Order Number 9419441

Contextual tonal variation in Mandarin Chinese

Xu, Yi, Ph.D.

The University of Connecticut, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

CONTEXTUAL TONAL VARIATION IN MANDARIN CHINESE

Yi Xu, Ph. D.

The University of Connecticut, 1993

Contextual tonal variation in Mandarin is examined through acoustic analyses and perceptual experiments in this study.

F0 analysis of trisyllabic words and phrases, as described in Chapter 1, finds that tonal contexts in which adjacent pitch values disagree across syllable boundaries may greatly change a tone from its canonical form, sometimes severely enough to even alter the direction of the tonal contour. It is also found that the F0 contour of a tone is affected more by the preceding tone than by the following tone. Perception of coarticulated tones is examined by removing relevant semantic information from these trisyllabic words and phrases through waveform editing and using the edited utterances as stimuli for tone identification. Mandarin speakers identify the tones presented in the original tonal contexts with high accuracy. Without the original context, however, correct identification drops below chance for tones that deviate much from the ideal contours due to coarticulation. When the original tonal context is altered, listeners compensate for the altered contexts as if they had been there originally. These results are interpreted as demonstrating listeners' ability to compensate for tonal coarticulation.

Chapter 2 examines F0 contours in disyllabic sequences with all the possible bitonal combinations of the four lexical tones in Mandarin produced in different carrier sentences. It is found that a tone is influenced by both carryover and anticipatory effects, but the former seems to dominate. The carryover effect is found to be assimilatory, and its

influence is seen to extend across the next and even the third syllables in a row. In contrast, the anticipatory effect is mostly dissimilatory, i.e., a low or high starting pitch raises or lowers the F0 values of the preceding tone, and its influence seems to be limited to the immediately preceding syllable.

Chapter 3 offers a unified view of contextual tonal variation in Mandarin that explains various non-phonological contextual tonal variations found so far in Mandarin in terms of interaction between anticipatory and carryover effects.

CONTEXTUAL TONAL VARIATION IN MANDARIN CHINESE

Yi Xu

B.A., Shandong University, 1982

M.A., Chinese Academy of Social Sciences, 1985

M.A., The University of Connecticut, 1990

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at

The University of Connecticut

1993

APPROVAL PAGE

Doctor of Philosophy Dissertation

CONTEXTUAL TONAL VARIATION IN MANDARIN CHINESE

Presented by

Yi Xu, B.A., M.A.

Major Adviser ______________________ Ignatius G. Mattingly

Associate Adviser ______________________ Arthur S. Abramson

Associate Adviser ______________________ David Michaels

Outside Reader ______________________ Chilin Shih

The University of Connecticut

1993

Dedication

To my fellow countrymen in China, who are striving for freedom and human rights.

Acknowledgment

This research was supported in part by NIH grant NICHD-HD-01994 to Haskins Laboratories while I was a graduate assistant at Haskins.

I would like to express my thanks to the members of my committee, Ignatius Mattingly, Arthur Abramson and David Michaels for their guidance, criticism, and counsel. Ignatius's constant request for theoretical depth and technical refinement helped to form this thesis to its present shape (Of course, I am responsible for any potential errors still in the thesis). At the same time, his patience, his readiness to help, and his kindness made my student experience at UConn an extremely pleasant one. I benefited tremendously from Arthur's expertise on tonal studies. Also it was Arthur who first suggested Experiment 3 described in Chapter 1. David's phonological insights were especially valuable, which often helped me to take a different angle, and made me more cautious about the claims I was making. I should also mention my debt to Chilin Shih who served as the outside reader of my thesis and provided many helpful comments.

I am indebted to many people at Haskins. Al Liberman gave me tremendous encouragement and stimulation for my research. Doug Whalen first showed me the advantages of "manual" measurements of F0, and how that can be done within practical limits. I also benefited from countless discussions with Doug Whalen during my years at Haskins. Carol Fowler, Cathy Best, Bruno Repp, and Freddie Bell-berti read an earlier version of Chapter 1 and provided very helpful comments. Also, technical support by Don Hailey, Vince Gulisano, Michael D'Angelo, and Ed Wiley is gratefully acknowledged.

I am also very grateful to my Chinese subjects. Without their cooperation, this study could not have been possible.

Finally, to my wife, Xihong, I express my thanks for her support, love, and understanding.

Table of Contents

List of Tables

List of Figures

Introduction

The aim of this thesis is to find answers to some of the questions concerning the nature of variation in Mandarin tones caused by tonal contexts in particular, and to further our understanding of coarticulation, perception, and intonation in general.

Mandarin, as a tone language, uses tones to differentiate words and morphemes that are otherwise identical in their phonological structures. Along with most of the other East Asian languages, Mandarin is said to have a contour tone system as opposed to the register tone systems found in many African languages and some of the American Indian languages (Pike, 1948). In Mandarin, a monosyllabic word or morpheme is specified phonologically not only by its consonants and vowels, but also by a tone, which is manifested mainly in terms of the rate of vocal fold vibration during the voiced portion of the syllable.[1]

Citation form of Mandarin Tones

Over the years, many studies have been done on the citation forms of Mandarin tones (e.g., Bai, 1934; Chao, 1948, 1968; Chuang, Hiki, Sone, & Nimura, 1971; Howie, 1970, 1974; M. Lin, 1965; Liu, 1924; Shih, 1986; Tseng, 1981).

According to those studies, there are four stressed lexical tones and one neutral tone in Mandarin. The four stressed tones have the pitch[2] contours high-level, mid-rising, falling-rising, and high-falling when the syllable carrying them is produced in isolation.

---

[1] Howie (1974) argues that the domain of tones in Mandarin does not include any initial voiced consonants or a non-syllabic vowel, because they "are merely anticipatory adjustments of the voice." This way of defining the domain of tones, however, may imply that the voiced consonants and non-syllabic vowels provide no information about the tone of a syllable, which is yet to be proved.

[2] Here and throughout this thesis, the term "pitch" refers to either underlying or perceived tonal value rather than actual fundamental frequency.

Table 0.1 lists the four stressed lexical tones in Mandarin and the different names and descriptions for them.

Table 0.1. The four stressed tones in Mandarin

Pinyin spelling	Tone name	Chinese name	Name used in this thesis	Pitch value	Tone letter by Chao (1930)	Symbol used in this thesis	Pitch targets in this thesis
mā	Tone 1	yin̄pinǵ[3]	high	55 or 44	˥ or ˦	¯	high-high
má	Tone 2	yanǵpinǵ	rising	35	˧˥	/	low-high
mǎ	Tone 3	shanǧshenḡ	low	214	˨˩˦	_	low-low
mà	Tone 4	qùshenḡ	falling	51	˥˩	\	high-low

The neutral tone in Mandarin occurs only on suffixes. It never occurs in isolation or on the first syllable of a word. There is no definite pitch contour for the neutral tone, and its actual contour is determined by the tone of the preceding syllable.

Tones in Context

When produced in context, lexical tones undergo variations. Some may change their tonal categories in certain tonal contexts. This kind of categorical tone shift is usually referred to as "tone sandhi." Other contextual tonal variations produce only allophonic changes or phonetic perturbations. The literature on contextual tonal variation covers both categorical tonal changes and phonetic perturbations. Among the earliest studies of contextual tonal variation were those by Y. R. Chao, who, with his keen ear for phonetic

[3] An alternative way of marking the tones is used here which puts the tone marks on the last letter in a syllable. Interested readers may contact the author for the reference.

subtleties, made many observations on contextual tonal changes (Chao, 1948, 1956, 1968). The following are his major observations.

a) When a low tone is followed by any tone except another low tone, only the first half of its pitch contour in citation form is used, thus it has the pitch contour 21 instead of 214;

b) A low tone changes into a rising tone when followed by another low tone;

c) At conversational speed, a rising tone changes into a high tone when preceded by a high or rising tone and followed by any other tone except the neutral tone.

d) When a falling tone is followed by another falling tone, the first only falls to the middle, i.e., having a pitch contour of 53.

Among the above four observations, (a) has met the least dispute. Observation (b) was questioned by some. Zee (1980a, p. 121), for example, observes that despite the similarity in contours, the F0 values are "higher for the first syllable in the compound with underlying /35 + 214/ than for the compound with underlying /214 + 214/." Shen (1990) also notices that the low tone after sandhi does not rise as high as the rising tone. However, perceptual experiments by Wang and Li (1967) demonstrate that for listeners, the derived rising tone after sandhi is the same as the real rising tone produced in the same tonal context. Also Xu (1991) finds that even in performing silent short-term memory tasks, Mandarin speakers confuse the derived rising tone with the original rising tone when it is followed by a low tone.

Chao's third observation has encountered the most challenges. Zee (1980b), for example, reports that after a high or rising tone and before a non-neutral tone, the F0 contour of the rising tone on the second syllable does not always change to High-Level, i.e., having the tonal value of [55]. Shih and Sproat (1992) find that, when preceded and followed by the high tone, the rising tone still has different F0 contours from the high tone. They argue that the rising tone sandhi rule

> is the result of a phonetic implementation rule, which applies to the low target of a rising tone in high tone context when the rising tone in question is in prosodically weak positions. The amount of pitch drop to the low target of a rising tone varies with the prosodic strength of the syllable. As a result, the pitch contour of an extremely weak rising tone in high tone context approaches the shape of a high level tone (p. 193).

