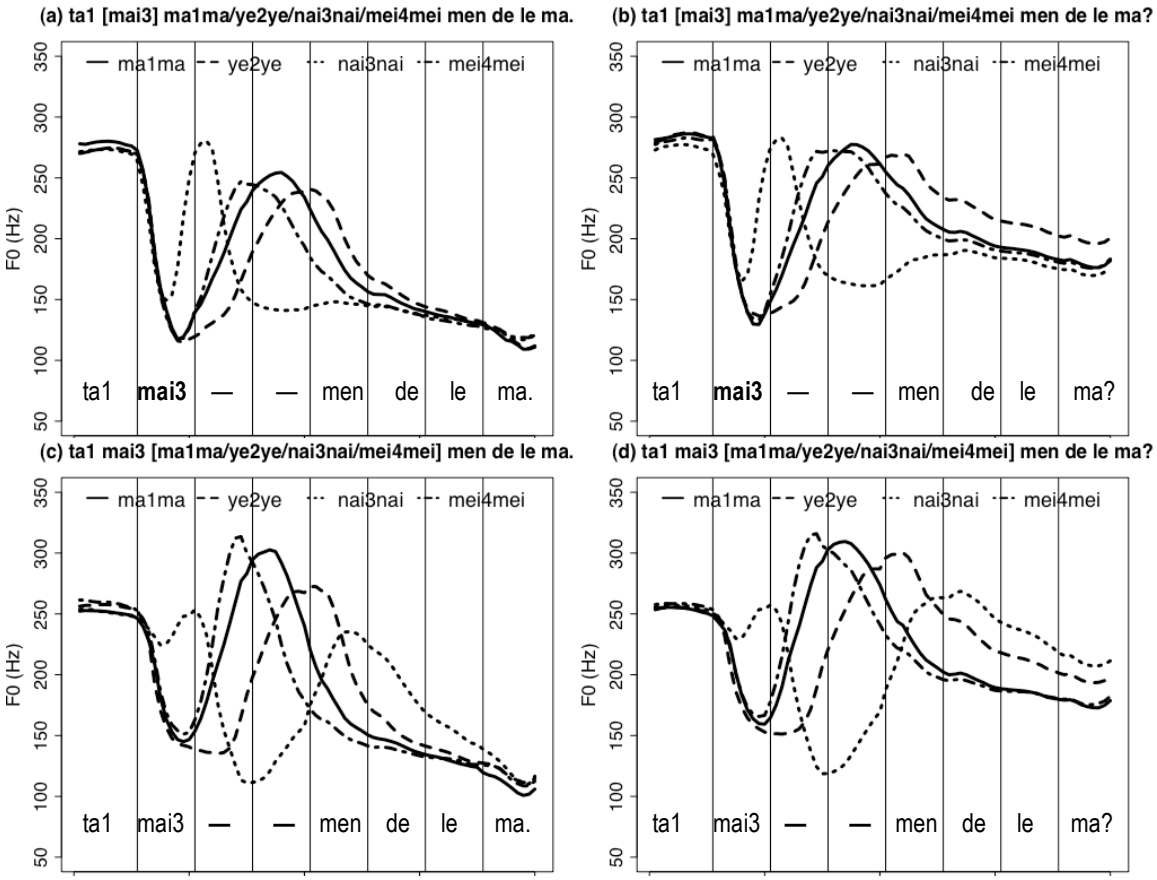

Figure 4.5. Time-normalized F_0 contours of Mandarin statements/questions alternating across "māma/yéye/năinai/mèimei" on the third and fourth syllables with focus on "măi" (in (a) and (b)) or on "mā/yé/năi/mèi" (in (c) and (d)) and with the neutral tone on the final 5 syllables.

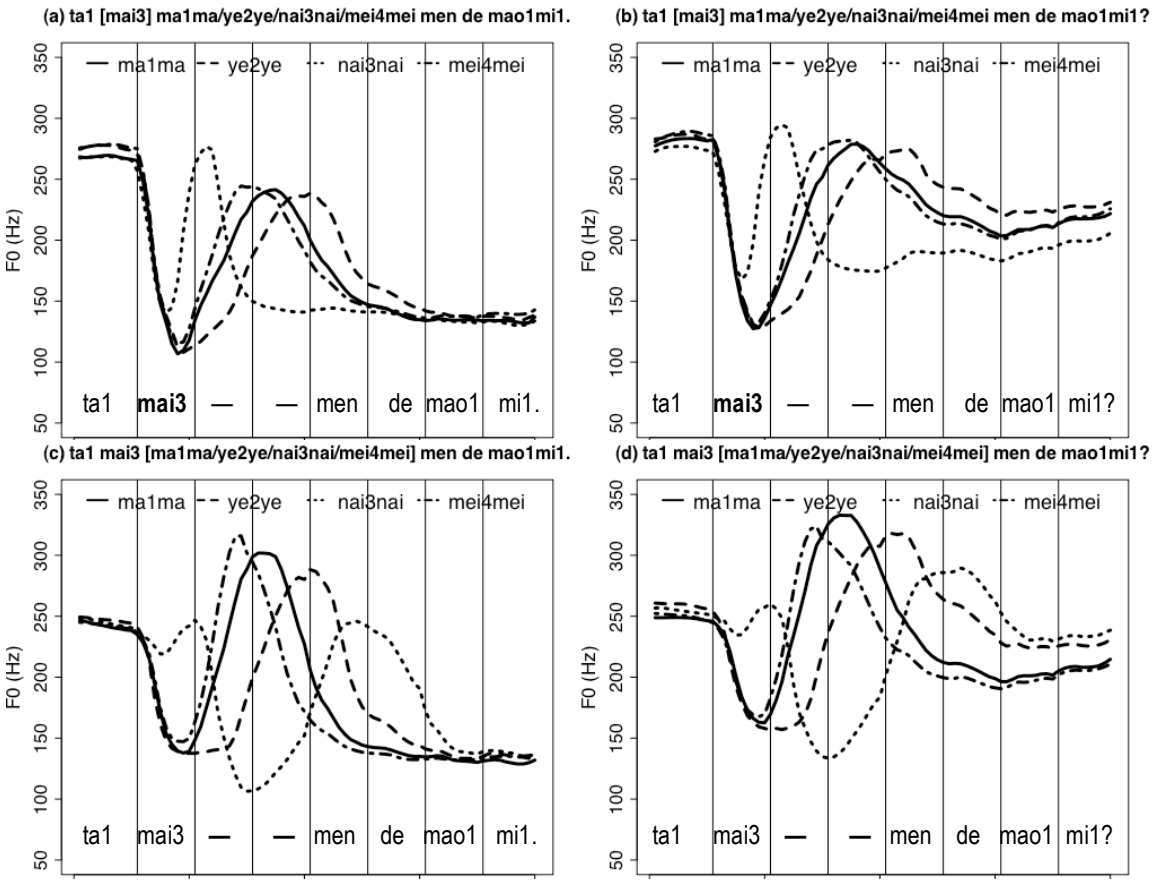

Figure 4.6. Time-normalized F_0 contours of Mandarin statements/questions alternating across "māma/yéye/nǎinai/mèimei" on the third and fourth syllable with focus on "mǎi" (in (a) and (b)) or on "mā/yé/nǎi/mèi" (in (c) and (d)) and with the High tone on the final two syllables.

Fig. 4.5 shows that the magnitude of the influence of the full tone on the neutral tones depends not only on the tonal target of the full tone, but also on the focus and sentence type conditions. In Fig. 4.5a, where the sentences are statements with focus on *mǎi*, F_0 of all the tones except the Low tone goes up after the focused Low tone in syllable 2. These rises also affect the F_0 of the following neutral tone sequence in the descending order of Rising > High > Falling. Despite such influence, however, by the time of the last two neutral tones F_0 has largely converged to a falling contour. In Fig. 4.5c, where the sentences are also statements but the focus is on syllable 3, the Low tone

has the most long-lasting raising effect on the trajectory of the neutral tone sequence, due to the post-L raising mentioned earlier, while the order of the effects from the other three tones (Rising > High > Falling) remains the same. Here the F_0 trajectories do not converge until the last syllable in the neutral tone sequences. In Fig. 4.5b and 4.5d, where the corresponding questions under different focus conditions are plotted, F_0 contours of the neutral tone sequences never converge, thanks to the Rising tone in Fig. 4.5b and the Low tone in Fig. 4.5d. In spite of the variations in the amount of convergence, however, an overall decreasing trend can be seen in the F_0 of all the neutral tone sequences in question intonation.

What is observed above for the neutral tone sequence also applies to the 3 neutral tones plus 2 High tones sequence in Fig. 4.6, except that the final High-tone sequences exhibit slight rising contours in questions, as opposed to the level contours in statements.

4.2.4.2 Statistical analysis

Mean F_0 (indicating pitch height, in st) and pitch span (= max F_0 – min F_0, in st) of the syllables in these sentences (grouped according to focus, sentence type, final tone, and preceding tone conditions) are listed in Table 4.1. Two sets of repeated measures ANOVAs were conducted using R (R Development Core Team, 2005) to see if the observations made on Figures 4.3-4.6 are statistically significant. In each of the ANOVA models, mean F_0 or pitch span of each of the eight syllables in the sentences is treated as the dependent variable, sex as the between-subjects factor, and focus, sentence type, final tone, and preceding tone as the within-subjects factors. For post-hoc analysis, Tukey's

Honest Significant Difference method (the Studentized range statistic) was used to do multiple comparisons between the means of multiple levels of the factors.

Table 4.1. Mean F_0 (in st) and pitch span (in st) of syllables in Mandarin statements and questions. Potential focus locations are bolded.

Sentence type	Question				Statement			
Focus	măi		māma/yéye/ năinai/mèimei		măi		māma/yéye/ năinai/mèimei	
Pitch range	Mean F_0	Pitch span	Mean F_0	Pitch span	Mean F_0	Pitch span	Mean F_0	Pitch span
ta/ta	97.8/97.6	1.0/1.0	95.9/95.5	1.0/1.0	97.5/96.8	1.0/1.0	95.7/95.1	1.1/1.0
măi/măi	**93.1/92.9**	**14.7/14.8**	92.6/92.5	8.0/7.5	**92.2/91.6**	**16.4/18.4**	91.8/91.0	9.8/9.7
mā/mā	92.0/91.9	10.3/10.6	**94.5/95.5**	**10.7/11.7**	90.5/90.0	10.6/11.3	**93.5/93.6**	**11.6/12.6**
ma/ma	96.9/97.0	2.0/1.9	**98.8/100.0**	**2.7/2.7**	95.4/94.4	2.8/3.2	**98.1/97.9**	**5.0/5.5**
men/men	94.6/95.2	3.8/3.0	94.5/95.4	5.0/5.3	91.3/89.5	6.6/6.0	90.6/89.4	7.6/7.5
de/de	92.0/93.1	1.5/1.4	91.6/92.4	1.5/1.3	86.8/85.8	2.2/1.8	86.1/85.5	2.3/1.2
le/māo	90.8/92.4	1.0/1.2	90.5/91.7	0.9/1.0	85.0/84.9	1.4/1.0	84.2/84.7	2.3/1.3
ma/mī	89.9/93.2	1.0/1.1	89.6/92.4	1.1/1.2	83.0/84.8	5.8/2.1	81.7/84.4	5.7/1.6
ta/ta	97.8/97.8	1.0/0.9	95.9/96.2	1.1/1.2	97.1/97.3	1.0/0.9	96.0/95.4	1.0/1.0
măi/măi	**93.1/93.0**	**13.8/14.3**	92.3/92.6	8.9/8.7	**91.9/91.8**	**16.5/18.0**	91.7/90.9	10.7/10.4
yé/yé	87.9/87.7	7.2/7.7	**88.3/89.1**	**6.0/6.5**	85.8/85.4	8.0/8.8	**86.2/86.5**	**6.2/6.0**
ye/ye	95.1/95.3	4.4/4.9	**96.4/97.5**	**5.6/5.7**	93.4/93.2	5.0/5.0	**95.1/95.5**	**6.5/7.1**
men/men	96.3/96.7	2.7/2.6	98.0/99.1	3.4/3.6	93.3/93.0	6.0/6.4	95.4/96.0	7.6/9.0
de/de	94.0/94.7	1.7/1.5	94.7/95.8	2.5/2.3	87.8/87.4	3.2/2.5	87.8/87.7	4.2/3.3
le/māo	92.6/93.6	1.0/1.1	92.8/93.8	1.4/1.2	85.4/85.4	1.8/1.3	85.0/85.2	2.0/1.9
ma/mī	91.7/94.0	1.0/0.9	91.5/93.9	1.2/0.8	83.4/85.1	5.9/1.9	83.0/85.1	4.9/2.0
ta/ta	97.3/97.3	1.0/1.0	96.0/95.9	1.0/1.0	97.1/96.8	1.1/1.0	95.7/95.2	0.9/1.1
măi/măi	**93.9/94.2**	**10.1/10.3**	95.2/95.4	3.1/2.9	**92.9/92.5**	**11.8/12.5**	94.9/94.3	3.7/3.3
năi/năi	95.5/96.6	8.9/8.1	**92.5/92.9**	**14.0/11.9**	95.0/95.0	11.1/11.5	**92.0/91.7**	**15.6/15.7**
nai/nai	88.3/89.7	1.8/1.8	**85.7/87.4**	**7.0/6.5**	86.0/86.2	2.1/2.1	**84.9/84.0**	**7.4/7.7**
men/men	89.8/90.5	2.5/1.8	94.5/96.0	7.5/7.5	86.3/85.9	1.1/1.0	93.0/93.5	7.0/8.2
de/de	90.6/90.7	0.9/0.8	96.4/97.7	1.8/1.9	85.9/85.5	1.1/0.8	92.3/93.7	5.0/4.5
le/māo	90.1/90.8	0.9/1.2	94.5/94.8	1.8/2.5	84.8/84.8	1.1/1.4	87.6/87.0	3.4/5.6
ma/mī	89.2/91.7	1.3/1.2	92.7/94.4	1.4/0.9	83.4/84.6	4.1/2.8	83.9/85.3	6.3/1.9
ta/ta	97.6/97.9	1.1/1.2	96.1/95.6	1.0/1.0	97.1/97.4	0.9/0.8	96.2/95.1	0.8/1.1
măi/măi	**93.2/93.2**	**13.9/14.8**	92.9/92.4	7.9/7.0	**92.2/92.4**	**16.9/17.0**	92.4/91.2	9.4/8.9
mèi/mèi	93.9/93.9	11.1/11.7	**96.8/97.1**	**11.1/11.2**	92.4/92.4	11.4/11.1	**96.3/96.5**	**12.5/13.0**
mei/mei	96.7/97.3	2.3/2.2	**97.7/98.0**	**4.3/4.8**	94.1/94.0	4.5/4.6	**95.5/95.4**	**9.0/10.0**
men/men	93.5/94.5	3.4/3.0	93.1/93.2	3.5/3.1	88.6/88.4	4.6/4.5	87.7/86.9	3.8/3.6
de/de	91.5/92.6	1.1/1.1	91.2/91.5	1.1/1.0	85.9/85.8	1.5/1.3	85.4/85.0	1.4/1.1
le/māo	90.6/92.3	0.9/1.3	90.4/91.3	0.7/1.2	84.5/85.3	1.4/1.2	84.3/84.6	1.3/1.0
ma/mī	89.8/93.3	1.2/1.3	89.7/92.2	1.2/1.2	83.4/85.5	3.9/1.8	82.8/84.9	5.2/1.6

Table 4.2. Results of the main effects from repeated measures ANOVAs of mean F_0 (in st) of each syllable on sex (Female vs. Male), focus (Nonadjacent: on *măi* vs. Adjacent: on *mā/yé/năi/mèi*), sentence type (Question vs. Statement), final tone (High vs. Neutral), and preceding tone (High, Rising, Low, or Falling). The effects with *p*-values less than 0.05 are bolded.