More recently, Moore (1993), after examining four Mandarin Chinese classifiers representing each of the four phonemic Mandarin tones embedded in the middle of a trisyllabic phrase, finds that these classifiers (one of them has the rising tone) do not completely lose their tones, but maintain their underlying duration properties and tonal targets.

A study by Wu (1988), however, finds that if the ending pitch of the first syllable and the starting pitch of the last syllable are both high, the tone of the second syllable assumes a contour that has a shape similar to that of the high tone. He also finds that if the ending pitch of the first syllable is high but the starting pitch of the last syllable is low, i.e., the first tone is a high or rising tone, and the last tone is a rising or low tone, the F0 contour of the second syllable goes from high to low, becoming a high-falling transition.

As for Chao's fourth observation, there is a dispute not so much over the phenomenon itself as over its nature. Chao (1968) attributes the half fall in the first falling tone to the lesser stress on the first syllable in a disyllabic sequence. Shen (1990), however, argues that the half fall of a falling tone before another falling tone results from tonal coarticulation.

The kind of tonal variation described by Chao is mostly detectable by the naked ear. When instruments are used, more subtle perturbations may be observed. A number of instrumental studies have been conducted on contextual tonal variations in Mandarin as well as in other tone languages of East Asia that are also said to have contour tones (Pike, 1948).

Han and Kim (1974) examine the influence of the six Vietnamese contrastive tones on the phonetic characteristics of another tone in an adjacent syllable by analyzing the acoustic measurements on disyllabic utterances. They find that the overall F0 height of a tone varies considerably depending on its immediate tone environment and also on its syllable position: a variant of a tone adjacent to a high tone is higher than a variant adjacent to a low tone. They find that in general the phonetic variation of the tones is greater in the second syllable than in the first syllable, indicating that in Vietnamese the carryover effect on tones is greater than the anticipatory effect. They also observe that,

> a phonetic overlap between two different phonemic tones occurs only in different environments. However, the phonetic overlap caused by such a purely phonetic conditioning factor does not affect the perceptual identities of the tones when they are presented with the relevant environments. Apparently the pitch information in the neighboring syllables serves as a basis on which the retrieval of the phonemic status of a tone is made (p. 232).

Abramson (1979a) studies all possible sequences of two tones from the five tones of Thai — mid, low, high, falling, and rising, and he finds that,

> (1) the full system of five tones is preserved on monosyllabic Thai words embedded in all possible preceding and following tonal contexts. (2) With citation forms taken as the standard, embedding a tone in running speech causes some perturbation, but generally not enough to damage its identifiability. (3) Laryngeal coarticulation is governed by the specific tonal context; nevertheless, in some contexts coarticulation does not occur for some speakers, and in such instances the larynx may have received instructions to reset itself to produce an approximation of the ideal contour for the tone (p. 7).

A recent study of tonal coarticulation in Thai (Gandour, 1992a) finds the carryover effects of the first syllable on F0 height and slope of the final syllable to be minimal. Gandour also observes, however, that the extent to which tonal coarticulation is triggered depends on the magnitude of the difference in F0 height and slope of the preceding tone. Of the three pairs of tone combinations (falling + low / mid + low, falling + mid /

low + mid, high + rising / rising + rising), in only one (falling + mid / low + mid) the two opposing tones in the first syllable differ in terms of the relationship of F0 height of the first syllable to that of the final syllable. And so it is only in this pair that the F0 slope of the mid tone varies depending on the tonal category of the first syllable, rising in the beginning portion when following a low tone, falling when following a falling tone.

In another study (Gandour, 1992b), Gandour finds clear evidence of anticipatory effects between Thai tones. Interestingly, the F0 contour of the falling tone is generally higher and steeper when occurring before the low and rising tones, both of which have a low F0 onset. But when occurring before another falling tone, which has a high F0 onset, its contour is lower and shallower. So, as Gandour observes (p. 121), "instead of a lowering effect, the low and rising tones induced a raising of F0 in the preceding falling tone."

H.-B. Lin (1988) conducts several experiments on the acoustics of tonal coarticulation in Taiwanese. She finds that "compared with what has been found in Thai and Vietnamese, Taiwanese showed much smaller coarticulatory effects (p. 137)."

Coming back to Mandarin, Shih (1986) investigates all the 16 tonal combinations in disyllable pairs, and observes both anticipatory and carryover effects. Because of their complexities, her observations will be discussed in detail later in relevant contexts.

Shen (1990) examines all possible combinations of the four Mandarin tones on trisyllabic sequences and makes the following observations (pp. 293-294).

(1) Mandarin tones are affected by both carryover and anticipatory coarticulations, and the bidirectional effects are symmetric.

(2) In Mandarin, as in Vietnamese and Thai, not only are tonal onsets and offsets affected by coarticulation, but the tonal contours are shifted up or down by the surrounding tones.

(3) Tonal coarticulation affects only F0 height, not F0 direction. However, F0 directions of the surrounding tones influence the variations of F0 heights. The registers of the tonal offsets being equal, tones following a rising offset are higher than those following a falling offset.

M. Lin and Yan (1991) investigate tonal coarticulation in all the tetrasyllabic combinations and draw the following conclusions.

(1) Tonal coarticulation only affects two adjacent syllables, that is, the coarticulatory effect does not go beyond a neighboring tone.

(2) Tonal coarticulatory effects are unidirectional, that is, Mandarin tones are affected by either the carryover coarticulatory effect or the anticipatory coarticulatory effect. More specifically, the starting points of tone 1, tone 2 and tone 3 are affected by carryover effects; and the ending points of tone 4 and tone 3 are affected by anticipatory effects.

(3) Tonal coarticulation is manifested in terms of changes in F0 height but not in F0 direction. Changes in F0 height do not affect the tonal identities.

Existing Perceptual Studies

Most of the perceptual studies in the literature focus on the perception of isolated tones (Abramson, 1962, 1976, 1978, 1979b; Chuang, et al., 1971; Gandour, 1981, 1983, 1984; Howie, 1972; M. Lin, 1988; H.-B. Lin & Repp, 1989; Shen & M. Lin, 1991; Whalen & Xu, 1992; Yang, 1989). In general, those studies find good recognizability for tones produced in isolation.

There are far fewer studies in the literature on the perception of tones produced in context. The study by Wang and Li (1967), mentioned before on the perception of the low tone after sandhi, is one of the earliest. Fang (1990) also studies the perception of tone sandhi in Mandarin. With two sets of continua in which the tones of the monosyllables and disyllables are varied in steps from the rising tone to the low tone, he finds that compared with the monosyllables, the phoneme boundaries between the rising and the falling tones of disyllabic sequences are shifted toward the second tone. Fang's finding indicates that perceptually, listeners expect a more rising contour for a low tone followed by another low tone than for a low tone produced in isolation, which in part confirms the finding of Wang and Li (1967).

T. Lin and Wang (1984) examine the effect of the relative F0 difference between a stressed syllable and the following neutral syllable on perception of the stressed tone. They find that a high tone is heard increasingly more frequently as high, rising or low tone as its F0 value relative to that of the following neutral syllable changes from about the same, to moderately lower, and to much lower. This finding indicates that listeners use the F0 information in the context when judging the tone category of a particular pitch pattern.

Summary of Previous Findings about Production and Perception of Tones in Context

To sum up, previous studies have observed various perturbations of the lexical tones produced in context. Some of the perturbations may or may not involve change of tonal categories, e.g., low tone before low tone, or rising tone in a trisyllabic sequence (Chao, 1948, 1968); some of them appear more like mere phonetic coarticulation, e.g., perturbations in F0 onsets and offsets or in overall F0 height (Shen, 1990; Shih, 1986); still others are somewhat mysterious, e.g., the raising of the falling tone by a following low or rising tones in Thai (Gandour, 1992b).

Previous perceptual studies, in general, confirm that the change of a low tone into rising tone before another low tone is categorical (Fang, 1990; Wang & Li, 1967), although it is still not settled whether the derived rising tone and the lexical rising tone are phonetically the same (Shen, 1990; Zee, 1980b). Existing perceptual studies also find evidence for the importance of the F0 level of the tonal context relative to that of the target tone for tone perception (T. Lin & Wang, 1984). However, the exact mechanism of contextual influence on tone perception still remains unclear.

A Brief Review of Coarticulation Models

Much of the contextual tonal variation may be viewed as a form of coarticulation — the phenomenon in speech that the actualization of the intuitively discrete and invariant phonetic units varies depending on their surrounding phonetic environment. The issue of coarticulation is particularly important for our understanding of the speech communication process because of the apparent discrepancy between the lack of invariability of the phonetic unit in the speech signal and the ease of listeners in correctly perceiving the linguistic units intended by the speaker; this discrepancy is now believed to be a result of, among other things, the process of coarticulation.

Many efforts have been made to explain coarticulation, especially its function as a link between the presumably concrete, invariant linguistic unit and the context sensitive, ever changing articulatory and eventually acoustic output. Among the earliest efforts were those that searched for invariant neural commands for individual segments (Cooper, Liberman, Harris, & Grubb, 1958; Liberman, Cooper, Harris, MacNeilage, & Studdert-Kennedy, 1967; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). It was hypothesized that the neural commands for phonemes are relatively invariant, and the contextual variability is only the result of "temporal overlap in the effects of adjacent commands on the mechanically sluggish articulatory system" (MacNeilage, 1985, p. 15). This hypothesis was found later not to be supported by the results of electromyographic studies, which showed that the instructions sent to the muscles for a given vowel or consonant varied with the identity of the surrounding segments (MacNeilage, 1970).