		Sex	Focus	Sentence type	Final tone	Preceding tone
tā	F	**15.30**	**7.73**	2.58	3.75	1.85
	p	**0.0079**	**0.0320**	0.1596	0.1008	0.1738
		F>M	**A<N**			
măi	F	**17.54**	0.00	**8.05**	**14.16**	**27.16**
	p	**0.0058**	0.9996	**0.0297**	**0.0094**	**< 0.0001**
		F>M		**Q>S**	**H<N**	**L>(F,H,R)**
mā/ yé/ năi/ mèi	F	**13.53**	**15.66**	**25.61**	5.32	**138.94**
	p	**0.0104**	**0.0075**	**0.0023**	0.0605	**< 0.0001**
		F>M	**A>N**	**Q>S**		**F>L>H>R**
ma/ ye/ nai/ mei	F	**14.67**	**13.14**	**35.74**	2.89	**108.67**
	p	**0.0087**	**0.0110**	**0.0010**	0.1400	**< 0.0001**
		F>M	**A>N**	**Q>S**		**H>F>R>L**
men	F	**20.64**	**69.31**	**51.81**	1.87	**58.08**
	p	**0.0039**	**0.0002**	**0.0004**	0.2200	**< 0.0001**
		F>M	**A>N**	**Q>S**		**R>H>(L,F)**
de	F	**27.91**	**56.07**	**79.82**	1.81	**51.60**
	p	**0.0019**	**0.0003**	**0.0001**	0.2274	**< 0.0001**
		F>M	**A>N**	**Q>S**		**(L,R)>(H,F)**
le/ māo	F	**33.16**	**11.73**	**116.73**	**7.19**	**35.51**
	p	**0.0012**	**0.0141**	**< 0.0001**	**0.0365**	**< 0.0001**
		F>M	**A>N**	**Q>S**	**H>N**	**(L,R)>(H,F)**
ma/ mī	F	**38.42**	0.41	**153.67**	**37.74**	**14.18**
	p	**0.0008**	0.5456	**< 0.0001**	**0.0009**	**< 0.0001**
		F>M		**Q>S**	**H>N**	**R>(F,H) & L>H**

The main effects of sex, focus, sentence type, final tone, and preceding tone on mean F_0 of each syllable in the ANOVA models are shown in Table 4.2 and summarized as follows. 1) Female speakers generally have higher mean F_0 than male speakers. 2) The effect of focus varies with its location: (i) When it is on *măi*, which is nonadjacent to the following neutral tone sequence, focus has an anticipatory raising effect on the mean F_0

of the preceding syllable *tā*. (ii) When focus is on *mā/yé/nǎi/mèi*, which is adjacent to the following neutral tone sequence, it raises the mean F_0 of both the focused and the post-focus (except the last) syllables. 3) Starting from the focused item, the F_0 of questions becomes higher than that of statements. The difference in F_0 height between the two is increasingly larger as the sentences approach the end. This is demonstrated by the more and more significant effect of sentence type on the mean F_0 of the eight consecutive syllables in the sentences. 4) There is no obvious anticipatory effect of the tone of the final two syllables (High vs. neutral) on the mean F_0 of the preceding syllables. Rather, the effect is strictly local: sentences ending with High tones have significantly higher final mean F_0 than those ending with neutral tones. 5) There is no anticipatory effect of the four full tones of the third syllable on the mean F_0 of the first syllable. According to the tone sandhi rule (L + L > R + L) in Mandarin, *mǎi* (L) becomes *mái* (R) before *nǎi* (L), which is why the mean F_0 of *mǎi* is the highest before the Low tone. Following *mǎi*, the mean F_0 of the four full tones shows the order of *mèi* > *nǎi* > *mā* > *yé*. This is likely due to the articulatory constraint that determines how fast F_0 can change even with full effort (Xu & Sun, 2002). Following the four full tones, there are four corresponding neutral tone syllables whose F_0 trajectory seems to be a continuation of its preceding full tone, thus having mean F_0 in the order of *ma* > *mei* > *ye* > *nai*. Following the ending point of *māma/yéye/nǎinai/mèimei*, the F_0 contour of the neutral tone syllable *men* is mostly falling after *māma/yéye/mèimei*, but mostly rising after *nǎinai* (especially when *nǎinai* is focused). This makes its mean F_0 the highest after *yéye*, but much lower after *māma* and *nǎinai*/*mèimei*. The F_0 contour of *de* is falling no matter what the preceding

full tone is, but the "post-L raising" effect of the focused *nǎi* on *nai men* raises the pitch range of the following *de* so that this *de* has a higher mean F_0 than those preceded by the other three full tones. Although most cases of the last two syllables *le ma/māomī* converge to a falling/level contour in statements, only those preceded by *māma men de/mèimei men de* converge in questions. In addition, *le ma/māomī* are in a relatively higher pitch range when following focused/non-focused *yéye men de* and focused *nǎinai men de*.

Table 4.3. Results of the main effects from repeated measures ANOVAs of pitch span (Max F_0 – Min F_0, in st) of each syllable on sex (Female vs. Male), focus (Nonadjacent: on *mǎi* vs. Adjacent: on *mā/yé/nǎi/mèi*), sentence type (Question vs. Statement), final tone (High vs. Neutral), and preceding tone (High, Rising, Low, or Falling). The effects with *p*-values less than 0.05 are bolded.

		Sex	Focus	Sentence type	Final tone	Preceding tone
tā	F	0.68	0.14	3.06	0.20	0.15
	p	0.4404	0.7195	0.1307	0.6715	0.9285
mǎi	F	0.10	**28.43**	**25.58**	0.37	**66.20**
	p	0.7613	**0.0018**	**0.0023**	0.5669	**<0.0001**
			A<N	**Q<S**		**(R,H,F)>L**
mā/ yé/ nǎi/ mèi	F	0.27	**8.13**	**10.31**	0.37	**49.52**
	p	0.6198	**0.0291**	**0.0184**	0.5669	**<0.0001**
			A>N	**Q<S**		**(L,F,H)>R & L>H**
ma/ ye/ nai/ mei	F	0.16	**39.01**	110.09	1.02	**11.49**
	p	0.7047	**0.0008**	**<0.0001**	0.3513	**0.0002**
			A>N	**Q<S**		**(R,F,L)>H & R>L**
men	F	0.03	**43.75**	**159.72**	0.00	**6.36**
	p	0.8648	**0.0006**	**<0.0001**	0.9676	**0.0040**
			A>N	**Q<S**		**(H,R,L)>F & H>L**
de	F	2.73	**82.61**	**32.56**	**13.03**	**46.52**
	p	0.1497	**<0.0001**	**0.0013**	**0.0112**	**<0.0001**
			A>N	**Q<S**	**H<N**	**R>L>H>F**
le/ māo	F	0.10	**13.36**	**9.92**	0.95	**10.43**
	p	0.7636	**0.0106**	**0.0198**	0.3673	**0.0003**
			A>N	**Q<S**		**L>(R,H,F)**
ma/ mī	F	0.18	0.46	**22.93**	**51.12**	1.99
	p	0.6866	0.5243	**0.0030**	**0.0004**	0.1511
				Q<S	**H<N**	

The main effects of sex, focus, sentence type, final tone, and preceding tone on pitch span (max F_0 – min F_0) of each syllable in the ANOVA models are shown in Table 4.3 and summarized as follows. 1) Pitch spans of male and female speakers are not significantly different. 2) The effect of focus is manifested by a) on-focus pitch range expansion: the pitch spans of *mǎi* and *mā/yé/nǎi/mèi* are enlarged when they are focused,

b) post-focus pitch range suppression: the pitch span of the neutral tone sequence (except the last syllable) is narrower when the focus is on the second syllable than when it is on the third syllable, indicating that post-focus pitch range suppression becomes increasingly larger as the distance between the focus and the following syllables becomes longer, and c) pre-focus pitch range neutralization: the pitch span of the pre-focus syllable *tā* is not significantly affected by the focus location. 3) The pitch span of the pre-focus syllable *tā* is not significantly different between statements and questions; however, those of on-/post-focus syllables are narrower in questions than in statements. 4) Pitch spans of *de* and *ma/mī* are narrower when the final two tones are High than when they are neutral, but those of other syllables are not significantly affected by the tones of the last two syllables. 5) Pitch spans of the first and last syllable are not significantly affected by the tone of the third syllable, but those of other syllables vary with the alternating tone on the third syllable. Due to the tone sandhi rule (L + L > R + L), *măi* (L) changes into *mái* (R) when followed by *năi* (L); it thus has the smallest pitch span when followed by the Low tone than by other tones. For the third syllable *mā/yé/năi/mèi*, pitch span of *yé* is the smallest, and that of *năi* is the largest, which is determined by the pitch target of its preceding syllable *măi* and their own pitch targets. For the neutral tone syllable *ma/ye/nai/mei*, *ye* has the largest pitch span and *ma* has the smallest, which is caused by the carry-over effect from the previous full tone *mā/yé/năi/mèi*. Similarly, pitch spans of *men*, *de*, and *le/māo* are greatly influenced by the carry-over effect from their preceding neutral tones.

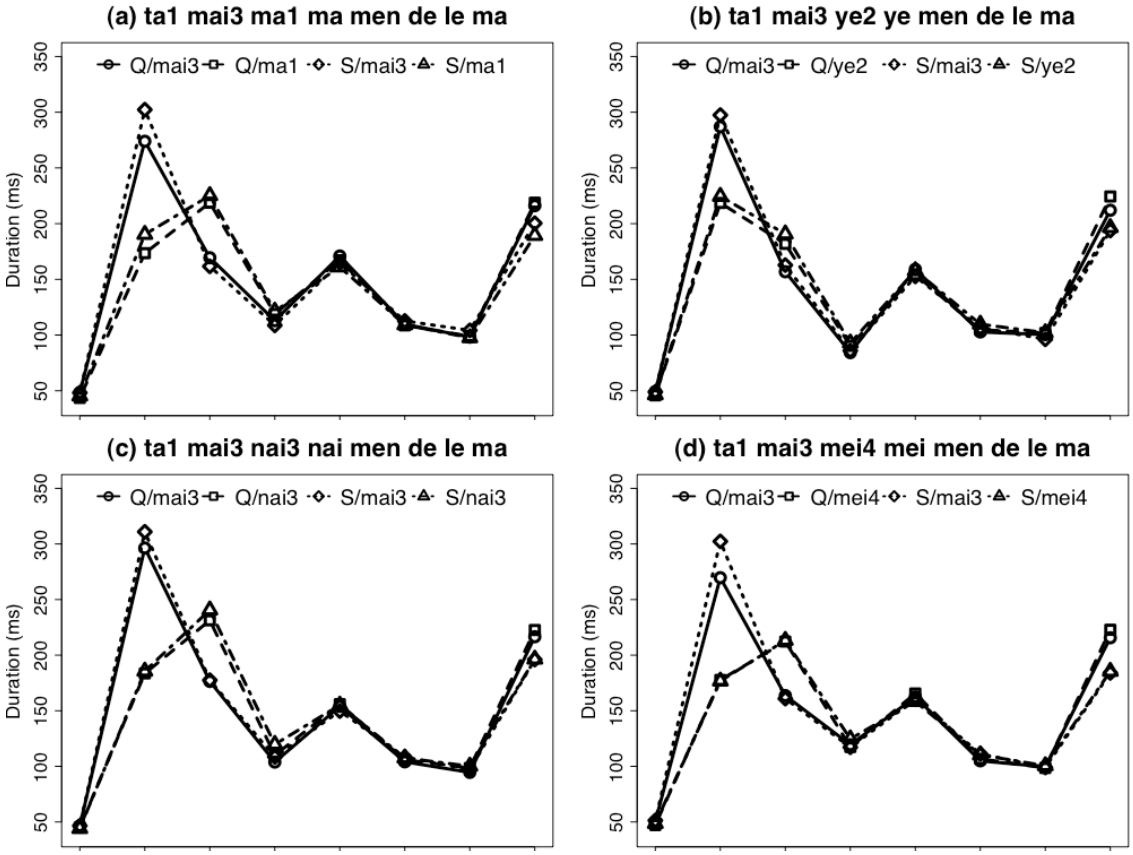

Figure 4.7. Durations of the syllables under different focus conditions and sentence types in the neutral-tone-ending sentences, which differ in the third and fourth syllable: (a) *māma*, (b) *yéye*, (c) *nǎinai*, and (d) *mèimei*. In the legend, "S" stands for statement, and "Q" for question. "Q/mai3" means a question with focus on "*mǎi*", and so on.

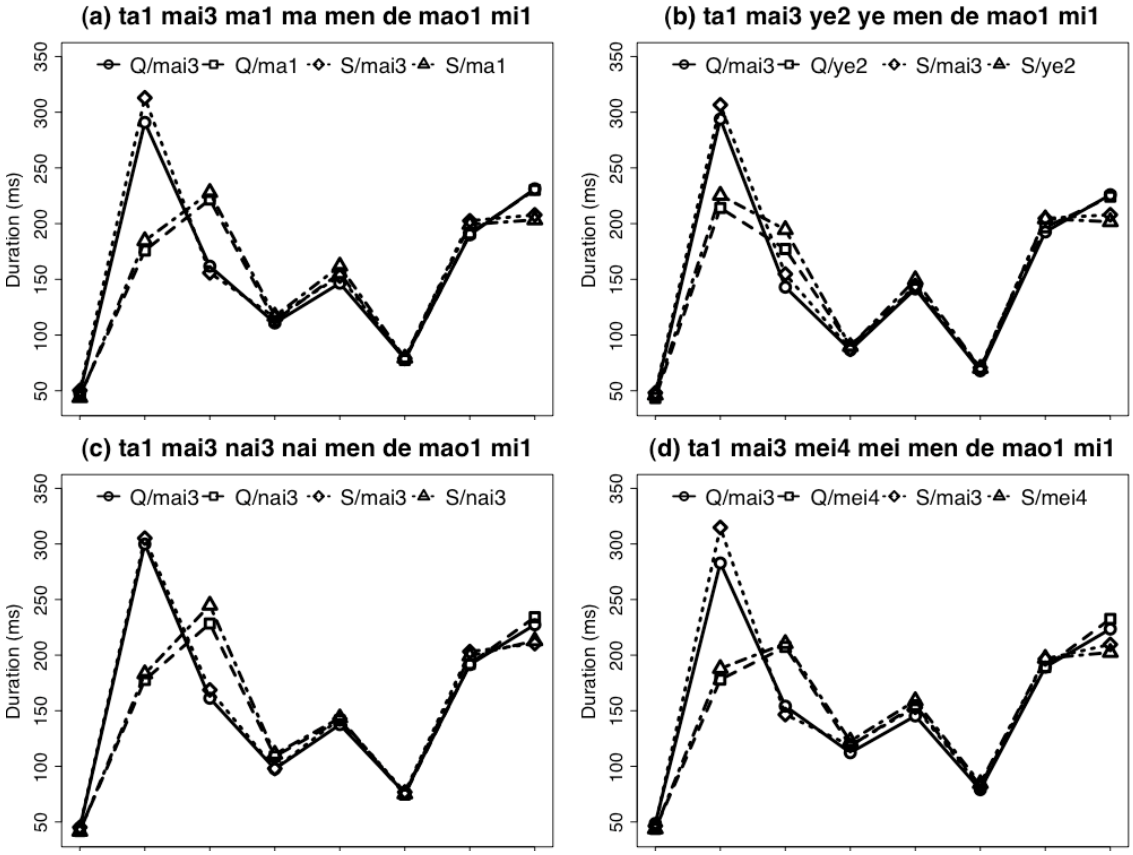

Figure 4.8. Durations of the syllables under different focus conditions and sentence types in the High-tone-ending sentences, which differ in the third and fourth syllable: (a) *māma*, (b) *yéye*, (c) *nǎinai*, and (d) *mèimei*.

Figures 4.7-4.8 show the duration of each syllable under different focus and sentence type conditions in the neutral-tone-ending and High-tone-ending sentences, respectively. Repeated measures ANOVAs of duration on sex, focus, sentence type, final tone, and preceding tone generated the following results (as shown in Table 4.4): 1) There is no significant difference in durations of the syllables between male and female speakers. 2) Focused words (*mǎi* and *māma/yéye/nǎinai/mèimei*) have longer durations than their non-focused counterparts. 3) The duration of the last syllable is longer in questions than in statements, although some earlier syllables show the opposite pattern. 4) When the tones of the final two syllables are High (as opposed to neutral), there seems to

be an anticipatory shortening effect on the durations of the two preceding neutral tones. Furthermore, the duration of *mā/yé/nǎi/mèi* is on average shorter in High-tone-ending sentences than in neutral-tone-ending sentences. 5) The tone of the third syllable shows no systematic effects on the durations of their preceding and following syllables.

Table 4.4. Results of the main effects from repeated measures ANOVAs of duration of each syllable on sex (Female vs. Male), focus (Nonadjacent: on *mǎi* vs. Adjacent: on *mā/yé/nǎi/mèi*), sentence type (Question vs. Statement), final tone (High vs. Neutral), and preceding tone (High, Rising, Low, or Falling). The effects with *p*-values less than 0.05 are bolded.