Hypotheses about coarticulation that assume invariant neural commands may also be questioned from another line of reasoning. If, as assumed by such a hypothesis, the overlap between phonetic units occurs fairly late in speech production, in cases where two adjacent phonetic units have to share the same articulator, and if those two units happen to have opposing requirements for that articulator, the conflict would have to be resolved as late as after the muscle contraction, and that articulator would have to be pulled simultaneously from different directions during the period of overlap. This would imply that much of the speech production process is that of creating conflicts between the neural commands and letting the conflicts be cushioned only by the sluggishness of the articulators. This would run against the notion that speech production is a highly skilled human behavior, and should be in general conducted harmoniously and with high efficiency.

A second group of accounts are those that assume features to be invariant properties of linguistic units and feature-spreading to be the mechanism for coarticulation (Benguerel & Cowan, 1974; Henke, 1966; Keating, 1985, 1988; Moll & Daniloff, 1971). The difficulty with feature-spreading accounts, however, is the frequently found gradient nature of the coarticulatory variation. By definition, features are supposed to be "distinctive". Hence, anything less than all-or-none would be incompatible with the very notion of distinctiveness inherent in the feature concept.

Another group of accounts try to explain coarticulation in terms of coproduction (Fowler, 1977, 1980), overlapping gestures (Liberman et al., 1967; Liberman & Mattingly, 1985), or gestural phasing (Browman & Goldstein, 1985, 1986). According to these accounts, gestures responsible for neighboring linguistic units actually overlap with one another, and are coproduced during the overlap. While they seem adequate in explaining most local coarticulation, those accounts have difficulty dealing with long distance coarticulations, particularly those between vowels (Magen, 1989; Whalen, 1990). One remedy is to define gestures at a more abstract level (Fowler & Smith, 1986), and assume, for each gesture, a prominence curve with a tail that can extend across several units (Whalen, 1990). However, although the tail hypothesis may sufficiently describe the long distance coarticulations, the concept it entails is somewhat incompatible with the articulatory nature of the coproduction theory, and thus seems ad hoc within its own framework.

Yet another approach is to explain coarticulation with so-called target-based (MacNeilage, 1970) or goal-oriented (Chistovich & Kozhevnikov, 1969) models. As hypothesized by MacNeilage (1970, p. 190, emphasis in the original), the articulation process is controlled by a two-loop system in which, "the open loop component emits a *context-independent* command for an articulator to reach a certain position," and "the closed-loop control circuits constantly sample the mechanical state of the articulator and adjust the command accordingly." One potential difficulty with this kind of models, as pointed out by Kent & Minifie (1977, p. 122), is that of deciding the degree and nature of the target specifications:

> If the targets are concrete and positionally well-defined, then the model probably could not account for such coarticulatory effects as the anteroposterior variation in dorsal stop constriction (that is, the difference in the lingual closure for /k/ in the words *key* and *coo*). On the other hand, if the targets are abstract and only loosely specified as to position, the control system may allow too much variation and thereby jeopardize phonemic identity.

While efforts to explain coarticulation usually seem to assume that there is only one kind of coarticulatory mechanism, it is conceivable that all coarticulatory phenomena might not result from a single mechanism. For example, whereas there is no way two adjacent segments can be produced without any coarticulation, because no articulator is capable of changing its position instantaneously, it is possible that two nonadjacent neighboring segments be produced with minimal coarticulation, as Whalen (1990) has shown for anticipatory coarticulation. In other words, while some cases of coarticulation is utterly unavoidable, others may be avoided at least under certain conditions. Furthermore, while short distance carryover coarticulation could be relatively easily explained in terms of mechanical inertia, it would be rather difficult to account for long distance anticipatory coarticulation in terms of inertia. It is thus quite conceivable that phenomena conventionally covered under the term coarticulation might turn out to be resulting from different mechanisms.

Coarticulation and Tonal Variation

Most of the studies of coarticulation are focused on consonants and vowels, the so-called segmental units. A segmental unit typically has more than one correlate in both acoustic and articulatory terms. For example, a vowel is typically specified by at least two resonance frequencies, namely, formants, and each of the formants is related in very complicated ways to the shape of the vocal tract (Fant, 1960, 1973). It would be interesting to study some phonetic units, such as tones, that have fewer dimensions than the segmental units. Tonal contours are defined in terms of pitch. The articulatory correlate of pitch, i.e., the number of open-and-closed cycles of the glottis occurring during a given period of time, has a simple relation with the acoustic correlate of pitch, namely, fundamental frequency (F0). This property simplifies the relation between articulation and its acoustic consequence. It makes the study of tonal coarticulation simpler than the study of segmental coarticulation, because, for example, while it usually takes several formants to infer the shape of the vocal tract,[4] fundamental frequency (F0) measurements alone is sufficient to determine the actual articulatory output of the vocal folds. Because of this, teasing apart the major factors that control the fundamental frequency in a tone may be relatively easy compared with doing the same thing for the segmental units. This is especially important in dealing with coarticulation, a process which is itself complicated.

Moreover, unlike the coarticulation between vowels and consonants, but similar to coarticulation between vowels, tonal coarticulation is the coarticulation between the same

[4] A recent study shows that to represent a vowel completely, the entire spectral shape rather than just the formant frequencies is needed (Zahorian & Jagharghi, 1993).

kind of gestures in nature. Usually, coarticulation refers to cases where the influenced segment has some feature or articulatory gesture that is not definitely defined, and its surface value would depend on the influencing segment in which that feature or gesture is more completely defined. In the case of tonal coarticulation, adjacent tonal gestures share the same articulator, i.e., the vocal folds. If two adjacent syllables are both stressed, and if their tones happen to have opposing pitch values at the syllable boundary, the two tones would have to compete with each other for the control of the vocal folds in deciding the final output — F0. It would then be intriguing to see how this kind of conflict is resolved. Although the present thesis is not an articulatory study, it is hoped that acoustic measurements as well as perceptual experiments would provide useful evidence for or against different models of coarticulation.

Finally, although the term "coarticulation" refers to phonetic variations that are due to contextual differences, not all context-dependent variations are clear-cut cases of coarticulation. Two points are particularly relevant for the present study.

First, since the issue of coarticulation came into being mainly due to the discrepancy between seemingly invariant phonetic units and their extremely context-sensitive acoustic realization, it is assumed, in this thesis, that any contextual change that actually results in a change of perceived phonological category (e.g., the tone sandhi involving the low tone in Mandarin, Chao, 1968) is not a typical form of coarticulation, although there may be cases where a particular phonological rule stems from mechanisms similar to that of coarticulation.

Second, the concept of coarticulation is given rise to by the finding that the time period during which the production of a given phonetic unit apparently dominates also carries characteristics of the neighboring units, and that properties of any given unit can often be found in other nearby units. To say that neighboring linguistic units are coarticulated implies that articulatory movements responsible for their production overlap one another in time. Hence, whatever its mechanism might be, it is assumed in this thesis that the effect of coarticulation is assimilatory: a phonetic unit would carry some of the properties of its neighboring units. In other words, contextual variations that are apparently not assimilatory are not considered covered under the term coarticulation.

Focus of the Thesis

Based on the above understanding of the state of the art in the study of contextual tonal variations as well as the current status in the study of speech coarticulation, this thesis will focus on the following questions.

1) The true nature of rising tone variation in trisyllabic sequences: is it a categorical change?

2) How are the coarticulated tones perceived? What is the mechanism of this perception?

3) What is the extent of tonal coarticulation? Does it affect only the onset and offset of a tone, or its overall F0 contour as well, or even the contour directions?

4) What is the relative strength of anticipatory and carryover effects? Are they both of the same nature? What is their scope?

This thesis will attempt to answer each of the above questions. The body of the thesis is divided into three chapters. Chapter 1 consists of one acoustic analysis and four perceptual experiments, which address questions (1), (2), and part of (3). Chapter 2, which consists of only acoustic analyses, tries to answer part of question (3) and question (4). Chapter 3 offers a unified view of contextual tonal variation in Mandarin.

Chapter 1 — Production and Perception of Coarticulated Tones in Trisyllabic Sequences

Background

Chao (1948, 1968) proposes a tone sandhi rule that states that, at conversational speed, a rising tone changes into a high tone when preceded by a high or rising tone and followed by any other tone except the neutral tone:

$$[\text{rising}] \rightarrow [\text{high}] \;/\; \left\{ \begin{array}{c} [\text{high}] \\ [\text{rising}] \end{array} \right\} \;\_\_\; [\text{-neutral}]$$

Although, as mentioned before, this observation is challenged by some acoustic analyses (Moore, 1993; Zee, 1980b), it is not without empirical support. Wu (1988) for example, finds that if the ending pitch of the first syllable and the starting pitch of the last syllable are both high, the tone of the second syllable changes into a tone with a F0 contour similar to that of a high tone. What is more, if the ending pitch of the first syllable is high but the starting pitch of the last syllable is low, the F0 contour of the second syllable goes from high to low, becoming a high-falling transition. Shih and Sproat (1992), while noticing that the rising tone still has a different F0 contours from the high tone when preceded and followed by the high tone, argue that the rising tone variation described by Chao is the result of phonetic implementation, and is attributable to the prosodically weak position the rising tone is in, because the amount of pitch drop to the low target of a rising tone varies with the prosodic strength of the syllable.

While prosodic stress is a very likely condition for the rising tone variation to occur, it does not seem to be the basic mechanism for the variation. Wu's finding, however, hints that it is the pitch value of the tones adjacent to the rising tone that is probably responsible for the variation. Take, say, the Mandarin word *conḡyoúbinǧ* 'green scallion pancake' for example. According to Chao's rule, the following change in tones would take place in fluent speech.

¯	/	_	→	¯	¯	_
high	rising	low		high	high	low

Notice that before the change, the ending pitch in the tone of the first syllable is high, while the starting pitch in the tone of the second syllable is low; likewise, the ending pitch in the tone of the second syllable is high, while the starting pitch in the tone of the third syllable is low. If the underlying tonal values of the rising tone were to be fully realized within the second syllable in connected speech, the articulators responsible for F0 control would need to change very quickly from one state to another at the syllable boundaries, which would be rather difficult due to inertia of the articulators. However, this could be avoided by having the neighboring phonetic units coarticulated. The consequence of coarticulation would be a compromise between adjacent phonetic units in the actually realized surface values, especially when they differ from one another substantially. In careful speech, especially when the syllable is fully stressed, this kind of compromise could result in variations in the onset and offset and even the overall height of the tone, as seen in Shen (1990), where speakers were asked to stress evenly the three syllables to be analyzed. In more natural running speech, however, there could be more compromise. In that case, the combined influence of the first and the last tones on the second tone would be in the direction of flattening the tonal contour or even reversing its direction.

If the tonal variation in the tone of the second syllable in words like *cōngyóubǐng* were a result of coarticulation, Mandarin listeners should also hear it as such. As shown by Fowler (1984) and Fowler and Smith (1986), coarticulatory variation is perceived by listeners as information about the phonetic context that causes the variation rather than as information about the phonetic unit that carries the variation. In the case of *cōngyóubǐng*, this would mean that the tonal variation in the second syllable should not be heard as a shift of phonological categories. Instead, listeners should still be able to identify the underlying tone.

The Coarticulation Hypothesis and its Predictions

Based on the above discussion, it is hypothesized that the rising tone variation in trisyllabic sequences is the result of tonal coarticulation. Following this coarticulation hypothesis, the following predictions are made.

1) Not only does the rising tone undergo perturbation, but also the falling tone is subject to perturbation under comparable conditions;

2) Other things being equal, the magnitude of the contextual tonal perturbation is positively related to speaking rate;

3) Listeners will treat the contextual perturbation as coarticulation, and will hence compensate for it with the help of the tonal context present with the target tone;

4) Listeners' performance in identifying the perturbed tones will not be directly proportional to different magnitude of the perturbation due to different phonetic contexts or speaking rates.

The implication of (1) from the coarticulation hypothesis is obvious. If the rising tone variation in words like *cōngyóubǐng* is due to coarticulation, the falling tone should be also subject to such variation under comparable conditions. That the amount of tonal variation is positively related to speaking rate can be predicted from both the coproduction account of coarticulation (Fowler, 1977, 1980) or the target-based model (MacNeilage, 1970). According to the coproduction account, adjacent gestures overlap in their articulatory trajectories. It is conceivable that the amount of gestural overlap is related to the rate of speech articulation — the faster the rate, the greater the overlap. According to the target-based model, at faster speaking rate, the articulators would have less time to approach a given target, thus would be less likely to reach that target.

Prediction (3) that listeners are able to compensate for coarticulation is based on findings by existing studies on the perception of segmental coarticulations. It has been shown that perception of one segmental phoneme depends on the adjacent phonemes (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Liberman, Delattre, & Cooper, 1952; Mann, 1980; Mann & Repp, 1980), indicating that listeners take coarticulation between adjacent phonemes into consideration when perceiving speech. Furthermore, it is suggested that listeners hear coarticulatory information as information for the influencing segment, rather than attributing it to the segment with which it co-occurs in time (Fowler, 1984; Fowler & Smith, 1986). If the rising tone variation is coarticulatory in nature, listeners should be able to treat it as such and compensate for it while correctly identifying the intended tone.

Prediction (4) that different magnitude of tonal perturbation due to different contexts would not directly affect tone perception is further derived from predictions (2) and (3). If listeners treat the tonal perturbation as coarticulation, they would be prepared for different amounts of tonal variation in different tonal contexts or at different speaking rates, and thus would be able to make an appropriate normalization when trying to identify the intended tone.

Compatible and Conflicting Contexts

To test the coarticulation hypothesis, acoustic analyses and perceptual experiments have been conducted. In those experiments, two kinds of phonetic environments are defined, namely, "compatible" and "conflicting." A "compatible" context is an environment in which adjacent phonetic units share identical or similar values along the phonetic dimension under scrutiny. A "conflicting" context is an environment in which adjacent phonetic units have very different values along that phonetic dimension. The word *conḡyoúbinǧ* 'green scallion pancake' (see Figure 1.1), for example, has conflicting tonal contexts both at the boundary between the first and second syllable, and at the boundary between the second and third syllable, since in both places the pitch values differ substantially. In contrast, the word *daǹbaízhǐ* 'protein', has compatible contexts at both syllable boundaries, because in both places the pitch values are similar.

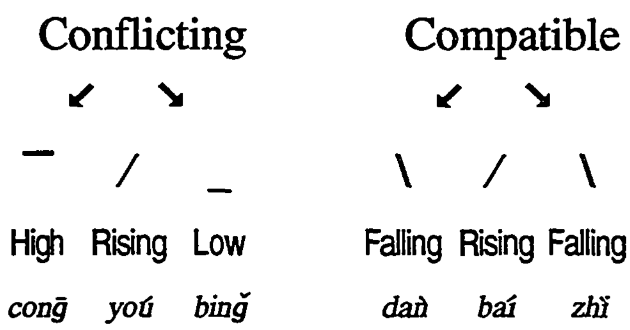


Figure 1.1. Two kinds of phonetic environments — compatible and conflicting contexts.

Different tonal variations are expected for tones produced in compatible and conflicting contexts based on the coarticulation hypothesis discussed above. A conflicting context, for example, those found in *conḡyoúbinǧ*, would force the adjacent phonetic units into a compromise in terms of the actually realized surface F0 values, especially when they differ from one another substantially, resulting in tonal contours that are close to flat or even with different F0 directions. On the other hand, in a compatible context, e.g., those found in *daǹbaízhǐ*, since adjacent tones share identical or similar pitch values at the syllable boundary, little or no compromise is necessary, and thus the pitch values shared by both tones should be realized to the fullest possible extent. Furthermore, the difference between tonal variation in compatible and conflicting contexts should vary depending on speaking rate — the faster the rate, the greater the difference.

As for perception of tonal variations, from predictions (3) and (4) above, it is expected that listeners are able to identify an intended tone when it is presented with the original context, regardless of whether the context is compatible or conflicting. When a tone is presented without the original context, or with a different context, however, the identification of the tone is expected to be affected, and the extent of the effect is expected to depend on the nature of the original context as well as that of the new context, if it is provided.

In the following sections, acoustical analyses and perceptual experiments will be described that are designed to test the above predictions about contextual tonal variation in Mandarin and its perception by native speakers of Mandarin.

Acoustic Analyses

Material

Trisyllabic words or phrases whose second syllables have either the rising tone or the falling tone were used in this experiment. These words had either compatible or conflicting tonal contexts, as defined earlier, for the target tone, i.e., the tone of the second syllable.

There are 16 possible tonal combinations in Mandarin, as shown in Figure 1.2, that satisfy the description of the two categories. All of them were used in the reading list.

¯ \ _	/ \ /	_ / ¯	\ / \	Compatible Context
¯ \ /	/ \ _	_ / \	\ / ¯	
¯ / _	/ / /	_ \ ¯	\ \ \	Conflicting Context
¯ / /	/ / _	_ \ \	\ \ ¯	

Figure 1.2. Tone patterns of trisyllabic sequences having compatible or conflicting tonal contexts.

Speakers

Five Chinese students from The University of Connecticut, three males and two females, served as speakers. They were all native speakers of Beijing Mandarin. For each of the patterns, two words or phrases were found. Five repetitions of each of the 32 words were printed in Chinese in random order.

Recording

The speakers recorded the whole list once with each word spoken in isolation, and a second time with each word spoken in a carrier sentence, *wǒ gaòsu nǐ* ___ *shǐ zeňmo huíshì* 'Let me tell you what ___ is all about.' It was expected that the presence of the carrier sentence would increase the speaking rate, thus increasing the magnitude of coarticulatory tonal variation. Each speaker produced 320 utterances.

During the recording, the overall speaking rate of the speakers was controlled by regularly occurring pure tone beeps played through a loudspeaker by a separate cassette player. The speakers were required to produce each sentence between two beeps occurring at an interval of 3 seconds. The recording was made with a condenser microphone and an Otari tape recorder at a tape speed of 7.5 in./sec.

F0 analysis

All the utterances were digitized by the Haskins PCM system (Whalen, Wiley, Rubin, & Cooper, 1990) at a 20 KHz sampling rate. The extraction of the F0 values was done manually to achieve high accuracy. The wave form of an utterance was displayed on the computer screen, and was stretched so that each glottal cycle is shown clearly. A label was placed at the beginning of each vocal pulse in the wave form of each target syllable displayed on the computer screen. Then, a computer program was used to measure the time intervals between neighboring labels and convert the intervals into frequency values.