		Sex	Focus	Sentence type	Final tone	Preceding tone
tā	F	1.60	2.50	0.12	4.77	2.36
	p	0.2528	0.1651	0.7415	0.0717	0.1056
mǎi	F	0.98	**37.73**	**16.28**	2.45	**7.42**
	p	0.3611	**0.0009**	**0.0068**	0.1687	**0.0019**
			A<N	**Q<S**		**R>(L,H,F)**
mā/ yé/ nǎi/ mèi	F	0.05	**33.95**	1.28	**24.32**	**7.47**
	p	0.8277	**0.0011**	0.3018	**0.0026**	**0.0019**
			A>N		**H<N**	**L>(H,F)>R**
ma/ ye/ nai/ mei	F	0.02	**64.06**	**10.15**	3.10	**27.8206**
	p	0.8938	**0.0002**	**0.0189**	0.1290	**<0.0001**
			A>N	**Q<S**		**(F,H)>L>R**
men	F	0.02	2.61	0.58	**21.64**	**3.98**
	p	0.9016	0.1570	0.4742	**0.0035**	**0.0244**
					H<N	**(H,F)>(R,L)**
de	F	1.20	0.15	4.93	**78.85**	**8.71**
	p	0.3157	0.7139	0.0683	**0.0001**	**0.0009**
					H<N	**H>R & F>(L,R)**
le/ māo	F	2.54	0.14	**7.68**	**138.02**	2.51
	p	0.1621	0.7219	**0.0323**	**<0.0001**	0.0916
				Q<S	**H>N**	
ma/ mī	F	5.29	0.30	**51.09**	4.33	2.01
	p	0.0612	0.6058	**0.0004**	0.0825	0.1487
				Q>S		

In summary, analyses of pitch range (indicated by mean F_0 and max F_0 – min F_0) and duration of the syllables under different focus and sentence type conditions indicate that 1) There is a tri-zone pitch range modification of the focus effect in both statements and questions: pre-focus pitch range neutralization, on-focus pitch range expansion, and post-focus pitch range suppression. Nevertheless, pitch ranges of on- and post-focus syllables are higher and narrower in questions than in statements. 2) Focus lengthens the focused words (which may contain a neutral-tone syllable) in both statements and questions. 3) The sentence-final syllable is longer in questions than in statements, whereas syllables in the earlier part of the sentences may be shorter in questions than in statements. 4) Mean F_0 and pitch span of the neutral tone syllables are greatly influenced by focus, sentence type, and the preceding full tone.

4.2.4.3 Pitch target of the neutral tone

Chen and Xu (2006) have shown evidence that the neutral tone in Mandarin has its own target, although the target is executed with a weak articulatory force. Similar evidence is seen in the present data. First, as shown in Figure 4.5, despite the strong carryover influence from the preceding full tones, the F_0 contours of the neutral tone gradually converge over time, and the convergence is virtually complete by the end of the third neutral tone syllable in Figure 4.5a. The convergence is not complete until the last neutral tone in Figure 4.5c, but only because the post-L raising effect mentioned earlier keeps F_0 higher after the Low tone than after the other three tones. Such convergence suggests that F_0 of the neutral tone is approaching a particular target rather than being

fully determined by the preceding full tone. The nature of this target, however, has to be determined by separate analysis.

According to the TA model, a pitch target, defined in terms of both F_0 height and slope, is best approximated by the end of the syllable to which it is associated, as shown in Figure 4.2. Thus measurements taken at the end of a syllable may best reflect its underlying pitch target (evidence can be also seen in Chapter 3). To determine the slope of the neutral tone target, final velocities of the F_0 contours of the syllables in the sentences were extracted at 30 ms before the conventional syllable offset as indicators of the pitch targets of the syllables.[2] Principally, a large positive velocity corresponds to a [rise] target, a large negative velocity a [fall] target, and a velocity close to zero a [high/mid/low] target. Repeated measures ANOVAs of the mean final velocity of each syllable on sex, focus, sentence type, final tone, and preceding tone indicate that final velocity of the sentence-final syllable *ma/mī* is only marginally affected by sentence type ($F(1,6) = 6.31$, $p = 0.0458$). Thus, the pitch target of the neutral tone can potentially be inferred from the final velocity value of the sentence-final neutral tone *ma*. Table 4.5 displays mean final velocities of the neutral tone *ma* and the High tone *mī*. In statements, the final velocities of the two tones are not significantly different from zero, or from each other. In questions, the final velocities of the two tones are again not significantly different from each other, although both are significantly greater than zero. Assuming

[2] Measuring final velocity at 30 ms before the conventional syllable offset is based on the finding of Xu and Liu (2007) that real syllable boundaries, i.e., where one syllable ends and the next one begins, is about 30-50 ms earlier than the conventionally used acoustic landmarks, e.g., onset of nasal murmur in /ma/. Evidence of such leftward shift is also seen in the prototypical velocity profiles of Mandarin tones obtained through unsupervised learning in Gauthier et al. (2007). See also Xu & Liu (2006) for further theoretical discussions.

that the pitch target of the High tone is static, which is uncontroversial, and assuming that tones in statements are closer to their canonical form, the pitch target of the neutral tone should also be a static one.

Table 4.5. Mean final velocities (st/s) of sentence-final neutral tone and High tone, and the *t*-tests indicating whether they are significantly different from zero or from each other.

Tone / Intonation	Neutral	High	Neutral vs. High
Statement	-4.03 t(63) = -1.59 $p = 0.1167$	1.16 t(63) = 1.22 $p = 0.2285$	t(63) = 1.71 $p = 0.0924$
Question	4.76 t(63) = 9.25 $p < 0.0001$	3.89 t(63) = 10.98 $p < 0.0001$	t(63) = -1.93 $p = 0.0584$

As for the height of this static target, Chen and Xu (2006) concluded that it should be [mid] because its height is half way between the maximum F_0 of the Falling tone and the minimum F_0 of the Low tone in the same sentence position. In the current data, the mean F_0 of the neutral tone *ma* is significantly lower than that of the High tone *mī* ($F(1,6) = 37.74$, $p = 0.0009$, 86.80 st < 89.05 st). However, unlike in Chen and Xu (2006), there is no Low tone at the end of the sentence in the current data to act as a reference to the floor of the pitch range. Moreover, one could argue that the difference in mean F_0 between the High-tone syllable *mī* and the neutral-tone syllable *ma* is attributable to the difference in intrinsic F_0 between high and low vowels. Cross-linguistically, high vowels (e.g. /i/) are known to have higher F_0 than low vowels (e.g. /a/) (Whalen & Levitt, 1995). Nevertheless, in Torng's (2000) study of tonal production in Mandarin Chinese, /i/ had exceptionally lower F_0 than the other high vowels /u, y, o/, and only one third of /i/ was produced with a higher F_0 than /a/. Therefore, the F_0 difference between *mī* and *ma* is

likely to be due to different tonal targets of the High and neutral tones, rather than caused by the difference in vowel qualities.[3]

4.2.5 Discussion

Traditionally considered to be "targetless", the neutral tone at the question-final position has been assumed to be fully free to carry the rising intonation and therefore should be high in pitch (Qi, 1956). However, the present data suggest that the neutral tone is not any better than a full tone in manifesting question intonation, because the same post-focus lowering as seen in the High-tone-ending questions occurs in the neutral-tone-ending questions as well.

In summary, Experiment 4 shows that 1) the F_0 trajectories of the neutral tone are heavily influenced by the preceding tone, but they nevertheless gradually converge over time, although in questions the convergence is incomplete even at the end of the sentence. According to its final velocity values, the neutral tone has an underlying static pitch target; 2) the effects of focus and sentence type are exerted on the neutral tone via global pitch controls. In this experiment, the neutral tone sequence occurs in the post-focus region and at the sentence boundary. Thus, the post-focus pitch range suppression due to focus and the nonlinear post-focus pitch increase due to question intonation (Liu & Xu, 2005) combine to form the pitch pattern of the neutral tone sequence (and the neutral tone

[3] One could also argue that the F_0 contours of the neutral tone merely reflect some natural resting state of the voice rather than due to gradual approximation of any specific target. Note that returning to a resting state is not a mechanism fundamentally different from target approximation. The only difference is whether such return is due to passive muscle relaxation or active muscle contraction. Judging from the upward movement due to question intonation or due to the rather drastic post-L raising effect in Figure 4.5c and 4.5d, and given the precise control of the F_0 height in all the conditions observed in the current study, pure passive relaxation is an unlikely mechanism. Nonetheless, it is plausible that in a neutral statement a [mid] target is close to speakers' most comfortable pitch level. In this sense, the notion of a resting state could offer a partial account for the source of the [mid] target, although it does not explain how this target is approached.

+ High tone sequence as well); 3) the F_0 of the sentence-final High tones remains level in statements, but becomes slightly rising in questions. This is in contrast to the F_0 of the final neutral tones, which keeps falling not only in statements but also in questions; 4) syllables under focus are lengthened in both statements and questions; however, compared to statements, questions seem to be produced with quicker tempo at first but slow down at the end of the sentence. Overall, the above patterns indicate that, in Mandarin, sentence type is an independent intonational function whose manifestation is achieved through modifying the local tonal pitch target specified by the lexical tones, including the neutral tone.

4.3 Experiment 5: Statements and Yes/no Questions in General American English

Experiment 5 aims to address three unresolved issues in English intonation: 1) What is the temporal scope of the acoustic contrast between statements and yes/no questions? 2) How does focus affect the realization of this contrast? 3) Are there any pitch target variations involved in this contrast?

4.3.1 Materials

Speech materials consist of three sets of sentences, within which the final syllable of the last word is either stressed (e.g., ***Elaine***) or unstressed (e.g., ***Alan***). Each sentence was uttered as two sentence types (statement vs. yes/no question) and with two focus conditions (medial vs. final). The sentence type (./?) and focus (in bold face) conditions were elicited by different leading sentences (in parentheses), as shown below. Each sentence was repeated eight times by each subject, resulting in 960 utterances in total.

Note that here the leading sentences are meant to be spoken by the same speaker in a dialogue rather than by alternating speakers.

1. (Not an **internship**./?)
   You want a **job** with Microsoft./?

   (Not an **internship**./?)
   You want a **job** with La Massage./?

   (Not **La Massage**./?)
   You want a job with **Microsoft**./?

   (Not **Microsoft**./?)
   You want a job with **La Massage**./?

2. (It's not **fate**./?) (It's not **you**/**me**./?)
   There is something **unmarriable** about **me/May**./?

3. (It's not **Sears**./?) (It's not **Elaine/Alan**./?)
   You're going to **Bloomingdales** with **Alan/ Elaine**./?

4.3.2 Subjects

Three female and two male speakers, aged 18 – 30, served as subjects. They were raised in either California or the Midwest and spoke General American English. They had no self-reported speech or hearing disorders.

4.3.3 Procedure

Recordings were done in a sound-treated booth in the Language Labs at the University of Chicago, Chicago, Illinois. There were 8 sentence blocks (for 8 repetitions of each sentence) and a different randomization was applied to each block. Recording procedure was similar to that in Experiment 4, except that the utterances were first recorded onto a digital memory card using a solid-state recorder and then transferred to a computer. The digitized sounds were re-sampled at 22.05 kHz for later analyses.

Measurements and calculations for graphical and statistical analyses were the same as in Experiment 4.

4.3.4 Results

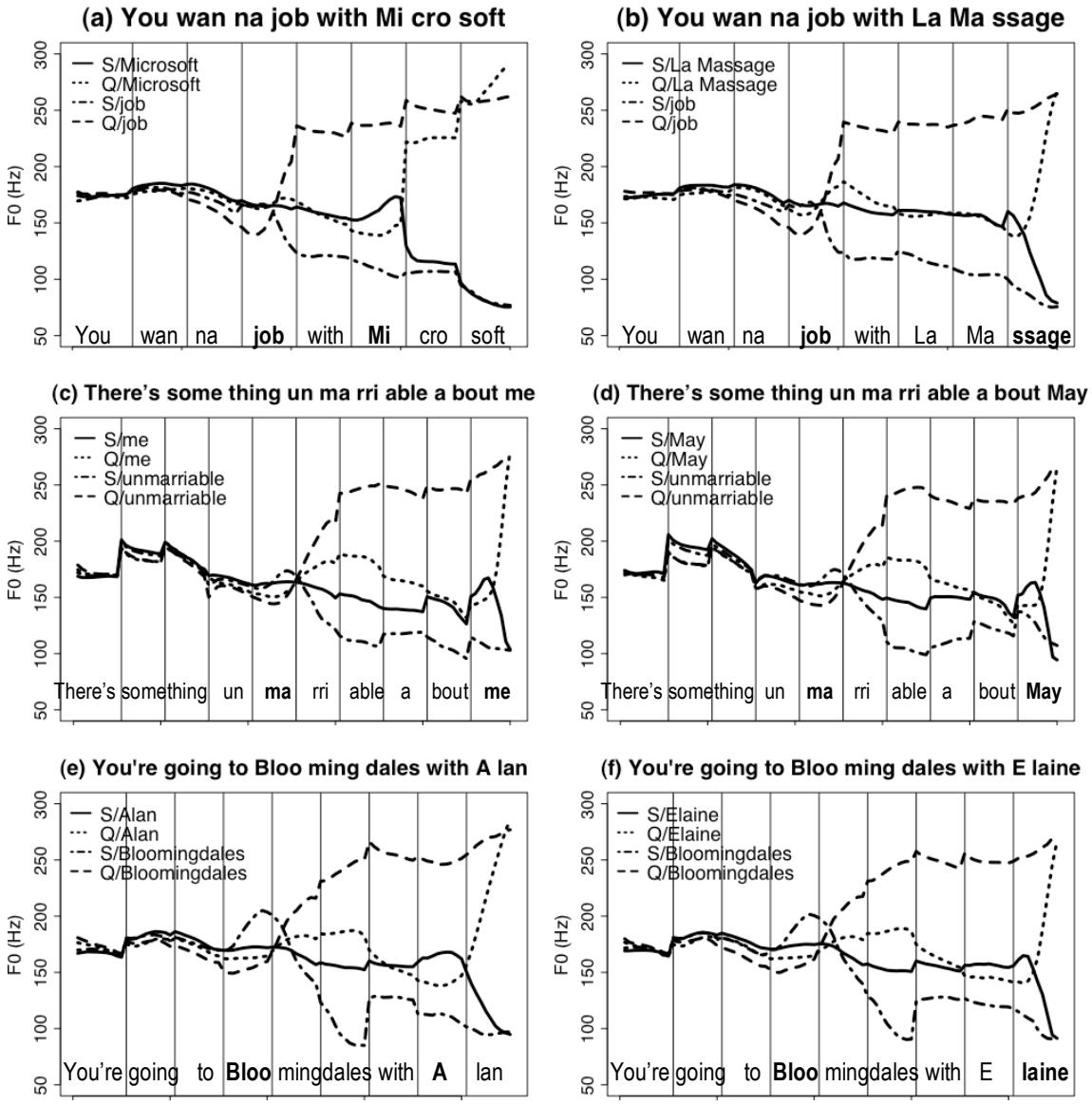

Figure 4.9. Time-normalized F_0 contours (averaged across 40 repetitions by 5 subjects) of speech materials in English. Vertical lines indicate syllable boundaries. In the legend, "S" stands for statement, and "Q" for yes/no question. "S/Microsoft" means a statement with focus on "Microsoft", and so on.

Fig. 4.9 displays mean time-normalized (with syllable as the normalization domain as done in Experiment 4) F_0 contours (averaged across 40 repetitions by 5 subjects) of the speech materials under different focus conditions and in different sentence types. The following patterns can be seen directly from the overlaid F_0 contours, which will be later verified with statistical analysis. 1) The difference in the F_0 contours between statements and yes/no questions in English seems to start from the stressed syllable of the first content word (focused or non-focused). 2) Focus has no effect on the pre-focus region in either statements or questions. In the on-focus region, the pitch range of the stressed syllable is expanded in both statements and questions. The post-focus pitch range is compressed and lowered in statements, but compressed and raised in questions. 3) The pitch target of the stressed syllable in a focused non-final-stressed word (e.g. *Bloomingdales* in (e) and (f)) seems to be high in statements, but rising in questions, and that in a focused final-stressed word (e.g., *May* in (d), *job* in (a) and (b), and *Elaine* in (f)) appears to be falling in statements, but rising in questions. In utterances with final focus (e.g., *You're going to Bloomingdales with* ***Alan*** in (e)), the pitch target of the stressed syllable in a pre-focus content word (e.g. *Bloomingdales*) seems to be high in statements but rising in questions. In utterances with medial focus (e.g., *You're going to* ***Bloomingdales*** *with Alan/Elaine* in (e) and (f)), the pitch target of the stressed syllable in a post-focus content word (e.g. *Alan/Elaine*), despite the pitch range compression, is still recognizable: *A-* in *Alan* is high in statements but rising in questions; *-laine* in *Elaine* is falling in statements but rising in questions.

4.3.4.1 Effects of sentence type, focus, and final stress on F_0

Table 4.6. Mean F_0 (in st) and pitch span (in st) of syllables in English statements and questions. Potential focus locations are bolded.