For syllables with initial aspirated stops or affricates, F0 measurement started from the onset of voice in the syllable. For syllables with initial fricatives, nasals, the lateral, or unaspirated stops, F0 measurement started from the point of consonantal release.[5] For all the target syllables, the F0 measurement ended right before the constriction for the initial consonant of the following syllable.

For each token, three raw measurements were obtained for the second syllable, namely, duration, mean F0, and slope of the F0 contour. Duration measures the time in milliseconds between the first marked vocal pulse and the last. Mean F0 is the average of the F0 values for all the vocal pulses. Slope is the coefficient of a simple linear regression line for each token, with time of the vocal pulses as independent variable.

Previous tone studies often used the location and F0 value of the turning point in a tone contour, i.e., the point where a contour changes from rising to falling or vice versa, as important measurements (Gårding, 1987; Gårding & Zhang, 1986; Shen, 1990; Shen & M. Lin, 1991). However, because the durations of the second syllables in the trisyllabic words examined in this study were rather short (around 100 ms), there were sometimes only a few vocal pulses in the vowel, and the contour was often very close to a straight line. So it was often difficult to decide where the turning point was. For this reason, turning points were not measured, and the use of slope as the major measurement for tonal contours ignored any possible curvature within a contour. Because this analysis was concerned mainly with comparing tonal variations under different conditions rather than examining the exact contour shapes, slope measurement should be sufficient.

---

[5] Notice that no F0 measurement was make on any initial voiced consonant. It was assumed that tonal alignment is the same for all syllables, regardless of whether the initial consonant is voiced. Since no F0 measurement can be made before the onset of vowel for syllables with voiceless initial consonants, there is no need in measuring F0 during the voiced initial consonant, for lack of counterparts.

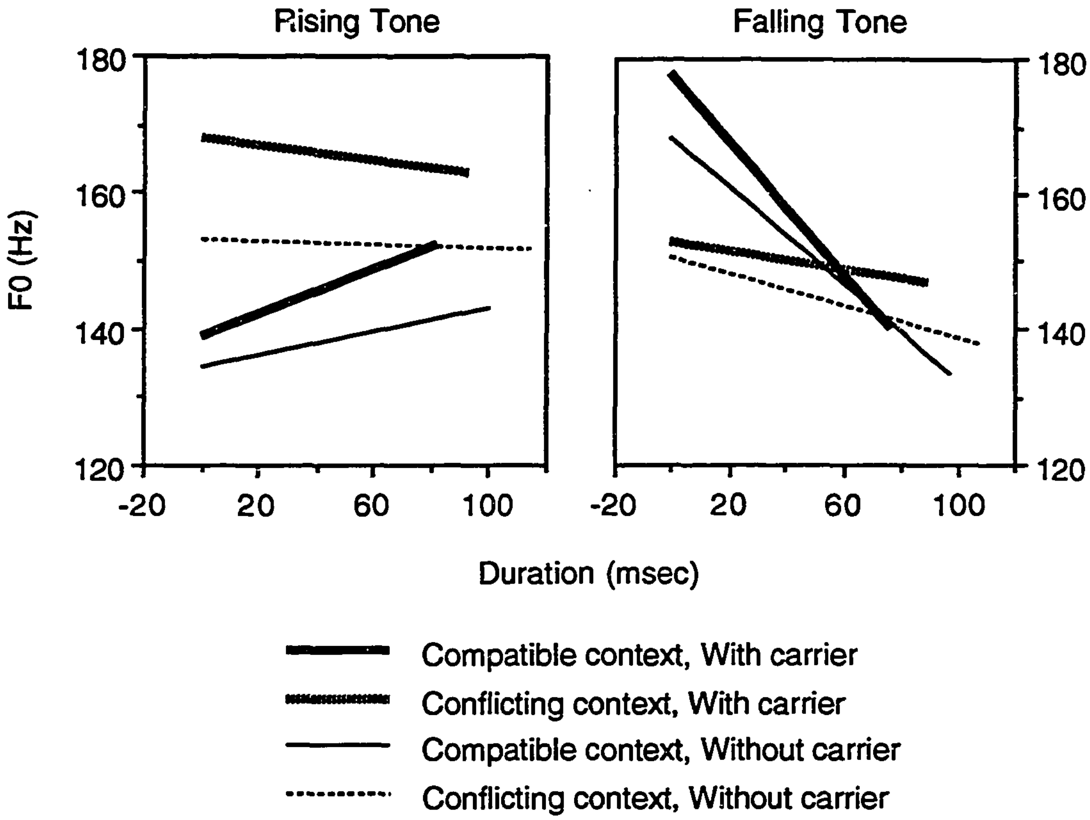


Figure 1.3. Mean F0 slopes, mean F0 height and mean duration for the rising tone and falling tone in compatible and conflicting contexts averaged over all the conditions and speakers. For each line, the mid point represents mean F0; the slope of the line represents the mean slope; and the projection of the length of the line on the horizontal axis represents mean duration.

For each word or phrase spoken by each speaker at one of the speaking rates (with or without carrier), the mean slope, mean F0, and mean duration were obtained by averaging over the individual values of those measurements for all five tokens of the word. Figure 1.3 gives a schematic representation of the mean values of all three measurements by target tone and tonal context for all conditions and speakers. For each line in the figure, the mid point represents mean F0; the slope of the line represents mean slope; and the projection of the length of the line on the horizontal axis represents mean duration. Several things are immediately apparent from this figure. First, for the rising tone, the slope is positive in compatible context but negative in conflicting context. For the falling tones, the slope is much steeper in compatible context than in conflicting context. Second, in compatible context, the slopes are steeper at a faster speaking rate (with carrier) than at a slower speaking rate (without carrier) for both tones; however, in conflicting context, the slopes are steeper for the rising tone at a faster speaking rate than at a slower speaker rate, whereas for the falling tone, the slopes are steeper at a slower speaking rate than at a faster speaking rate. Third, while the rising tone has higher mean F0 in conflicting context than in compatible context, the falling tone has lower mean F0 in conflicting context than in compatible context. Fourth, the duration is shorter at a faster speaking rate than at a slower speaking rate. Fifth, the mean F0 is higher at a faster speaking rate than at a slower speaking rate.

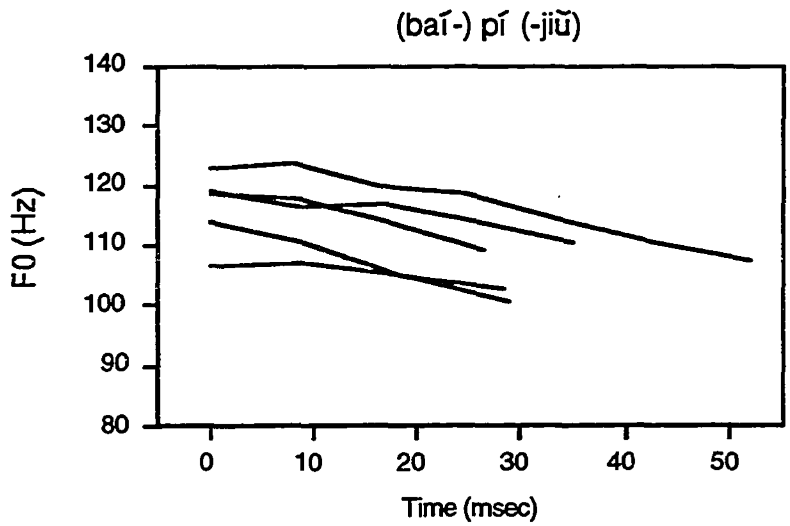


Figure 1.4. F0 contours of 5 tokens of /pí/ in the word "baǐpíjiǔ" (light beer) produced by speaker LC.

Figure 1.4 provides some examples of the rising tone produced with negative slopes in case of conflicting context. The five tokens were produced by one of the male speakers who participated in this experiment. It may be noticed that the duration of those contours are very short. Some of them consists of only four vocal pulses. However, these are only durations of the vocalic segment in the syllable /pí/, not of the entire syllable.

Figure 1.5 shows the grand means over all the speakers for the slopes of rising tone and falling tone in compatible context versus conflicting context. In the conflicting context, the slope values for both tones approach zero, whereas in the compatible context, the slope values for the two tones fall far apart. More interestingly, the mean slope of the rising tone becomes negative under the influence of the conflicting context, although it is still very close to zero. In fact, 11 out of the 16 words in this condition (rising tone in conflicting context) had negative mean slopes. For the falling tone, although the overall mean slope remained negative even in the conflicting context, 3 out of the 16 words in the conflicting context had positive mean slopes.