Sentence type	Question				Statement			
Focus	Medial		Final		Medial		Final	
Pitch range	Mean F_0	Pitch span	Mean F_0	Pitch span	Mean F_0	Pitch span	Mean F_0	Pitch span
you/you	89.5/89.6	1.8/1.8	89.4/89.2	1.7/1.8	89.2/89.3	1.9/1.8	89.3/89.3	1.6/1.7
wan/wan	89.7/89.7	1.6/1.5	89.7/89.6	1.4/1.3	89.9/89.8	1.3/1.4	90.2/90.1	1.4/1.1
na/na	88.2/88.0	2.9/2.8	89.5/89.6	1.8/1.8	89.0/88.7	2.2/2.2	89.8/89.8	2.0/1.8
job/job	**88.1/88.1**	**7.0/7.2**	88.6/88.6	2.2/3.2	**87.9/87.9**	**6.9/7.5**	88.5/88.6	1.5/1.4
with/with	94.2/94.5	1.1/1.2	87.8/89.4	2.6/2.8	82.5/82.0	1.8/2.1	87.8/88.0	1.6/1.6
Mi/La	94.7/94.7	1.2/1.1	**85.8**/87.6	**2.6**/1.9	81.1/82.3	4.0/2.9	**88.0**/87.9	**3.2**/1.5
cro/Ma	95.7/94.9	1.1/1.0	93.6/87.3	1.4/2.1	80.3/80.2	1.5/2.3	82.0/87.3	2.7/2.2
soft/ssage	96.2/95.8	1.4/2.0	96.8/**91.0**	2.5/**11.7**	75.5/77.5	5.7/8.1	74.8/**85.1**	7.0/**15.3**
there's/there's	89.2/89.0	2.1/1.6	89.0/88.8	1.6/1.9	88.9/89.0	1.4/1.9	88.7/89.1	1.9/1.5
some/some	90.5/90.2	2.1/2.2	90.5/90.2	2.0/2.0	91.0/91.1	1.9/2.0	91.2/91.5	1.9/1.8
thing/thing	90.3/90.3	3.1/2.8	90.7/90.3	2.7/2.7	90.5/90.6	3.1/2.6	90.7/90.9	2.7/3.1
un/un	87.6/87.5	3.2/2.7	88.1/87.9	2.0/2.5	88.3/88.6	2.1/2.5	88.6/88.5	2.0/2.8
ma/ma	**86.6/86.3**	**2.1/2.1**	87.1/87.2	1.4/1.4	**88.4/88.5**	**2.3/2.2**	88.1/88.1	1.0/1.2
rri/rri	91.4/91.2	5.6/5.6	89.4/89.3	2.5/2.0	86.6/87.0	6.4/5.4	87.7/87.6	1.7/2.0
able/able	95.3/95.2	1.2/1.5	90.4/90.1	1.8/1.7	80.8/79.6	3.6/3.6	86.7/86.2	2.2/2.0
a/a	95.3/94.4	1.2/1.4	88.2/88.0	1.6/1.8	82.5/81.1	1.8/2.9	85.2/86.7	1.6/1.2
bout/bout	95.4/94.5	1.8/1.5	86.6/86.1	3.7/4.0	80.5/82.9	4.4/3.3	86.0/86.5	4.1/3.5
me/May	96.5/95.5	2.0/2.9	**90.3/89.6**	**12.6/11.6**	80.4/83.9	3.8/6.0	**87.0/86.6**	**10.3/11.9**
you're/you're	89.4/89.3	2.2/2.2	89.1/89.1	1.9/2.0	88.8/89.0	1.9/1.8	88.6/88.8	1.7/1.6
going/going	89.6/89.8	2.0/2.0	89.6/89.7	1.6/1.8	90.0/90.1	1.6/1.6	90.2/90.2	1.7/1.7
to/to	88.6/88.3	2.5/2.1	89.3/89.3	1.7/1.7	89.4/89.3	1.6/1.7	89.8/89.8	1.8/1.4
Bloo/Bloo	**87.1/87.3**	**2.2/2.1**	88.2/88.2	1.3/1.3	**90.7/90.5**	**3.8/3.5**	89.0/89.2	1.3/1.5
ming/ming	91.5/91.7	5.2/5.0	89.7/89.6	2.2/2.1	89.5/89.4	6.6/6.4	88.7/88.8	1.9/2.1
dales/dales	94.9/94.9	1.9/2.0	90.4/90.5	1.4/1.4	80.7/81.8	8.0/7.8	87.4/87.1	1.7/1.5
with/with	96.0/95.6	1.4/1.3	87.6/88.4	2.7/2.7	83.5/83.5	2.0/2.1	87.4/87.4	1.3/1.5
A/E	95.5/95.6	1.7/1.0	**85.7**/86.3	**3.0**/3.0	80.9/82.7	5.0/1.9	**88.4**/87.5	**4.2**/2.1
lan/laine	96.7/96.1	1.9/2.0	93.8/**90.4**	11.3/**11.7**	78.7/80.0	4.8/7.5	83.5/**86.5**	10.3/**12.2**

Mean F_0 and pitch span of the syllables in these sentences (grouped according to focus, sentence type, and final stress) are listed in Table 4.6. Two × three sets of repeated

measures ANOVAs were conducted, where mean F_0 or pitch span of each syllable in the three sets of sentences is treated as the dependent variable, sex as the between-subjects factor, and focus, sentence type, and final stress as the within-subjects factors.

Table 4.7. Effects of sentence type and sentence type × focus on mean F_0 in the repeated measures ANOVAs for the three sets of sentences. The effects with *p*-values less than 0.05 are bolded. Here, "Q" stands for question, and "S" for statement.

Sentence set 1		Sentence type	Sentence type × focus	Sentence set 2		Sentence type	Sentence type × focus	Sentence set 3		Sentence type	Sentence type × focus
you	F	0.06	0.76	There's	F	0.07	0.15	You're	F	1.52	0.00
	p	0.8167	0.4486		*p*	0.8151	0.7281		*p*	0.3059	0.9720
wan	F	0.20	0.30	some	F	2.04	0.46	going	F	0.58	0.24
	p	0.6862	0.6204		*p*	0.2487	0.5445		*p*	0.5011	0.6599
na	F	0.38	0.36	thing	F	0.22	0.06	to	F	1.62	0.64
	p	0.5832	0.5919		*p*	0.6708	0.8211		*p*	0.2931	0.4815
job	F	0.01	0.02	un	F	1.53	0.85	Bloo	F	6.15	**17.99**
	p	0.9112	0.9003		*p*	0.3041	0.4256		*p*	0.0894	**0.0240**
with	F	**25.26**	**36.21**	ma	F	3.09	9.38	ming	F	3.10	7.79
	p	**0.0152**	**0.0092**		*p*	0.1771	0.0549		*p*	0.1764	0.0683
		Q>S									
Mi /La	F	**207.97**	**139.91**	rri	F	**10.21**	6.58	dales	F	**23.47**	**22.53**
	p	**0.0007**	**0.0013**		*p*	**0.0495**	0.0829		*p*	**0.0168**	**0.0177**
		Q>S				**Q>S**				**Q>S**	
cro /Ma	F	**482.41**	**53.70**	able	F	**20.83**	**28.61**	with	F	**138.12**	**215.16**
	p	**0.0002**	**0.0053**		*p*	**0.0197**	**0.0128**		*p*	**0.0013**	**0.0007**
		Q>S				**Q>S**				**Q>S**	
soft /sagge	F	**113.14**	**20.96**	a	F	**41.32**	**28.42**	A/E	F	**94.26**	**182.44**
	p	**0.0018**	**0.0196**		*p*	**0.0076**	**0.0129**		*p*	**0.0023**	**0.0009**
		Q>S				**Q>S**				**Q>S**	
				bout	F	**140.86**	**45.79**	lan /laine	F	**62.44**	**265.51**
					p	**0.0013**	**0.0066**		*p*	**0.0042**	**0.0005**
						Q>S				**Q>S**	
				me /May	F	**101.93**	**128.60**				
					p	**0.0021**	**0.0015**				
						Q>S					

As seen in Table 4.7, repeated measures ANOVAs on mean F_0 indicate that statements have significantly lower mean F_0 than questions after the stressed syllable in

the first content word (focused or non-focused). More specifically, in sentence set 1, the first four syllables do not differ in mean F_0 in statements and questions. However, those syllables after *job* all show higher mean F_0 in questions than in statements. Similarly, in sentence set 2, there is no significant difference in mean F_0 for the first five syllables in statements and questions. However, the syllables following the stressed syllable *ma* (as in *unmarriable*) all reach higher mean F_0 in questions than in statements. Again, in sentence set 3, mean F_0's of the first four syllables are not significantly different between statements and questions. However, after the stressed syllable *Bloo* (as in *Bloomingdales*), except for *ming*, all the other syllables have significantly higher mean F_0 in questions than in statements.

There is no significant main effect of focus on mean F_0's of the syllables in either set of sentences. However, the interaction of sentence type × focus on mean F_0 is significant for the syllables in the later part of these sentences (especially after the medial focus). This is due to the facts that post-focus pitch range is raised in questions but lowered in statements, and that pitch targets of the stressed syllables in the content words (focused or non-focused) are [high] or [fall] in statements but [rise] in questions (as will be discussed in 4.3.4.3). Specifically, in sentence set 1, mean F_0's of the syllables *with Mi/La cro/ma soft/ssage* differ dramatically according to their relative position with the medial (*job*) and final (*Mi* or *sagge*) focus in statements and questions. The same phenomena occur for the later syllables in sentence sets 2 and 3, as shown in Table 4.7.

The main effect of final stress (i.e., the sentence-final syllable is either unstressed or stressed, as in *Micro**soft***/*La Ma**ssage***, ***me***/***May***, and ***A**lan*/*E**laine***) on mean F_0's of the

syllables is not statistically significant in either set of the sentences. However, since the interaction of sentence type × focus becomes salient from the stressed syllable of the focused word, the three-way interaction of sentence type × focus × final stress on mean F_0 is significant for the following syllables: *cro/ma* ($F(1,3) = 13.72$, $p = 0.0342$) and *soft/ssage* ($F(1,3) = 445.98$, $p = 0.0002$) as in *Microsoft*/*La Massage* in sentence set 1, *me/May* ($F(1,3) = 10.47$, $p = 0.0480$) in sentence set 2, and *lan/laine* ($F(1,3) = 76.02$, $p = 0.0032$) as in *Alan*/*Elaine* in sentence set 3. That is, mean F_0 of these syllables is high in medial-focused questions, but low in medial-focused statements. However, in final-focused sentences, their mean F_0 values are affected by both sentence type and the stress pattern of the final word.

In the repeated measures ANOVAs on pitch span of each syllable in the three sets of sentences, the effects of sex, sentence type, and final stress are in general not statistically significant. The effect of medial focus seems to not only enlarge the pitch span of the focused stressed syllable (*job* ($F(1,3) = 16.01$, $p = 0.0280$) in sentence set 1, and *ma* ($F(1,3) = 20.50$, $p = 0.0202$) in focused *unmarriable* in sentence set 2), but to enlarge the pitch span of the first post-focus unstressed syllable, such as *rri* ($F(1,3) = 24.35$, $p = 0.0160$) in focused *unmarriable* in sentence set 2 and *ming* ($F(1,3) = 16.14$, $p = 0.0277$) in focused *Bloomingdales* in sentence set 3. This immediate post-focus pitch span expansion seems due to the F_0 lowering by the creaky voice that occurs right after the focused stressed syllable in statements, and due to the inertia after the [rise] target of the focused stressed syllable in questions. The effect of final focus is to enlarge the pitch span of the sentence-final syllable, stressed or unstressed, such as *soft/ssage* ($F(1,3) =$

376.03, $p = 0.0003$) in sentence set 1, *me/May* ($F(1,3) = 1113.78$, $p < 0.0001$) in sentence set 2, and *lan/laine* ($F(1,3) = 61.76$, $p = 0.0043$) in sentence set 3. One thing that is noteworthy in regard to the focus effect in English is that the pitch span of some syllables is not necessarily significantly greater when they are focused than not focused. Examples include *Mi/La* ($F(1,3) = 0.01$, $p = 0.9434$) in *Microsoft/La Massage* in sentence set 1, *Bloo* ($F(1,3) = 9.53$, $p = 0.0539$) in *Bloomingdales* and *A/E* ($F(1,3) = 2.61$, $p = 0.2048$) in *Alan/Elaine* in sentence set 3. For *Mi/La* and *A/E*, this is due to the significant interaction of sentence type × focus on their pitch span (*Mi/La*: ($F(1,3) = 12.93$, $p = 0.0369$; *A/E*: $F(1,3) = 16.65$, $p = 0.0266$): in statements, *Mi/La* and *A/E* do not have greater pitch span when focused than when non-focused, although this is always true in questions. Again, this phenomenon is likely caused by pitch target shift between statements ([high] or [fall]) and questions ([rise]), as will be discussed in 4.3.4.3.

In summary, F_0 differences between statements and yes/no questions in English become salient starting from the stressed syllable of the first content word, whether or not it is focused. The interaction of sentence type × focus on F_0 height of post-focus syllables is significant, with those in statements being low and those in questions being high. With different pitch targets in statements and questions, focused stressed syllables may or may not have greater pitch spans than their non-focused counterparts. The pitch span of the first post-focus unstressed syllable (when in the same word as the focused stressed syllable) is enlarged in both statements and questions, presumably to implement the large post-focus lowering in a statement or raising in a question.

4.3.4.2 Syllable durations in statements vs. yes/no questions in English

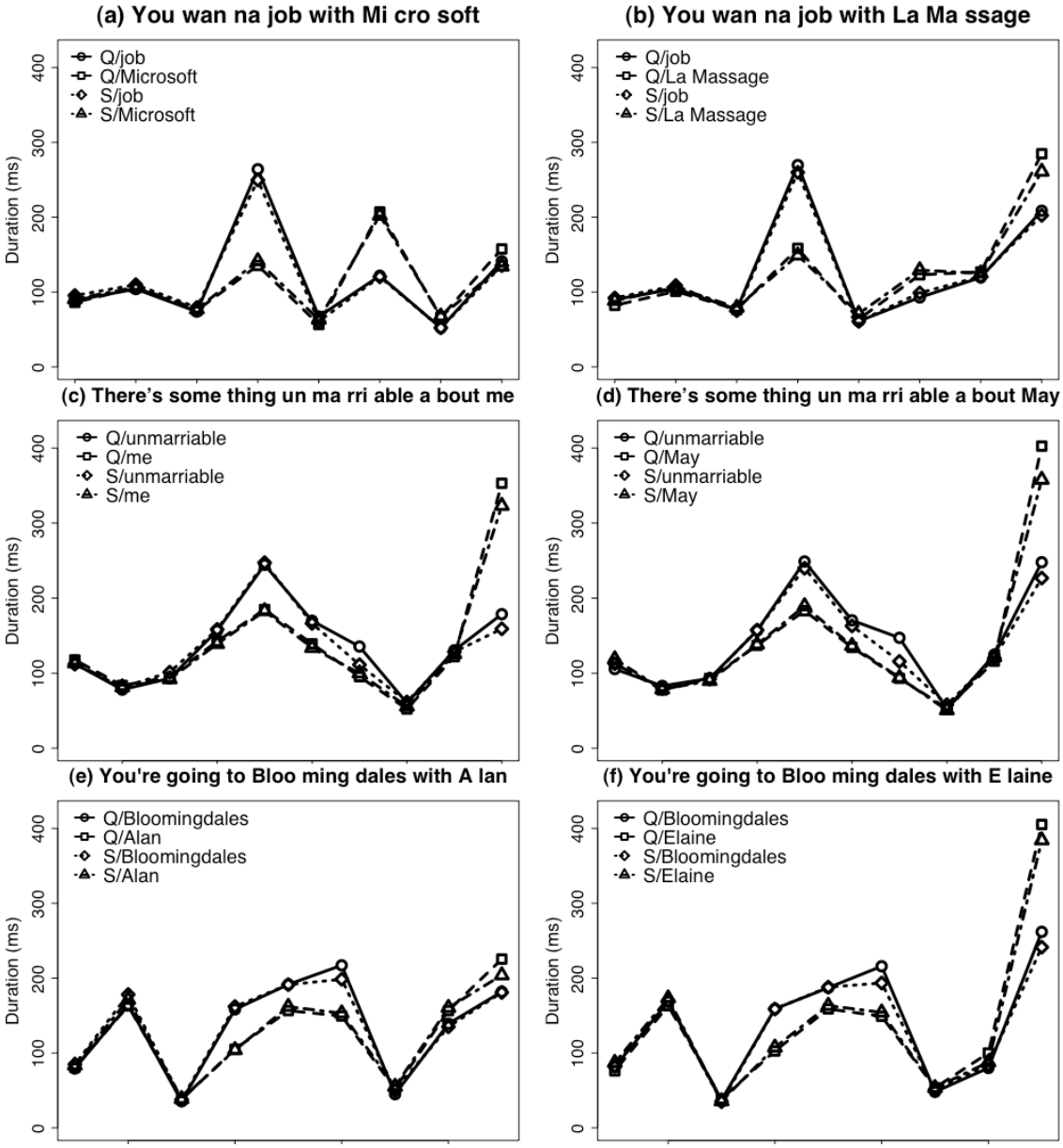

Figure 4.10. Durations of the syllables under different focus conditions and sentence types in English.