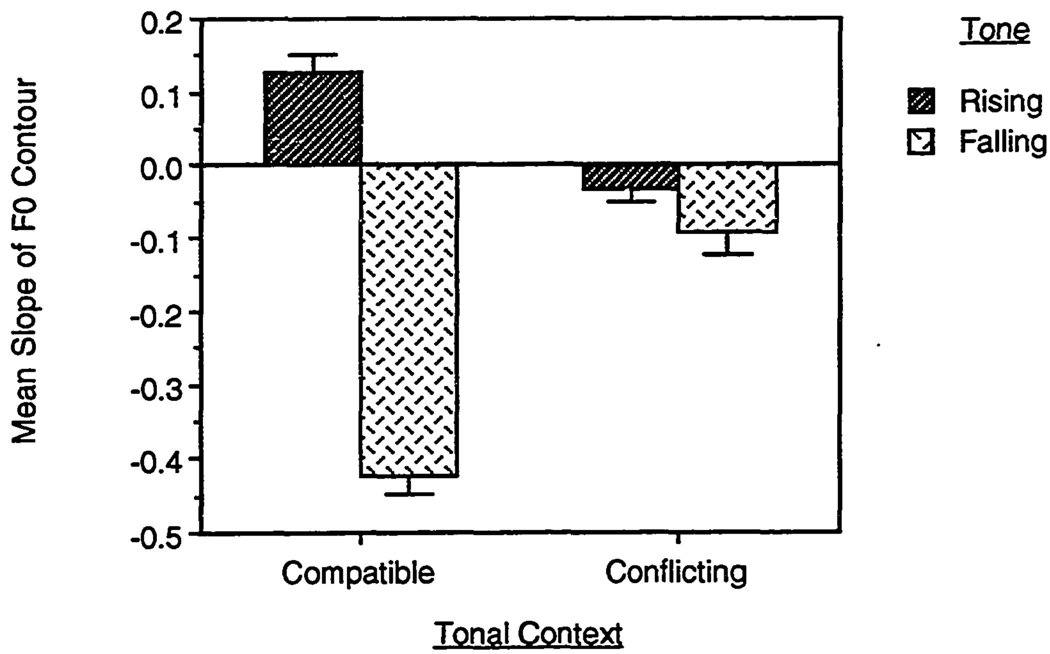


Figure 1.5. Mean slopes of rising tone and falling tone in compatible and conflicting tonal context collapsed over five speakers. The horizontal bars indicate standard error.

To make the slopes of the rising tone and the falling tone directly comparable for statistical analysis, the slope values should be normalized in terms of the direction of the slopes. One way to do this is to take absolute values of all the slopes. However, simply taking absolute values of the slopes would improperly treat cases of "over flattening," i.e., cases in which a rising tone is produced with a negative slope, or a falling tone produced with a positive slope. To take care of those cases, the signs of all the slope values for the falling tone were reversed, while the signs of the slope values for the rising tone remained unchanged. The resulting slope values are referred to as adjusted slopes and are used in the statistical analysis described below. Note that the signs of the adjusted slopes thus obtained indicate whether there is over-flattening, both for the rising tone and for the falling tone, while the absolute values of the adjusted slopes reflect the steepness of the slopes regardless of their directions.

An ANOVA was conducted with context (compatible or conflicting), carrier sentence (with or without), and tone (rising or falling) as factors. The dependent variable was adjusted slope. Two additional ANOVAs were conducted with mean F0 and mean duration as dependent variables respectively.

Overall, the mean adjusted slope for the compatible context (0.276) is steeper than for the conflicting context (0.030), $F(1,56) = 109.39$, $p < 0.001$, indicating a strong influence of the tonal contexts upon the slope of the target tones. The mean adjusted slope for the falling tone (0.260) is also steeper overall than for the rising tone (0.045), $F(1,56) = 83.02$, $p < 0.001$.

There is a significant interaction ($F(1, 56) = 13.29$, $p < 0.001$) between tone and context. More specifically, the difference between the falling and rising tones is greater in a compatible context than in a conflicting context. This is probably because a compatible context helps the F0 contour of a tone to approach its canonical form, whereas the conflicting context reduces the slope of both tones to close to zero. Since the canonical form of the falling tone has a much steeper slope than that of the rising tone, the reduction of the slopes needs to go further for the falling than for the rising tone.

The only significant effect on duration is presence/absence of carrier. As expected, tones produced in a carrier sentence are shorter (85 ms) than those produced in isolation (104 ms), $F(1,56) = 6.25$, $p = 0.015$, indicating a difference in speaking rate between the two conditions.

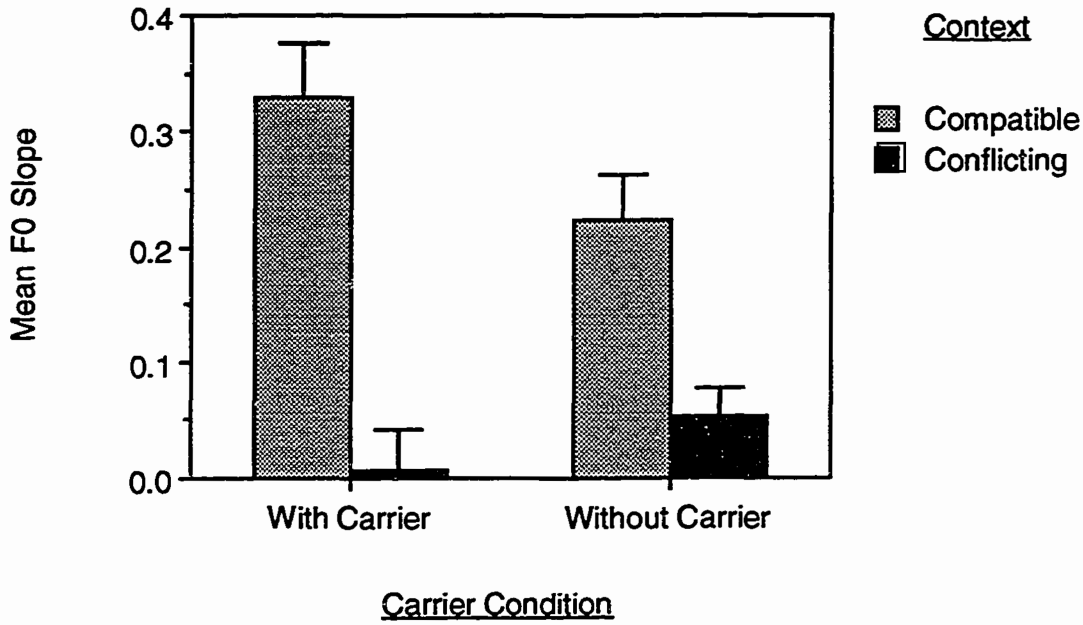


Figure 1.6. Interaction between context and carrier. The horizontal bars indicate standard error.

The main effect of carrier on adjusted slope is not significant. However, there is a significant interaction between carrier and context, $F(1,56) = 10.26$, $p < 0.01$. Figure 1.6 shows that with a carrier, the difference in adjusted slope between compatible and conflicting contexts is larger than without a carrier. Since the comparison of duration confirms the difference in speaking rate between the two carrier conditions, it is clear that tonal variation due to context is greater at a faster speaking rate than at a slower speaking rate. Note that the slopes for the compatible context are greater in the carrier sentences (i.e., when they are short), which shows that the smaller value for the conflicting context is not simply a matter of articulatory undershoot.

The main effect of speaking rate on mean F0 is highly significant, $F(1,56) = 49.12$, $p < .001$. As can be seen in Figure 1.3, regardless of tone or context, mean F0 is higher at faster speaking rate (thicker line in Figure 1.3) than at slower speaking rate (thinner line in Figure 1.3). This indicates a general trend of raising the overall F0 value by a faster speaking rate.

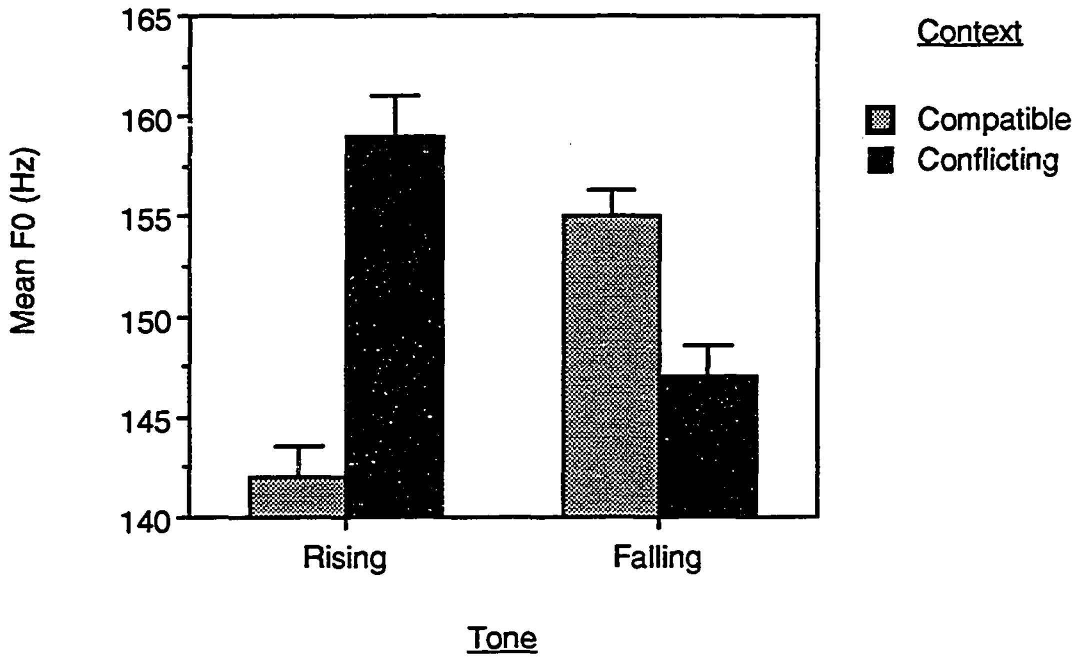


Figure 1.7. Mean F0 under the effect of tone and context. The horizontal bars indicate standard error.

There is a highly significant interaction between the effects of tone and context on the mean F0 of the target tones (See Figure 1.7). The rising tone had a higher mean F0 in the conflicting context than in the compatible context, whereas the falling tone had higher mean F0 in the compatible context than in the conflicting context. This does not make much sense until we take into consideration what the contexts are actually like for the two tones. For the rising tone in the conflicting context, the preceding tone ends high. Since in this case the mean F0 is high, it seems that the flattening of the contour is mainly accomplished through raising the beginning value of the rising tone rather than through lowering its ending value (also refer to Figure 1.3). Conversely, for the falling tone in the conflicting context, the low mean F0 seems more like a result of pulling down its beginning value by the low ending pitch of the preceding tone rather than a result of raising its ending value by the high starting pitch of the following tone. In both cases, it seems the preceding tone exerts more influence on the target tone than the following tone. Likewise, in the compatible contexts, the lower mean F0 in the rising tone indicates greater influence from the preceding falling tone that has a low ending F0 value; and the higher mean F0 in the falling tone indicates greater influence from the preceding rising tone that has a high ending F0 value. This apparent asymmetry in the influence of the tonal context upon a target tone may indicate a greater amount of carryover than anticipatory tonal coarticulation in Mandarin. Alternatively, this asymmetry of contextual influences might indicate closer rhythmic relation of the second syllable with the first syllable than with the last syllable in a trisyllabic word or phrase in Mandarin, as suggested by Wu (1984), who proposes that trisyllabic units in Mandarin are made of a disyllabic unit followed by a monosyllable rather than with three syllables equally bound together or with a monosyllable followed by a disyllabic unit. However, because only the tonal influence of the first and the last syllables upon the second syllable was examined here, the present results do not distinguish between these two possibilities. More conclusive answers to this question will be provided by the acoustic analyses in Chapter 2.

In summary, the results of the F0 analyses show that there are significant differences between tones produced in two kinds of contexts: the slopes for tones in a compatible context are much steeper than in a conflicting context, indicating that tonal environment does effectively change the tonal contours of the target tones. Furthermore, in a conflicting context, there occurred instances in which the direction of the tonal contours differed from that of the underlying contours. Thus, for example, a falling contour might have occurred when the underlying tone was rising. The interaction between context and carrier indicates that variation due to context increases at faster speaking rate. The interaction between the effect of tone and context on mean F0 indicates possible greater carryover than anticipatory tonal coarticulation in Mandarin.

Perceptual Experiments

The F0 analysis showed that in connected speech, a tone varied its F0 contour extensively in different tonal contexts, sometimes even to the extent of changing the direction of its slope. It would be interesting then to see how this kind of variation is perceived by listeners. Would they be able to recover the tone categories intended by the speaker? Or, would they rather take the F0 contours at their face value and perceive tones that are different from the intended ones in case the F0 contours deviated much from the canonical forms? If listeners are able to recover the intended tones, how do they do it? The following four perceptual experiments are designed to answer these questions. Experiments 1 and 2 look at identification of tones presented in the original contexts. Experiment 3 examines perception of tones produced in context but presented in isolation. Experiment 4 tests identification of tones presented in tonal contexts that are different from the original ones.

Perceptual Experiment 1

Stimuli

In the production study, the purpose of using real words and common phrases in the reading list was to guarantee that there would be maximum coarticulation, because they would be produced more naturally than nonsense material. To examine the perception of coarticulated tones, however, linguistic information other than purely phonetic information should be avoided. In other words, the design of the perceptual tests should prevent subjects from using their non-phonetic linguistic knowledge in accomplishing the task. To achieve this goal, certain modifications of the acoustic signal were made on each word so that the phonetic structure of one of the syllables was changed. The phonetic modifications were of two kinds, excision or substitution, and they were applied either to the first or to the last syllable in a word, but not to the second syllable. In the process of excision, the initial part of the frication noise in a syllable was cut out so that the phonetic identity of the first syllable was effectively changed. For example, a [s] may be changed into a [ts^h], [ts] or a [t], depending on how much of the original frication noise was cut out, and what vowel followed that consonant. These consonants share almost the same place of articulation in Mandarin, and thus resemble one another in the spectral shapes of their frication noise and the formant transitions into the following vowels. In the process of substitution, the transient noise in a stop or the frication noise in an affricate was replaced by frication noise from an affricate or a fricative. This process could change, for example, a [t] into a [ts], [t^h], [ts^h], or a [s].

The modifications of the acoustic signals were done on a VAX 11/780 computer using WENDY, a wave form editing program developed at Haskins Laboratories.

Table 1.1 is an illustration of the phonetic modifications performed on the trisyllabic words or phrases and the resulting nonsense strings written in Chinese as response choices. In this case, the beginning part of the frication noise of [tɕʰ] was removed from the signal so that it sounded like a stop [t^h]. It was necessary to use Chinese characters as response choices because the Pinyin system is not the conventional way of representing Mandarin speech, and the subjects would not be at ease with it. The syllable that has undergone phonetic modification is represented by a character different from the original one. The other two syllables were also represented by different characters from the original ones, but with the same pronunciation. This is to guarantee that the lexical identity of the original word is totally disguised. This is possible because in Chinese homophonous morphemes are often written with different characters. Because the morphemes themselves are not necessarily words, and their combinations into words are determined by convention rather than by the speaker's choice, the arbitrary combinations of syllables in the response choices as illustrated below can not remind the subject of any real words even if the meanings of the individual morphemes might be combined to make some sense.

Table 1.1. Illustration of phonetic modification and the resulting nonsense strings used as response choices for subjects in Experiment 1.

Characters	transcription	character glosses	word gloss
气象站	[tɕʰì ɕiaŋ̀ tʂaǹ]	'air situation station'	'weather station'
替香战	[tʰì ɕiaŋ̄ tʂaǹ]	'replace incense fight'	nonsense strings used as response choices
替详战	[tʰì ɕiaŋ́ tʂaǹ]	'replace detail fight'	
替想战	[tʰì ɕiaŋ̌ tʂaǹ]	'replace thought fight'	
替向战	[tʰì ɕiaŋ̀ tʂaǹ]	'replace toward fight'	

When recording the stimuli back onto magnetic tape for perceptual tests, the carrier sentences for the target words were not altered in any way and each remained with the (now-altered) words or phrases they carried.

Tokens produced by only one of the speakers were used as stimuli. All modified utterances were rerecorded onto magnetic tapes with five repetitions each in random order. For each tape, an answer sheet was prepared on which the response choices were the nonsense strings of Chinese characters as illustrated in Table 1.1. For the four choices in a trial, all the first characters were the same, and all the last characters were the same. For the first (or last) syllable that had undergone phonetic modification, the character used represented the new syllable. For the last (or first) syllable that had not been modified, the character represented the same syllable, but was a different character from the original character. The four alternative characters in the middle all had the same CV structure as the original character, but each had a different tone (See Appendix 1 for a complete list of the target words and their corresponding nonsense strings used as response choices.).

The modification of the original acoustic signal of the test words, together with the substitution of the characters on the answer sheet, insures that the trisyllabic sequences are no longer meaningful words or phrases, while all the tonal properties characteristic of fluent speech are well preserved.

Subjects

Ten Chinese students from The University of Connecticut who are native speakers of Beijing Mandarin participated in the identification test.

Procedure

A subject listened to the test tape through headphones, and decided, for each trial, which of the four choices printed on the answer sheet was the one he had just heard, and marked his choice on the answer sheet accordingly.

Results

Figure 1.8 shows tone responses for the target tones in this experiment. Each of the four stacked bars represents the distribution of tone responses for a given target tone in a given context. The falling tone was heard as falling tone 99.7% of the time in the compatible context and 97% of the time in the conflicting context. The rising tone was heard as rising tone 94% of the time in the compatible context and 81% of the time in the conflicting context.

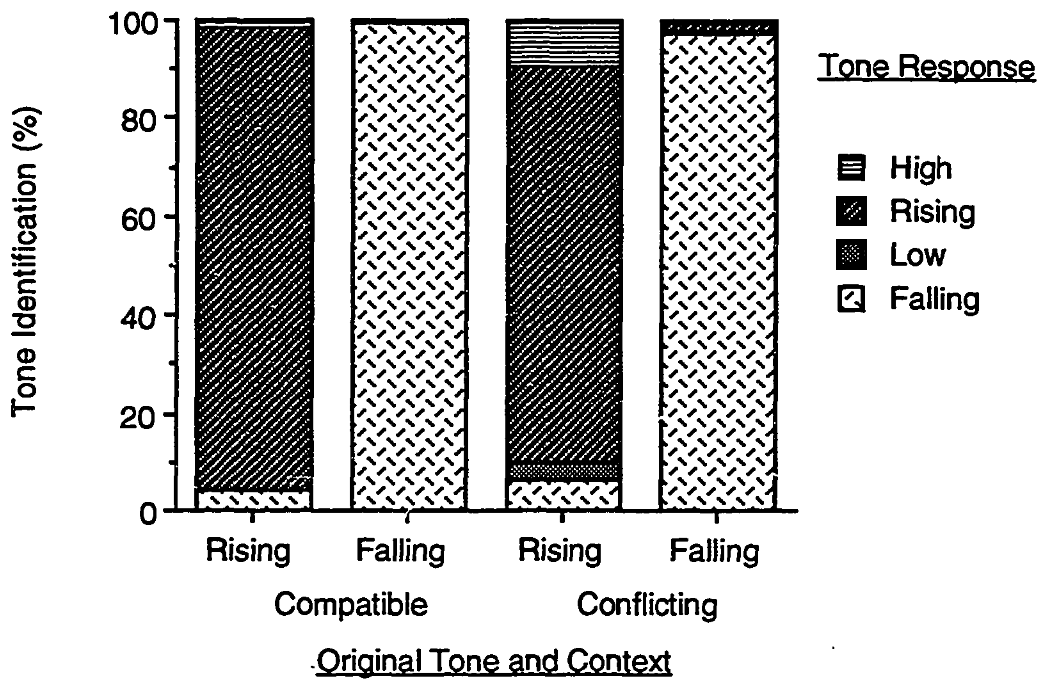


Figure 1.8. Tone responses for the rising and falling tones produced in the compatible and conflicting tonal contexts in Experiment 1.