Figure 4.10 displays the duration of each syllable under different focus and sentence type conditions in the three sets of sentences. Repeated measures ANOVAs

were conducted, where syllable duration is the dependent variable, sex is the between-subjects factor, and focus, sentence type, and final stress are the within-subjects factors.

Table 4.8. Effects of sentence type and focus on duration in the repeated measures ANOVAs for the three sets of sentences. The effects with *p*-values less than 0.05 are bolded.

Sentence set 1		Sentence type	Focus	Sentence set 2		Sentence type	Focus	Sentence set 3		Sentence type	Focus
you	F	1.31	1.13	There's	F	0.12	1.63	You're	F	2.57	0.00
	p	0.3358	0.3649		*p*	0.7562	0.2915		*p*	0.2070	0.9843
wan	F	3.89	0.40	some	F	0.02	0.01	going	F	5.47	0.11
	p	0.1431	0.5732		*p*	0.8906	0.9384		*p*	0.1014	0.7579
na	F	1.12	0.10	thing	F	0.02	0.38	to	F	0.17	0.45
	p	0.3679	0.7676		*p*	0.9038	0.5822		*p*	0.7060	0.5497
job	F	1.60	**843.05**	un	F	0.00	4.88	Bloo	F	0.48	**49.72**
	p	0.2951	**<0.0001**		*p*	0.9945	0.1141		*p*	0.5390	**0.0059**
			Medial>Final								**Medial>Final**
with	F	0.61	0.04	ma	F	0.09	**15.84**	ming	F	1.42	**22.95**
	p	0.4907	0.8597		*p*	0.7865	**0.0284**		*p*	0.3195	**0.0173**
							Medial>Final				**Medial>Final**
Mi /La	F	0.06	**301.35**	rri	F	4.61	**15.06**	dales	F	0.55	**969.87**
	p	0.8215	**0.0004**		*p*	0.1211	**0.0303**		*p*	0.5109	**<0.0001**
			Final>Medial				**Medial>Final**				**Medial>Final**
cro /Ma	F	0.01	4.26	able	F	**23.58**	**23.40**	with	F	1.56	4.66
	p	0.9260	0.1311		*p*	**0.0167**	**0.0169**		*p*	0.3005	0.1197
						Q>S	**Medial>Final**				
soft /sagge	F	1.26	**38.52**	a	F	1.13	6.35	A/E	F	0.37	**12.29**
	p	0.3433	**0.0084**		*p*	0.3650	0.0863		*p*	0.5859	**0.0393**
			Final>Medial								**Final>Medial**
				bout	F	1.13	0.53	lan /laine	F	3.14	**115.23**
					p	0.3663	0.5203		*p*	0.1747	**0.0017**
											Final>Medial
				me /May	F	3.73	**34.89**				
					p	0.1490	**0.0097**				
							Final>Medial				

Results of the ANOVA models indicate that (1) syllable duration does not vary systematically with either speaker sex or sentence type; (2) as seen in Table 4.8, focus tends not only to lengthen the duration of the focused stressed syllable, but to lengthen

the durations of its following syllables that are also within the focused word; (3) the effect of final stress on syllable duration is confounded with the intrinsic durations of the phonemes in the final words (*Microsoft/La Massage*, *me/May*, and *Alan*/*Elaine*), and therefore will not be discussed further.

In summary, syllable durations are not significantly different in statements and yes/no questions in English in the present data. Furthermore, focus tends to lengthen the durations of the on- and post-focus syllables in the focused word in both statements and questions.

4.3.4.3 Pitch targets of stressed syllables in English

Similar to Experiment 4, the mean of final velocities from the 8 repetitions for each syllable by each subject was calculated, and Table 4.9 lists the mean final velocities (averaged across the means of 5 subjects) of the on-focus stressed syllables grouped by stress pattern and sentential position of the word. Because final velocities of on-focus stressed syllables do not vary with different focus locations ($F(1,3) = 6.12$, $p = 0.0898$), medial and final focus are not treated separately in Table 4.9.

Table 4.9. Mean final velocities (st/s) of on-focus stressed syllables, and the *t*-tests indicating whether they are significantly different from zero. Here and subsequently, "wfinal_nonsfinal" stands for word-final and non-sentence-final, and "wfinal_sfinal" for word-final and sentence-final.

Intonation	Non-final	W-final Non-s-final	W-final S-final
Question	29.70 $t(29) = 5.95$ $p < 0.0001$	62.83 $t(9) = 9.51$ $p < 0.0001$	41.00 $t(19) = 12.72$ $p < 0.0001$
Statement	-7.57 $t(29) = -1.80$ $p = 0.0819$	-52.10 $t(9) = -6.03$ $p = 0.0002$	-53.49 $t(19) = -6.64$ $p < 0.0001$

Repeated measures ANOVAs with sex as the between-subject factor and sentence type and stress pattern of the on-focus word as within-subject factors indicate that on-focus stressed syllables have significantly different final velocities in statements and questions ($F(1,3) = 232.46$, $p = 0.0006$). Furthermore, although the stress pattern of the on-focus word has no significant main effect on final velocity ($F(2,6) = 2.84$, $p = 0.1357$), its interaction with sentence type is statistically significant ($F(2,6) = 9.80$, $p = 0.0129$).

To determine the pitch targets of on-focus stressed syllables in different sentence types, *t*-tests were conducted to see if their final velocities are significantly different from zero. Results in Table 4.9 suggest that the pitch target of the on-focus stressed syllable is likely to be [high] (in non-final-stressed words) or [fall] (in final-stressed words) in statements, but [rise] in questions.

Table 4.10 shows the mean final velocities of stressed syllables in pre- and post-focus content words in statements and questions, and the *t*-test results indicate whether they are significantly different from zero. Pre- and post-focus conditions are grouped together because final velocities of stressed syllables in content words do not differ according to their position relative to focus ($F(1, 3) = 5.69$, $p = 0.0971$).

Table 4.10. Mean final velocities of pre/post-focus stressed syllables, and the corresponding *t*-test results.

Intonation	Non-final	W-final Non-s-final	W-final S-final
Question	10.24 $t(29) = 3.68$ $p = 0.0009$	17.72 $t(9) = 2.47$ $p = 0.0353$	8.48 $t(19) = 5.11$ $p < 0.0001$
Statement	-9.63 $t(29) = -2.20$ $p = 0.0357$	-2.05 $t(9) = -0.66$ $p = 0.5238$	-15.96 $t(19) = -4.38$ $p = 0.0003$

Repeated measures ANOVAs with sex as the between-subject factor and sentence type and stress pattern of the pre-/post-focus content word as within-subject factors indicate that final velocities of stressed syllables in pre- and post-focus content words differ in different sentence types ($F(1,3) = 12.14$, $p = 0.0399$). The corresponding *t*-tests indicate that the pitch target of the pre- or post-focus stressed syllable in a content word is [high] (if the stress is word-final but non-sentence-final) or [fall] (if the stress is non-final or word-final and sentence-final) in statements, but [rise] in questions. Note that the velocity values of the non-final stressed syllables in pre-/post-focus content words in statements are only marginally significantly different from zero ($t = -2.20$, $p = 0.0357$, as in Table 4.10), so they may have a [high] rather than [fall] target as per perception.

In summary, these results suggest that the pitch targets of the stressed syllables in English are [high] or [fall] in statements, and [rise] in questions, depending on the focus condition and the stress pattern of the content words to which the stressed syllables belong.

4.3.5 Discussion

Results of Experiment 5 suggest that the intonational contrast between statements and yes/no questions in General American English involves variations of both pitch range of post-focus words and pitch targets of stressed syllables in content words (focused or non-focused). Specifically, the pitch range of post-focus words is compressed and lowered in statements, but compressed and raised in questions, which agrees with the general findings in Eady and Cooper (1986) and Pell (2001), but is verified here by a more detailed syllable-by-syllable (instead of word-by-word) analysis. Statements and

yes/no questions also differ in underlying pitch targets of stressed syllables in content words, with the former having a [high] or [fall] pitch target (depending on the stress pattern, sentential position, and focus condition) and the latter having a [rise] pitch target. Eady and Cooper (1986) also noticed pitch contour differences in content words between statements and questions. However, being concerned only with surface forms and without distinguishing words with different stress patterns, they did not reach a clear conclusion regarding the local pitch variations related to the statement/question contrast.

Furthermore, similar to what is found in Eady and Cooper (1986) but with a more detailed syllable-by-syllable analysis, Experiment 5 shows that focus lengthens the duration of the focused word in both statements and questions. In addition, there seems to be no significant difference in syllable duration between statements and yes/no questions with medial/final focus in English as shown by the current data. While Eady and Cooper (1986) reached similar conclusions (based on a word-by-word analysis) for their stimuli with initial/final focus, they nonetheless found that sentence-final words are significantly longer in questions with neutral focus than in statements with neutral focus.

In summary, the following conclusions regarding English intonation seem plausible: 1) The F_0 difference between statements and yes/no questions becomes salient starting from the stressed syllable of the first content word, whether or not it is focused. This manifestation of the sentence type difference is due to a pitch target shift that changes the underlying pitch targets of the stressed syllables from [high] or [fall] (depending on their position within the word and sentence, and on whether they are focused or not) in statements to [rise] in questions. 2) Focus expands the pitch range of

the focused word (whose duration is also lengthened), compresses and lowers (in statements) or compresses and raises (in questions) that of the post-focus words, and leaves that of pre-focus words largely unaffected.

4.4 Summary of Experiments 4 and 5

Experiments 4 and 5 examined statement and question intonation in Mandarin and English using the PENTA model, which simultaneously identifies communicative functions and articulatory mechanisms in F_0 production. Results of the F_0 analysis indicated that the two languages differ at least in the following aspects: A) the functional domain of statement/question intonation, B) the role that focus plays in differentiating sentence types, and C) the interaction between lexical tone/stress, focus, and sentence type. More specifically, 1) statement and question intonation in Mandarin start to diverge from the focused word, with F_0 of questions becoming increasingly higher than that of statements. In English, however, due to a pitch target shift (from [high] or [fall] in statements to [rise] in questions), the F_0 difference between statements and yes/no questions becomes salient starting from the stressed syllable of the first content word, whether or not it is focused; 2) in Mandarin and English, there exist both pre-focus pitch range neutralization and on-focus pitch range expansion, but post-focus pitch range behaves differently in statements and questions in the two languages. In Mandarin, post-focus suppression (compression and lowering) occurs in both statements and questions, although the latter exhibiting a smaller magnitude. In English, however, the pitch range of the post-focus words was compressed and lowered in statements but compressed and raised in questions; 3) in Mandarin, although tonal contours are slightly modified by

focus and sentence type, tonal targets remain intact. In English, however, the underlying pitch target of the stressed syllable in a content word varies systematically both with syllable position in word/sentence, and with focus and sentence type. Finally, focused words have longer duration than their non-focused counterparts in statements and questions of both languages. In Mandarin, while syllables in the early part of the sentence are produced with shorter duration in questions than in statements, the last syllable shows the opposite pattern. In English, however, syllable durations are not significantly different in the medial/final-focused statements and questions examined here.

In summary, by simultaneously identifying several contributing factors to speech intonation, including language, sentence type, focus, and lexical tone/stress, this chapter shows that (1) in both Mandarin and English, statement/question intonation is realized in parallel with focus and lexical items that also use pitch for their encoding, and (2) the similarities and differences between Mandarin and English intonation are essentially caused by the way sentence type interacts with focus and lexical tone/stress in the two languages. Theoretical significance of these findings will be further discussed in the following chapter.

5 INTONATION SYSTEMS OF MANDARIN AND ENGLISH: A FUNCTIONAL APPROACH

Using data from previous chapters, this chapter focuses on comparing intonation systems of Mandarin and English through a functional framework – the PENTA model. The theoretical differences between the PENTA model and the AM theory of English intonation will also be discussed.

5.1 General Discussion — Comparing the two languages

The previous chapter investigated how lexical tone/stress, focus, and sentence type interact in Mandarin and General American English. More specifically, the following research questions were addressed: 1) would statement/question intonation override lexical tone/stress, thus causing pitch target variations in the realization of lexical tone/stress in different sentence types? 2) how does focus affect the realization of lexical tone/stress and sentence type?

To answer these questions, one has to first make the intonation systems of Mandarin and English comparable. To achieve this goal, Experiment 5 adopts the notion of "English as a tone language" (Goldsmith, 1981) in analyzing F_0 contours of stressed syllables in English using underlying pitch targets, as is done for lexical tones in Mandarin. As can be seen in Figure 5.1, F_0 contours of Mandarin tones (Here in the sentence-final position, and data were extracted from Experiment 1) are modified by both sentence type (statement vs. yes/no question) and focus (initial, medial, final, and neutral). That is, the heights and slopes of sentence-final lexical tones are enlarged by both question and final/neutral focus, but reduced by statement and initial/medial focus.

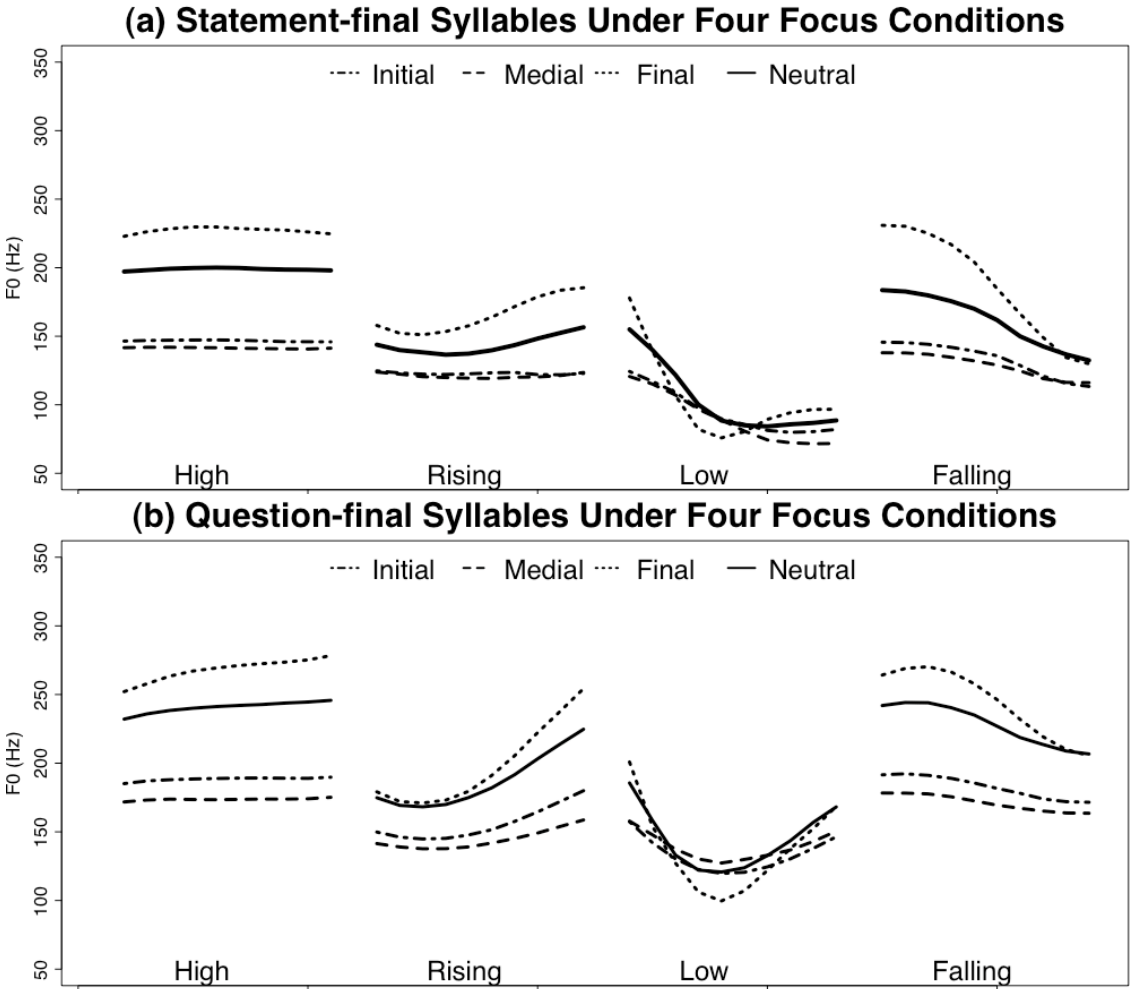

Figure 5.1. Time-normalized mean F_0 contours (averaged across 40 repetitions by 8 subjects from Experiment 1) of statement-/question-final syllables (with four full tones: High, Rising, Low, and Falling) under four focus conditions in Mandarin: (a) statement-final, and (b) question-final.

To assess pitch targets of lexical tones in Mandarin with measurable values, mean final velocities of the sentence-final syllables in statements and yes/no questions under initial/medial/final/neutral focus (data illustrated in Figure 5.1) were calculated and plotted in Figure 5.2. As can be seen, lexical tones in Mandarin tend to have greater final velocities in questions (Figure 5.2b) than in statements (Figure 5.2a), and when under final/neutral focus than when under initial/medial focus (in terms of absolute values for

the High tone in statements and the Falling tone in statements/questions). This indicates that pitch targets of lexical tones in Mandarin are modified under different focus and sentence type conditions. Nevertheless, even under different conditions as shown in Figure 5.2, final velocities of the High tone are always around zero, those of the Rising tone mostly positive, those of the Falling tone always negative, and those of the Low tone mostly positive (indicating the Falling-Rising allophone) and sometimes around zero (indicating the Low allophone). Therefore, there seems to be no categorical target shift for the Mandarin tones due to either focus or sentence type, except for the Low tone, which seems to show a shift between two alternative targets: [low] and [low+rise]. Similarly for the neutral tone, although showing a slight fall in surface form in both statements and questions in Experiment 4, its similar final velocity values to those of the High tone as shown in Table 4.5 indicate that it may indeed have a static target implemented with a weak articulatory strength in both sentence types (as discussed in section 4.2.4.3).

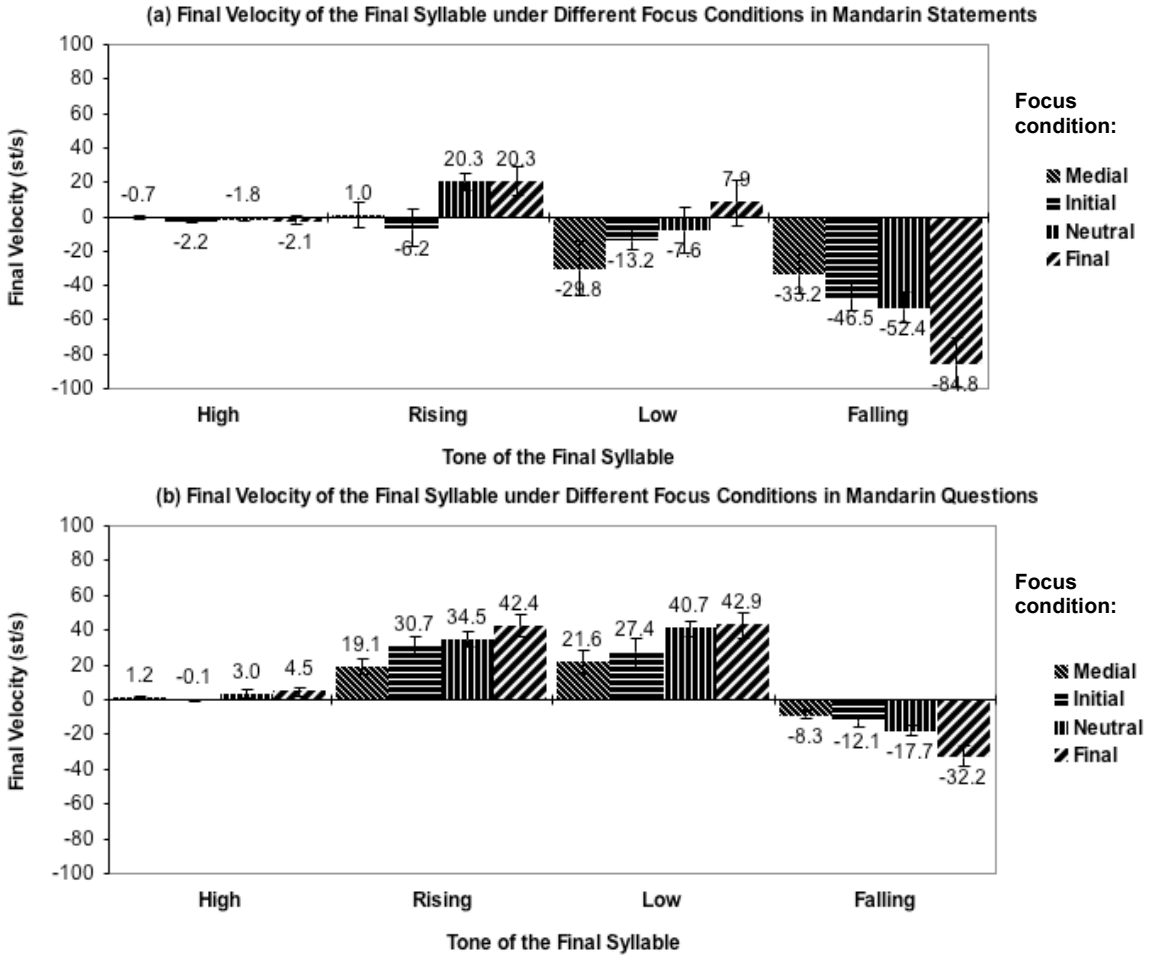

Figure 5.2. Final velocities of the sentence-final syllables in Experiment 1 (with four full tones: High, Rising, Low, and Falling) under different focus conditions (initial/medial/final/neutral) in Mandarin: (a) statements, and (b) yes/no questions.

To compare pitch targets of stressed syllables in English with pitch targets of lexical tones in Mandarin, final velocities of the F_0 contours of the on- and pre-/post-focus stressed syllables were plotted in Figure 5.3 according to three groups: non-final, word-final but non-sentence-final, and sentence-final (see also Tables 4.9-4.10 in section 4.3.4.3). As can be seen in Fig. 5.3a, all three groups of on-focus stressed syllables have large positive values in questions, some even larger than those of the Mandarin Rising tone in similar focus and sentence type conditions, thus indicating a [rise] target.

However, in statements, non-final stressed syllables seem to be produced with a [high] target, whereas word-/sentence-final stressed syllables with a [fall] target. For pre-/post-focus stressed syllables in Fig. 5.3b, again, they all have a [rise] target in questions. But only sentence-final stressed syllables show a [fall] target in statements. Therefore, unlike Mandarin, pitch targets of the stressed syllables in English vary with their position in the word and in the sentence, and with focus as well as sentence type.

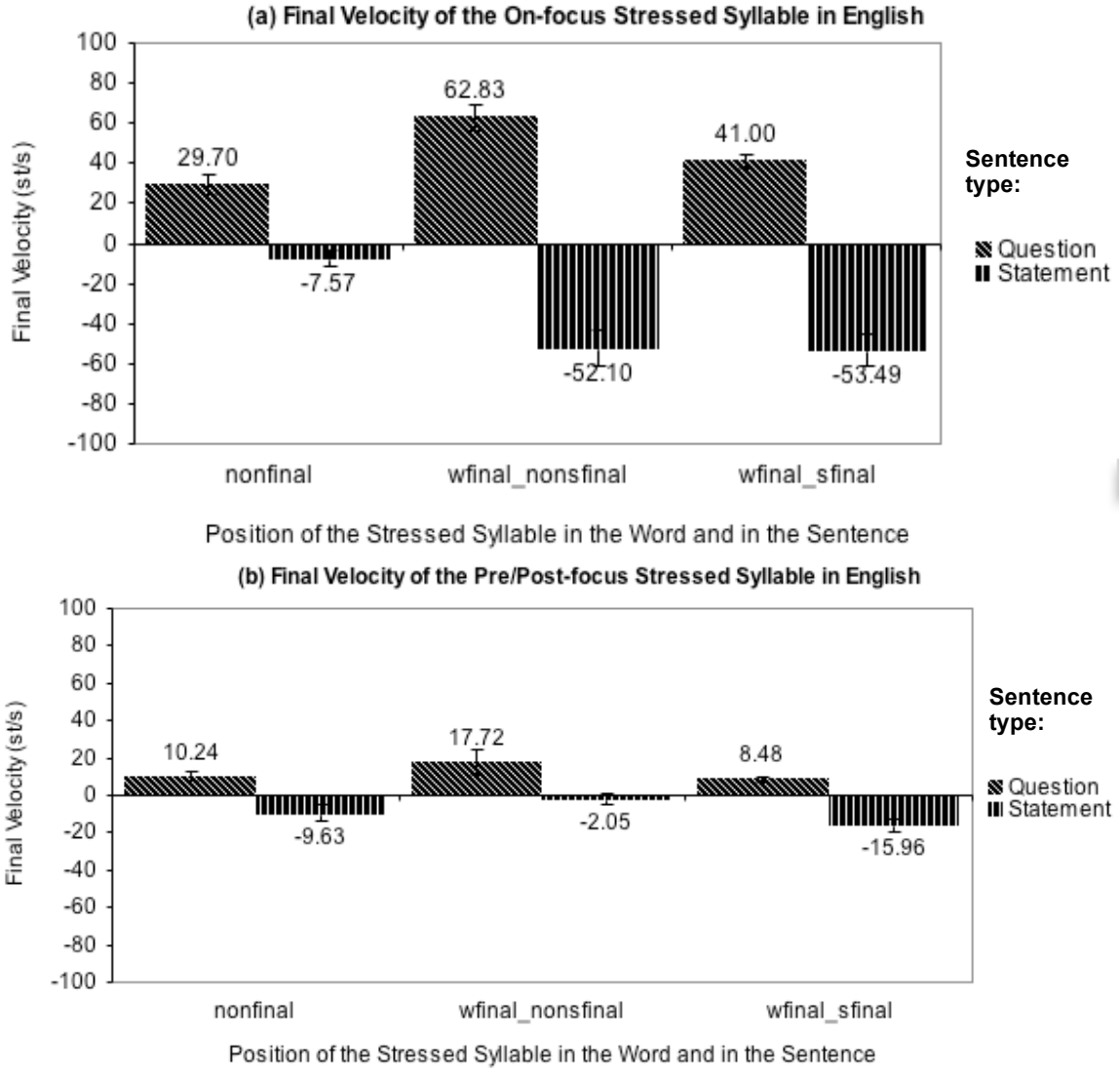

Figure 5.3. Final velocities of the stressed syllables in different positions (non-final/word-final but non-sentence-final/word-final and sentence-final) in English: (a) on-focus, and (b) pre-/post-focus. Data from Experiment 5.

Figure 5.4 is a quick reminder of the difference in statement/question intonation between Mandarin Chinese and General American English, where pitch contours of a pair of sentences are displayed in the two languages. The left panel shows an English statement vs. question with focus on *unmarriable*, and the right panel shows a Mandarin statement vs. question with focus on the Rising-tone syllable *yé*, whose following

syllables all bear the neutral tone. As can be seen, the on-focus stressed syllable *ma* in *unmarriable* seems to have a [high] target in the statement (because F_0 starts to rise from syllable onset and throughout most of the syllable, but starts to fall shortly before syllable offset), but a [rise] target in the question (because F_0 starts to drop at syllable onset, but changes direction in mid syllable, and rises sharply in the later half of the syllable); the post-focus content word *me* seems to have a [fall] target in the statement, but a [rise] target in the question. In Mandarin, however, pitch targets of the full tones (*tā mǎi yé*) and the neutral tone (*ye men de le ma*) are unchanged across different sentence types. Furthermore, focus modifies pitch range differently over the post-focus region in the two languages. In Mandarin, post-focus pitch range is lowered as well as compressed in both statements and questions, although the lowering is of smaller magnitude in questions. In English, however, post-focus pitch range is compressed and lowered in statements, just as in Mandarin, but compressed and *raised* in questions. Finally, the F_0 trajectory of the first post-focus unstressed syllable (*rri* as in *unmarriable*) in the English question resembles that of the first post-focus neutral tone (*ye* as in *yéye*) in the Mandarin sentences, with them all showing a rising trend that seems to be caused by the inertia from the [rise] target of the focused syllable.

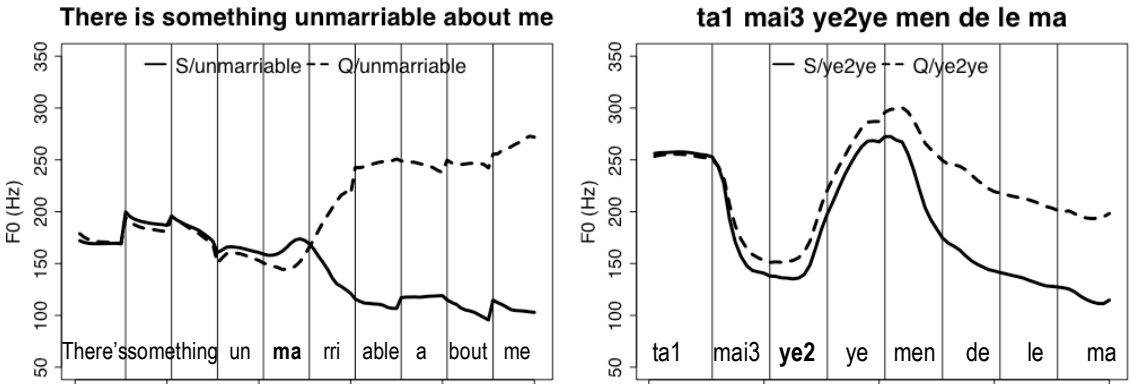

Figure 5.4. An illustrative comparison between statement/question intonation in English (Experiment 5) and Mandarin (Experiment 4).

In regard to the effects of focus and sentence type on syllable duration in the two languages, there seem to be both similarities and differences. In both languages, focus appears to lengthen the duration of the focused word as a whole, i.e., not only the on-focus stressed/full-tone syllable but also the post-focus unstressed/neutral-tone syllable(s) within the focused word are lengthened. In Mandarin, the sentence-final syllable tends to be longer in questions than in statements, whereas syllables in the earlier part of the sentences tend to be shorter in questions than in statements (which are all medial-focused). In English, however, there seems to be no difference in syllable duration between medial/final-focused statements and questions.

5.2 Further Discussion

The above analyses and interpretations are done in the framework of PENTA. Alternative interpretations are of course also possible and have been offered before. For Mandarin, the closest alternative interpretation is X.-N. S. Shen's (1990: 48) claim about the F_0 of the neutral tone being "the algebraic sum of the F_0 value of the preceding tone

and the sentence intonation". The present data in Experiment 4 indeed suggest that the F_0 of the neutral tone is determined by both the preceding tone and sentence type. But the resulting F_0 contour is unlikely a product of an algebraic summation. This is because articulation is not an algebraic operation, but a nonlinear dynamic process more akin to the simulation shown in Figure 4.2b of Chapter 4. Thus the influence of the preceding tone is imposed in the form of the final height as well as velocity of its F_0 contour; the effect of this influence depends also on the property of the current pitch target, including its height, slope and strength. The neutral tone may show a greater amount of influence from the preceding tone than a full tone due to its weak articulatory strength (Chen & Xu, 2006). If this weak strength account is valid, it may also account for the continuous gradual fall in the neutral-tone-ending questions in Experiment 4, which is absent in the questions consisting of all High tones in Experiment 1. Xu & Xu (2005) argued for English that the weak strength in an unstressed syllable makes it less effective in implementing post-focus F_0 lowering required by the focus function. Similarly, instead of an immediate post-focus lowering in a High tone sequence even in a question as shown in Experiment 1, the post-focus lowering occurs rather gradually over the course of a sequence of neutral tone syllables in Experiment 4.

In the case of English, one alternative theoretical account of the present data in Experiment 5 can be found in the AM theory of English intonation in the form of the upstep rule, echo accent and boundary tone (Pierrehumbert, 1980). First, regarding the upstep rule, as discussed in section 4.1.2 of Chapter 4, in AM the question intonation is transcribed as L* H-H%. To account for the final rising contour in a question, however,

Pierrehumbert (1980) proposed an upstep rule, which raises the F_0 of the H% boundary tone after the H- phrase accent in an intonational phrase. A schematic illustration of the upstep rule is shown in the lower right plot in Figure 5.5a, adapted from Pierrehumbert & Hirschberg (1990: 281). Comparing Figure 5.5a with Figure 5.5b we can see that the upstep pattern corresponds quite well to questions with both non-final and final focus, which means that the upstep rule seems to describe the surface F_0 with some accuracy. Note, however, the theoretical interpretations of the two accounts (AM and PENTA) are quite different. According to AM, the final rise is due to the H- phrase accent to which is attached a non-phonological (hence non-functional) phonetic rule that raises the F_0 of the following boundary tone, whereas in the perspective of PENTA the final rise is only part of the post-focus pitch range raising that encodes the question function, though it seems to be the most prominent part. In fact even the entire post-focus pitch range raising is seen in PENTA as only part of the encoding scheme of question intonation, because at least the target shift from [high] or [fall] to [rise] in the stressed syllables is also another important part of the encoding scheme.

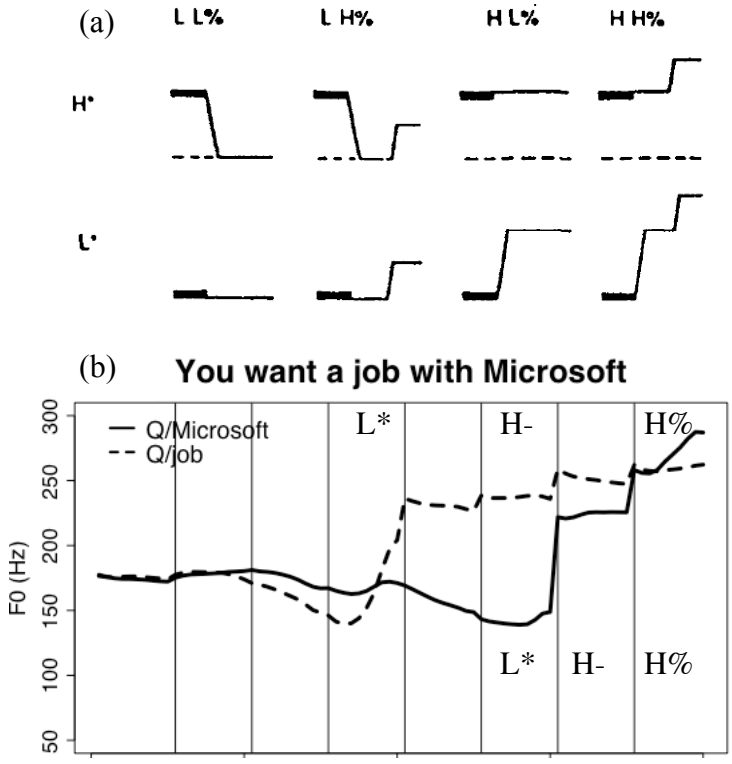

Figure 5.5. (a) A schematic illustration of the upstep rule, adapted from Pierrehumbert & Hirschberg (1990: 281). (b) F_0 contours of a pair of English sentences from the present data (Experiment 5).

Now consider the issue of echo accent. In the most widely adopted version of the AM theory and in ToBI, F_0 of syllables after the nuclear accent is assumed to be fully deaccented, i.e., no pitch events occur except the phrase accent. Pierrehumbert (1980: 223), however, described what she called echo accents: "Accentable syllables past the nuclear accent often carry a miniature replica of the nuclear accent. That is, in H* L- contours, one may see small peaks on accentable syllables following the H* nuclear accent; in L* H- contours, one [may] see small dips." Pierrehumbert (1980: 224) further suggests that "it would be appropriate to handle the phenomenon by a rule copying tones from the nucleus to sufficiently strong syllables on the right." For unknown reasons the echo accents are not included in the later presentation of the Pierrehumbert model or as

part of ToBI. Xu and Xu (2005) have shown that in English post-focus pitch range suppression in a statement does not eliminate local pitch targets, and that as a result, local F_0 contours in the post-focus region are similar to pre-focus region, and are only smaller in magnitude than in the on-focus region. The results of Experiment 5 in the present study show that the same is true for question intonation. What is even more clearly seen in the present data is that the type of pitch target of the stressed syllable in English corresponds to sentence type, and is thus functionally determined. Thus again from the perspective of PENTA there is unlikely a process of post-focus syllables copying the pitch target of the focused syllable, because all the pitch targets of a sentence are jointly determined by the lexical function and the sentence type function.

Finally, in regard to boundary tone and phrase accent, AM treats them as self-contained phonological units, which are independent of other intonational components. What the current data show is that the F_0 contours supposedly accounted for by boundary tone and phrase accent are more likely to be subcomponents of independent communicative functions. In relation to the boundary tone, the raising of the sentence-final F_0 is only part of the continuous pitch range increase starting from the focused word, and this is true of both English and Mandarin. For English, additionally, the question function also changes the pitch targets of the stressed syllables of all pre-, on- and post-focus content words. Thus from the PENTA perspective, for both languages, the final pitch raising is only part of the encoding scheme of the question function, albeit a very prominent part. As for the phrase accent, which has been problematic ever since Pierrehumbert (1980) because it is used to transcribe a stretch of nearly flat F_0 plateau of

varying lengths, there have been attempts within the AM theory to use a secondary association to account for the temporal domain of the plateau (Barnes et al., 2006; Grice et al., 2000; Pierrehumbert & Beckman, 1988). Again from the PENTA perspective, neither phrase accents nor their primary and secondary associations are needed because the plateaus, whether low (in statements) or high (in questions) are directly due to post-focus pitch range compression, and so are part of the encoding scheme of focus rather than independent units. And, as seen in both the present data (Experiment 5) and those of Xu & Xu (2005), in English at least, post-focus pitch range compression does not eliminate local pitch targets related to lexical stress. Thus, phrase accents cannot even fully account for the detailed F_0 contours in the post-focus regions.

Therefore, in regard to all three issues, the difference between the alternative accounts is that in the AM theory phonological components and their implementation rules are self-sustaining, and they together directly specify the surface F_0 contours, while in PENTA the encoding schemes are functionally defined and their links to surface F_0 contours are through the TA process rather than being direct.

In this dissertation, PENTA is used as a theoretical framework to guide experimental design, data analysis and interpretation of the results. It is also possible to use PENTA to generate functionally appropriate F_0 contours with data obtained in an empirical study. This has recently been done with the development of the qTA model—a quantitative implementation of PENTA (Promon, Xu & Thipakorn, in press). F_0 contours of tone in Mandarin, lexical stress in English and focus in both languages have been simulated though automatic analysis-by-synthesis using experimental data from Xu

(1999) and Xu & Xu (2005). Data obtained in the present study can therefore be used in future research to generate question intonation in Mandarin and English through analysis by synthesis using qTA.

5.3 Conclusions

In summary, by comparing question intonation in both Mandarin and English while controlling for focus, lexical tone and lexical stress, I have found similarities as well as differences between the two languages. In both languages, the question function raises the pitch range of all words starting from the location of focus, but the greatest increase occurs at the end of the sentence. In Mandarin, however, this pitch range raising does not change the basic focus pattern of on-focus pitch range expansion and post-focus pitch range lowering and compression. In English, in contrast, the post-focus pitch range is not only compressed, but also raised above the pitch of the focused word and stays high till the end of the sentence. Finally, in Mandarin, the question function does not seem to change the pitch targets of individual syllables associated with the lexical tones, and this is true of both full tones and the neutral tone. In English, in contrast, the question function seems to change the pitch targets of the stressed syllables of all content words from [high] or [fall] to [rise]. As a result, the English question/statement contrast often starts earlier than Mandarin in a sentence (unless when the focus is sentence initial, in which cases the contrast does start from the beginning of the sentence in Mandarin).

These findings were interpreted in terms of the theoretical framework of PENTA (Xu, 2005), which views communicative functions as the driving force of intonation and the articulatory process of target approximation as the encoding mechanism. It is further

assumed in PENTA that individual communicative functions have to be conveyed through encoding schemes that are defined in terms of the TA parameters they specify and that these encoding schemes are language dependent and thus have to be empirically discovered. Thus the differences between Mandarin and English summarized above are in terms of the specifics of their respective encoding schemes for a similar set of contrastive communicative functions.

Putting the PENTA model in the context of alternative interpretations (the AM theory of English intonation in particular), I have found theoretical similarities and differences between the two accounts. Specifically, the most basic difference between the two is that the AM account treats intonational components and their operational rules as self-sustaining while PENTA views intonation as driven by communicative functions that are encoded by an articulatory process. Future research into other communicative functions may help to further clarify the differences between the theoretical approaches.

6 CONCLUSIONS

Much controversy in the literature of speech prosody results from lack of recognition of various communicative functions. As an effort to achieve a better understanding of intonation systems of Mandarin Chinese and General American English, this dissertation explores the contributions of lexical tone/stress, focus, and sentence type in declarative and interrogative sentences in the two languages, with an approach that prioritizes the functional aspect of speech (Hirst, 2004; Kohler, 2005; Xu, 2005, 2006).

In Experiment 1, native speakers of Mandarin produced statements, yes/no questions, particle questions, wh-questions, rhetorical questions and confirmation questions with narrow focus on the initial, medial or final word of the sentence, or on none of the words. Detailed F_0 contour analyses showed that focus generated the same pitch range modification in questions as in statements, i.e., expanding the pitch range of the focused word, suppressing (compressing and lowering) that of the post-focus words, but leaving that of the pre-focus words largely unaffected. When the effects of focus (as well as other functions also potentially present) was controlled by subtracting statement F_0 contours from those of corresponding yes/no questions, the resulting difference curves resembled exponential or even double-exponential functions. Further F_0 analyses also revealed an interaction between focus and interrogative meaning in the form of a boost to the pitch raising by question starting from the focused word. Finally, subtle differences in the amount of pitch raising were also observed among different types of questions, especially at the sentence-final position.

Experiment 2 investigated whether listeners could detect both focus and sentence type in the same utterance. Results showed that listeners could identify both in most cases, indicating that F_0 variations related to the two functions could be simultaneously transmitted. Meanwhile, the lowest identification rates were found for neutral focus in questions and for statements with final focus. In both cases the confusions seemed to arise from the competing F_0 adjustments by interrogative meaning and focus at the sentence-final position.

In Experiment 3, decision trees with three different sets of feature vectors were implemented to determine the most significant elements in an utterance that signify its sentence type (statement vs. yes/no question in Mandarin). For the 10-syllable utterances, the highest correct classification rate (85%) was achieved when normalized (to remove the effects of speaker, tone, and focus) final F_0's of the 7th and the last syllables were included in the tree construction. This performance was close to human performance (89%) for the same testing set in Experiment 2. The results confirmed the finding in Experiment 1 that the difference between statement and question intonation in Mandarin was manifested by an increasing departure from a common starting point toward the end of the sentence. Furthermore, Experiment 3 demonstrated that features from direct surface curve fitting worked no better than tonal features extracted from final F_0 of each syllable, which was in turn worse than normalized syllable-final F_0's. This indicates that tonal targets and the speaker and focus effects should be accounted for in intonation modeling.

Experiments 4 and 5 examined question intonation in Mandarin and English, two languages known to be very different in their use of pitch at the lexical level. Speakers of both languages produced statements and yes/no questions with alternating lexical tone/stress and focus conditions. F_0 and duration patterns were analyzed in terms of how they jointly encode these communicative functions in light of the Parallel Encoding and Target Approximation (PENTA) model. Differences as well as similarities are found in the way the two languages simultaneously encode lexical, focal and sentential information: 1) In both languages F_0 of questions becomes increasingly higher than that of statements over the course of the sentence, starting from the focused word, although the magnitude of the raise is much bigger in English. 2) In both languages, on-focus pitch range is expanded. Post-focus pitch range is compressed and lowered in Mandarin in both statements and questions, although the latter is smaller in magnitude. In English, however, post-focus pitch range is compressed and lowered in statements but compressed and raised in questions. 3) In Mandarin, tonal targets remain unchanged regardless of focus or sentence type. In English, in contrast, the underlying pitch target of the stressed syllable in a content word changes from [high] or [fall] in statements to [rise] in questions. As a result, the sentence type contrast becomes salient starting from the stressed syllable of the first content word, whether or not it is focused. The comparison between Mandarin and English highlights the importance of understanding intonation based on specific communicative functions rather than on surface acoustic forms only.

Regarding the pitch target shift of stressed syllables in English content words, the current finding is in agreement with what was described in O'Shaughnessy and Allen

(1983: 1156), where the authors identified two forms ([high]: "an upward obtrusion of F_0", and [fall]: "a steep fall") of "primary emphasis" and one form ([high]: "relatively level F_0") of "secondary emphasis" in English statements, but noticed that "emphasis was mostly marked by rises" in English yes/no questions (p. 1167). Similarly, Grice (1995) noted that in English, pitch accents with leading tones (L+H* and H+L*) are tone clusters (meaning that the leading tones L+ and H+ belong to pre-nuclear accents and that the two pitch accents are equivalent to H* and L*), whereas those with trailing tones (L*+H and H*+L) are contour tones (meaning that the two pitch accents are actually [rise] and [fall]). Therefore, the six pitch accent types (H*, L*, L+H*, L*+H, H+L*, H*+L) in Pierrehumbert's system (Beckman & Pierrehumbert, 1986) are actually surface realizations of three functionally motivated local pitch targets in PENTA terms: [high] (which include H*, L+H*, and H+L*), [fall] (H*+L), and [rise] (which include both L* and L*+H). Furthermore, according to the current finding, the pitch target shift in the stressed syllable (denoted by different pitch accents in Pierrehumbert's framework) in the content word is jointly determined by the stress pattern of the word, the word's position in sentence, focus structure and sentence type. By ignoring intonational meanings when deriving the phonological system, Pierrehumbert (1980) did not find the rule that governs the occurrences of different pitch accents. Finally, as found in Experiment 5, the interaction between the functional components, namely, lexical stress, focus, and sentence type largely determines the detailed F_0 contours in English statements and yes/no questions that the AM theory has tried to describe in terms of pitch accents, phrase accents, and boundary tones. This means that, from a functional perspective, these non-

functionally defined units are virtually redundant, as they are not indispensable when it comes to generating surface contours, and their links to communicative meanings have to be subsequently established (as attempted by Pierrehumbert & Hirschberg, 1990).

Furthermore, the Pierrehumbert system only recognizes local pitch accents and edge tones as phonological units, while denying any phonological role of what they referred to as pitch scaling. The acoustic analysis in Experiment 1 showed that focus and sentence type in Mandarin are realized through pitch range variations without changing the local tonal targets. But the results of Experiment 2 showed that the recognition rates of both focus and sentence type were very high. This means that the pitch range variations due to focus and sentence type are highly contrastive, which is consistent with the contrastiveness requirement for phonological components. Also as mentioned in section 1.1 of Chapter 1, the components of the AM theory of intonational phonology that have high inter-transcriber consistency are those related to focus and sentence type, namely, presence/absence of pitch accents and the type of boundary tones. In other words, the functionally defined pitch range variations are actually far more contrastive than the visually defined accent types, and are thus more like genuine phonological categories.

In conclusion, the overall findings of this dissertation are consistent with the functional view of intonation (Xu, 2005), according to which components of intonation are defined and organized by individual communicative functions that are independent of each other but are encoded in parallel. These findings not only improve our

understanding of speech intonation in general, but also have significant implications for speech technology and second language acquisition.

REFERENCES

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, **36,** 3-25.

Atterer, M. & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F_0: Evidence from German. *Journal of Phonetics*, **32,** 177-197.

Barnes, J., Shattuck-Hufnagel, S.; Brugos, A. & Veilleux, N. (2006). The domain of realization of the L- phrase tone in American English. *Proceedings of Speech Prosody 2006*, Dresden.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, **3**, 255-309.

Boersma, P., & Weenink, D. (2008). Praat: doing phonetics by computer (Version 5.0.08) (Computer program). Retrieved February 11, 2008, from http://www.praat.org/.

Bolinger, D. (1978). Intonation across languages. In J. H. Greenberg (ed.), *Universals of human language, Phonology V. 2*, pp. 471-523. Stanford University Press.

Bolinger, D. (1986). *Intonation and its parts: melody in spoken English*. Stanford University Press, Palo Alto.

Botinis, A., & Bannert, R. (1997). Tonal perception of focus in Greek and Swedish. *Proceedings of an ESCA Workshop — Intonation: Theory, Models and Applications,* pp. 47-50. Athens, Greece.

Botinis, A., Bannert, R., & Tatham, M. (2000). Contrastive tonal analysis of focus perception in Greek and Swedish. In A. Botinis (ed.), *Intonation: analysis, modelling and technology*, pp. 97-116. Kluwer Academic Publishers, Boston.

Botinis, A., Fourakis, M., & Gawronska, B. (1999). Focus identification in English, Greek and Swedish. *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 1557-1560. San Francisco.

Chang, N. T. (1958). Tones and intonation in the Chengtu dialect (Szechuan, China). *Phonetica*, **2**, 59-85.

Chao, Y. R. (1930). A system of "tone letters". *Le Maître Phonétique*, **45,** 24-27.

Chao, Y. R. (1932). A preliminary study of English intonation (with American variants) and its Chinese equivalents. *Bulletin of the National Research Institute of History and Philology of the Academia Sinica*, **Supplementary volume I,** 105-156.

Chao, Y. R. (1933). Tone and Intonation in Chinese. *Bulletin of the Institute of History and Philology*, **4**, 121-134.

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

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech -- Evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica*, **63,** 47-75.

Clark, L. A., & Pregibon, D. (1992). Tree-based models. In J. M. Chambers, & T. J. Hastie (eds.), *Statistical Models in S*, pp. 377-419. Wadsworth & Brooks.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts, *Journal of the Acoustical Society of America*, **77**, 2142-2156.

Dainora, A. (2001). *An empirically based probabilistic model of intonation in American English*. Ph.D. dissertation, University of Chicago.

Dainora, A. (2002). Does intonational meaning come from tones or tunes? Evidence against a compositional approach. In *Proceedings of the 1st International Conference on Speech Prosody*, pp. 235–238. Aix-en-Provence, France.

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America*, **80,** 402-416.

Fok-Chan, Y. Y. (1974). *A perceptual study of tones in Cantonese.* University of Hong Kong Press, Hong Kong.

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

Gandour, J., Potisuk, S., & Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics*, **22,** 477-492.

Gårding, E., & Abramson, A. S. (1965). A study of the perception of some American English intonation contours. *Studia Linguistica*, **19**, 61-79.

Gauthier, B., Shi, R., & Xu, Y. (2007). Learning phonetic categories by tracking movements. *Cognition*, **103,** 80-106.

Goldsmith, J. (1981). English as a tone language. In D. Goyvaerts (Ed.), *Phonology in the 1980's* (pp. 287-308). Ghent: Story-Scientia. Circulated in 1974.

Grice, M. (1995). Leading tones and downstep in English. *Phonology,* **12**, 183-233.

Grice, M., Ladd, D. R., & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology*, **17**, 143-185.

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

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

Haan, J. (2002). *Speaking of Questions: An Exploration of Dutch Question Intonation*, Ph. D. dissertation, University of Nijmegen.

Hastie, T. (1992). Generalized additive models. In J. M. Chambers, & T. J. Hastie (eds.), *Statistical Models in S*, pp. 249-307. Wadsworth & Brooks.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Hedberg, N., Sosa, J. M., & Fadden, L. (2004). Meanings and Configurations of Questions in English. In *Proceedings of Speech Prosody 2004*, pp. 309-312. Nara, Japan.

Herman, R. (1996). Final lowering in Kipare. *Phonology*, **13**, 171-196.

Hirschberg, J. (2004). Pragmatics and intonation. In L.R. Horn, & G.L. Ward (eds.), *The Handbook of Pragmatics*, pp. 515-537. Oxford: Blackwell.

Hirschberg, J., & Ward, G. (1992). The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics,* **20**, 241-251.

Hirschberg, J., (2000). A corpus-based approach to the study of speaking style. In M. Horne (ed.), *Prosody: theory and experiment. Studies presented to Gösta Bruce*, pp. 335-350. Kluwer Academic Publishers.

Hirst, D. J. (2005). Form and function in the representation of speech prosody. *Speech Communication*, **46**, 334-347.

Ho, A. T. (1976a). Mandarin tones in relation to sentence intonation and grammatical structure. *Journal of Chinese Linguistics*, **4,** 1-13.

Ho, A. T. (1976b). The acoustic variation of Mandarin tones. *Phonetica*, **33,** 353-367.

Ho, A. T. (1977). Intonation variation in a Mandarin sentence for three expressions: interrogative, exclamatory and declarative. *Phonetica*, **34,** 446-457.

Honorof, D. N. & Whalen, D. H. (2005). Perception of pitch location within a speaker's F_0 range. *Journal of the Acoustical Society of America*, **117,** 2193-2200.

Hu, F. (2002). A prosodic analysis of wh-words in Standard Chinese. In *Proceedings of the 1st International Conference on Speech Prosody*, pp. 403-406. Aix-en-Provence, France.

Inkelas, S., & Leben, W. R. 1990. Where phonology and phonetics intersect: the case of Hausa intonation. In J. Kingston, & M. E. Beckman (Eds.), *Papers in Laboratory Phonology 1 — Between the Grammar and Physics of Speech* (pp. 17-34). Cambridge: Cambridge University Press.

Ishihara, S. (2002). Syntax-Phonology Interface of Wh-Constructions in Japanese. In *Proceedings of the Third Tokyo Conference on Psycholinguistics*, pp. 165-189. Tokyo.

Jin, S. (1996). *An Acoustic Study of Sentence Stress in Mandarin Chinese.* Ph.D. dissertation, The Ohio State University.

Jun, S.-A., & Oh, M. (1996). A Prosodic analysis of three types of wh-phrases in Korean. *Language and Speech*, ***39***, 37-61.

Kohler, K. J. (2004). Prosody revisited: function, time, and the listener in intonational phonology. In B. Bel, & I. Marlien (eds.), *Proceedings of International Conference on Speech Prosody 2004*, pp. 171-174. Nara, Japan.

Kohler, K. J. (2005). Timing and communicative functions of pitch contours. *Phonetica,* **62**, 88-105.

Ladd, D. R. (1983). Phonological features of intonational peaks. *Language,* **59**, 721-759.

Ladd, D. R. (1987). Review of Bolinger 1986. *Language*, **63**, 637-43.

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

Ladd, D. R., Mennen, I., & Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America*, **107,** 2685-2696.

Laniran, Y. O. & Clements, G. N. (2003). Downstep and high raising: interacting factors in Yoruba tone production. *Journal of Phonetics*, **31,** 203-250.

Lee, O. J. (2005). *The prosody of questions in Beijing Mandarin*. Ph.D. dissertation, Ohio State University.

Lee, W.-S. (2004). The Effect of Intonation on the Citation Tones in Cantonese. In *Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, pp. 107-110. Beijing.

Li, C. N., & Thompson, S. A. (1979). The pragmatics of two types of yes-no questions in Mandarin and its universal implications, *Papers from the Fifteenth Regional Meeting of the Chicago Linguistic Society*, pp. 197-206.

Lin, M. (2004). On production and perception of boundary tone in Chinese intonation. In *Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, pp. 125-129. Beijing.

Lindau, M. (1986). Testing a model of intonation in a tone language. *Journal of the Acoustical Society of America*, **80**, 757-764.

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

Liu, F., & Xu, Y. (2007a). Question intonation as affected by word stress and focus in English. In *Proceedings of The 16th International Congress of Phonetic Sciences*, pp. 1189-1192, Saarbrücken.

Liu, F., & Xu, Y. (2007b). The neutral tone in question intonation in Mandarin. In *Proceedings of Interspeech 2007*, pp. 630-633. Antwerp.

Liu, F., Surendran, D., & Xu, Y. (2006). Classification of statement and question intonations in Mandarin. *Proceedings of Speech Prosody 2006*. Dresden, Germany.

Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2004). The effects of intonation patterns on lexical tone production in Cantonese. In *Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, pp. 133-136. Beijing.

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**, 159-174.

Mixdorff, H. (2002). Speech technology, ToBI and making sense of prosody. In *Proceedings of Speech Prosody 2002*, pp. 31-38. Aix-en-Provence, France.

Mixdorff, H. (2004). Quantitative tone and intonation modeling across languages. In *Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, pp. 137-142, Beijing.

Ni, J.-F., & Kawai, H. (2004). Pitch targets anchor Chinese tone and intonation patterns. In *Proceedings of International Conference on Speech Prosody 2004*, pp. 95-98. Nara, Japan.

O'Connor, J. D., & Arnold, G. F. (1961). *Intonation of colloquial English*. London: Longmans.

O'Shaughnessy, D. (1979). Linguistic features in fundamental frequency patterns. *Journal of Phonetics*, **7,** 119-145.

O'Shaughnessy, D., & Allen, J. (1983). Linguistic modality effects on fundamental frequency in speech. *Journal of the Acoustical Society of America*, **74**, 1155-71.

Palmer, H. E. (1922). *English intonation, with systematic exercises*. Cambridge: Heffer.

Pell, M. D. (2001). Influence of emotion and focus on prosody in matched statements and questions. *Journal of the Acoustical Society of America*, **109,** 1668-1680.

Peng, S.-H., Chan, M. K.M., Tseng, C.-Y., Huang, T., Lee, O. J., & Beckman, M. E. (2005). Towards a Pan-Mandarin system for prosodic transcription. In S.-A. Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing,* pp. 230-270. Oxford: Oxford University Press.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, MIT, Cambridge, MA. [Published in 1987 by Indiana University Linguistics Club, Bloomington].

Pierrehumbert, J. (1981). Synthesizing intonation. *Journal of the Acoustical Society of America*, **70,** 985-995.

Pierrehumbert, J. (2000). Tonal elements and their alignment. In M. Horne (ed.), *Prosody: Theory and Experiment. Studies Presented to Gosta Bruce,* pp. 11-26. Kluwer, Dordrecht.

Pierrehumbert, J., & Beckman, M. (1988). *Japanese Tone Structure*. The MIT Press, Cambridge, MA.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (eds.), *Intentions in Communication*, pp. 271-311. Cambridge, Massachusetts: MIT Press.

Pierrehumbert, J., & Steele, S. (1989). Categories of tonal alignment in English. *Phonetica* 46:181–196.

Prieto, P. & Torreira, F. (2007). The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics*, **35,** 473-500.

Prom-on, S., Xu, Y., & Thipakorn, B. (in press). Modeling tone and intonation in Mandarin and English as a process of target approximation. To appear in *Journal of the Acoustical Society of America*.

Qi, S. (1956). Hanyu de zidiao, tingdun yu yudiao de jiaohu guanxi [The interaction among lexical tones, pause, and intonation in Chinese]. *Zhongguo Yuwen [Journal of Chinese Linguistics]*, **1956(10)**, 10-13.

R Development Core Team. (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rumjancev, M. K. (1972). *Ton i intonacija v sovremennom kitajskom jazyke [Tone and Intonation in Modern Chinese]*. Izdatel'stvo Moskovskogo Universiteta, Moscow. Reviewed by Lyovin, A. (1978). *Journal of Chinese Linguistics,* **6**, 120-168.

Rump, H. H., & Collier, R. (1996). Focus conditions and the prominence of pitch-accented syllables. *Language and Speech*, ***39***, 1-17

Saussure, F. de. (1998). *Course in General Linguistics*. Bally, C., & Sechehaye, A. (eds.), Harris, R. (trans.). Duckworth.

Selkirk, E., & Shen, T. (1990). Prosodic domains in Shanghai Chinese. In S. Inkelas, & D. Zec (eds.), *The Phonology-Syntax Connection,* pp. 313-37. Chicago, IL: University of Chicago Press.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research,* **25(2)**, 193-247.

Shen, J. (1985). Beijinghua shengdiao de yinyu he yudiao [Pitch range of tone and intonation in Beijing dialect]. In T. Lin, & L. Wang (eds.), *Beijing Yuyin Shiyanlu [Working Papers in Experimental Phonetics],* pp. 73-130. Beijing: Beijing University Press.

Shen, J. (1992). Hanyu yudiao moxing chuyi [Preliminary discussion of intonational models in Mandarin]. *Yuyan yanjiu*, **1992.4**, 16-24.

Shen, J. (1994). Hanyu yudiao gouzao he yudiao leixing [Intonational structures and patterns in Mandarin]. *Fangyan*, **1994.3**, 221-228.

Shen, X.-N. S. (1990). *The Prosody of Mandarin Chinese*. Berkeley, CA: University of California Press.

Shriberg, E., Bates, R., Stolcke, A., Taylor, P., Jurafsky, D., Ries, K., Coccaro, N., Martin, R., Meteer, M., & Van Ess-Dykema, C. (1998). Can prosody aid the automatic

classification of dialog acts in conversational speech? *Language and Speech,* **41(3-4)**, 439-487.

Silverman, K., Beckman, M., Pierrehumbert, J., Ostendorf, M., Wightman, C., Price, P., & Hirschberg, J. (1992). ToBI: A standard scheme for labeling prosody. In *Proceedings of the 2nd International Conference of Spoken Language Processing*, pp. 867-869. Banff, Canada.

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry,* **31(4)**, 649-689.

Studdert-Kennedy, M. & Hadding, K. (1973). Auditory and linguistic processes in the perception of intonation contours. *Language and Speech,* **16**, 293-313.

Sullivan, L. H. (1896). The tall office building artistically considered. *Lippincott's Magazine*, March 1896.

Syrdal, A., & McGory, J. (2000). Inter-transcriber reliability of ToBI prosodic labeling. In *Proceedings of International Conference On Spoken Language Processing*, vol. 3, pp. 235-238. Beijing, China.

't Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual Study of Intonation — An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Taylor, P. (2000). Analysis and synthesis of intonation using the Tilt model. *Journal of the Acoustical Society of America*, **107,** 1697-1714.

Taylor, P., King, S., Isard, S., & Wright, H. (1998). Intonation and dialogue context as constraints for speech recognition. *Language and Speech,* **41(3-4)**, 493-512.

Thorsen, N. G. (1978). An acoustical investigation of Danish intonation. *Journal of Phonetics*, ***6***, 151-175.

Thorsen, N. G. (1979). Lexical stress, emphasis for contrast, and sentence intonation in advanced standard Copenhagen Danish. In *Proceedings of Ninth International Congress of Phonetic Sciences*, pp. 417-423. Institute of Phonetics, Copenhagen.

Thorsen, N. G. (1980). A study of the perception of sentence intonation - Evidence from Danish. *Journal of the Acoustical Society of America*, **67**, 1014-1030.

Torng, P.-C. (2000). *Supralaryngeal articulator movements and laryngeal control in Mandarin Chinese tonal production*. Ph.D. dissertation. University of Illinois at Urbana-Champaign.

Tsao, W.-Y. (1967). Question in Chinese. *Journal of Chinese Language Teachers' Association*, **2**, 15-26.

Vance, T. J. (1976). An experimental investigation of tone and intonation. *Phonetica*, ***33***, 368-392.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer.

Wang, A. (2003). *Putonghua Yudiao Yingao Jiangshi Yanjiu [Research on the Pitch Downtrend of Intonation in Putonghua]*. Ph.D. dissertation, Beijing University.

Ward, G., & Hirschberg, J. (1985). Implicating uncertainty. *Language*, **61**, 747-76.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F_0 of vowels. *Journal of Phonetics*, **23**, 349-366.

Wightman, C. (2002). ToBI or not ToBI. In *Preceedings of the International Conference on Speech Prosody 2002*, pp. 25-29. Aix-en-Provence, France.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F_0 contours. *Journal of Phonetics*, **27**, 55-105.

Xu, Y. (2004). Separation of functional components of tone and intonation from observed F_0 patterns. In G. Fant, H. Fujisaki, J. Cao, & Y. Xu (eds.), *From Traditional Phonology to Modern Speech Processing: Festschrift for Professor Wu Zongji's 95th Birthday*, pp. 483-505. Beijing: Foreign Language Teaching and Research Press.

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, **46,** 220-251.

Xu, Y. (2005-2008). _TimeNormalizeF0.praat. Available from: <http://www.phon.ucl.ac.uk/home/yi/tools.html>.

Xu, Y. (2006). Speech prosody as articulated communicative functions. In *Proceedings of Speech Prosody 2006*. Dresden, Germany.

Xu, Y., & Kim, J. (1996). Downstep, regressive upstep, H-raising, or what? -- Sorting out the phonetic mechanisms of a set of "phonological" phenomena. *Journal of the Acoustical Society of America,* **100(Pt 2)**, 2824.

Xu, Y., & Liu, F. (2006). Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics*, **18,** 125-159.

Xu, Y., & Liu, F. (2007). Determining the temporal interval of segments with the help of F0 contours. *Journal of Phonetics*, **35,** 398-420.

Xu, Y., & Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America*, **111,** 1399-1413.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, **33,** 319-337.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, **33,** 159-197.

Xu, Y., Xu, C. X., Sun, X. (2004). On the Temporal Domain of Focus. In *Proceedings of International Conference on Speech Prosody 2004*, pp. 81-84. Nara, Japan.

Yuan, J. & Jurafsky, D. (2005). Detection of Questions in Chinese Conversational Speech. Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop, pp. 47-52. Cancun, Mexico.

Yuan, J. (2004). *Intonation in Mandarin Chinese: Acoustics, Perception, and Computational Modeling*. Ph.D. dissertation, Cornell University.

Yuan, J., Shih, C., & Kochanski, G. P. (2002). Comparison of declarative and interrogative intonation in Chinese. In *Proceedings of the 1st International Conference on Speech Prosody*, pp. 711-714. Aix-en-Provence, France.

Zeng, X.-L., Martin, P., & Boulakia, G. (2004). Tones and intonation in declarative and interrogative sentences in Mandarin, In *Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, pp. 235-238. Beijing.

Zue, V. (2007). On Organic Interfaces. *Proceedings of Interspeech 2007,* pp. 1-8. Antwerp, Belgium.