A three factor ANOVA was conducted with context, tone, and carrier as independent variables (The identification scores were arcsin transformed for the ANOVA to offset the ceiling effect (New_Score = 2 • arcsin • sqrt (Old_Score), (Winer, 1962, p. 400)).

Figure 1.9 shows the effect of context on the accuracy of responses plotted against the effect of carrier. The overall accuracy was high for both contexts (97% for the compatible context, and 88% for the conflicting context) and both carrier conditions (92% for with carrier, 94% for without carrier). While the effect of context was significant, $F(1,9) = 26.73$, $p < 0.001$, the effect of carrier is not. There was no interaction between the two factors.

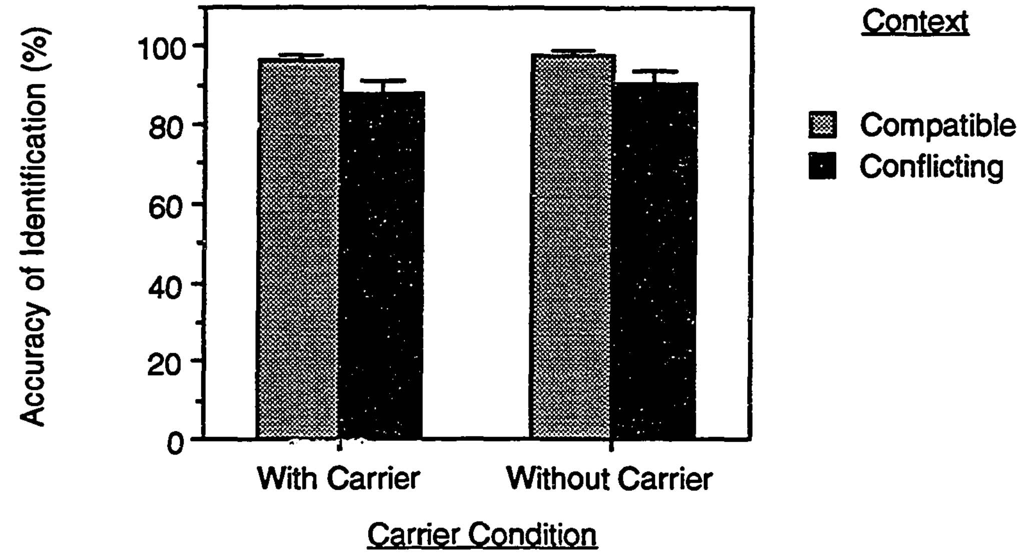


Figure 1.9. Accuracy of tone identification under influence of tonal context and carrier sentence in Experiment 1. The horizontal bars indicate standard error.

Figure 1.10 shows the effect of context plotted against the effect of tone. Both effects were significant. The context effect was discussed earlier. The falling tone had higher identification accuracy (99%) than the rising tone (87%), $F(1,9) = 113.82$, $p < 0.001$. There is also significant interaction between the effects of context and tone, $F(1,9) = 13.39$, $p < .01$. This interaction is partly attributable to the sandhi-triggered low tone responses. For the rising tone in conflicting context, 6% of the low tone responses were for the rising tone in the tonal context of ¯ / _ and / / _ . These low tone responses should be considered as ambiguous between low tone and rising tone percepts. This is because the second syllables were followed by syllables with low tone. Since it is known that a lexical rising tone preceding a low tone is nondistinct from a rising tone derived from a low tone (Wang and Li, 1967) due to the tone sandhi rule in Mandarin, a percept of a low tone in this context would imply that the listener might actually have heard a rising contour. However, the interaction between context and tone remained significant even when those sandhi-triggered low tone responses were treated as correct responses, $F(1,9) = 6.33$, $p = .033$ (The remaining errors consists of 9% high tone responses and 7%

falling tone responses). This indicates that listeners have even more difficulty compensating for the tonal variation due to conflicting context for the rising tone than for the falling tone.

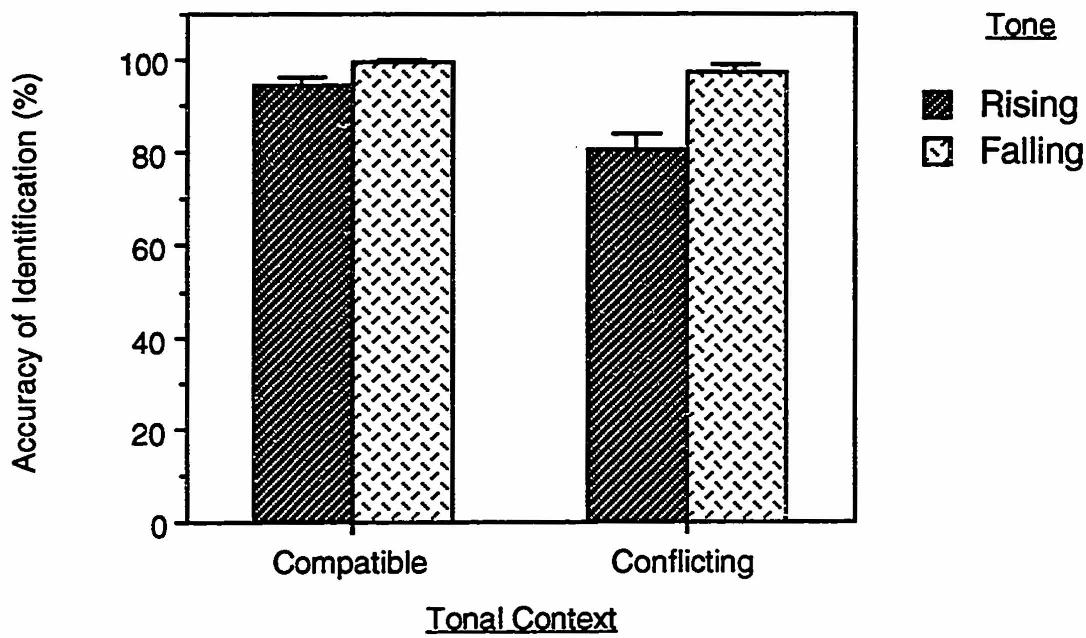


Figure 1.10. Tone identification accuracy under the influence of tonal context and tone in Experiment 1. The horizontal bars indicate standard error.

Discussion

The results are interesting in two respects. First, although as shown in the acoustical analysis, there are large differences between tones produced in compatible and in conflicting context, and for some tokens, the effect of the context is big enough even to change the direction of the F0 contours, the overall accuracy of response is very high for both contexts. This is especially significant when we remember that what the subjects heard was fluent speech stripped of all relevant semantic information. The apparent implication of this performance is that listeners allow for coarticulation when perceiving tones produced in connected speech.

Second, despite the overall high accuracy for both compatible and conflicting contexts, there was a significant difference between the two. When F0 contours were closer to the underlying ones due to the benefit of a compatible tonal context, the recognition was better than when F0 contours were distorted farther away from the underlying forms due to the effect of conflicting tonal context. This indicates that coarticulatory variation may not always be treated as such, and thus be fully compensated for by listeners.

Perceptual Experiment 2

In Experiment 1, only dynamic tones were used as target tones to test whether they could be correctly identified when presented with the original context. Since the static tone that those dynamic tones may be confused with, namely, the high tone, was not used as a stimulus, there is a lack of contrast in the stimuli. It could be argued that a lack of confusing static tone targets may have biased listeners toward hearing the dynamic tones rather than the high tone, because they had nothing to compare them with. To make sure that the high accuracy in tone identification found in Experiment 1 is not simply a result of response bias, Experiment 2 presented the high tone together with the rising and the falling tones in the same tonal context. Also, to verify that the high accuracy of identification is not unique to the speaker, utterances by another male speaker who also served as one of the speakers for the acoustic analysis were used as stimuli for this experiment.

Stimuli

Figure 1.11 displays the tone patterns used in this experiment. The first and the third rows consist of tone patterns containing only conflicting tonal contexts for the second syllable. They are exactly the same tone patterns used in Experiment 1 for the conflicting context condition. The second and the fourth rows consist of tone patterns that are identical with those tone patterns directly above them except for the tone in the middle, which is always the high tone. This arrangement provides direct comparison between the static tone (the high tone) with the dynamic tones (rising and falling tones) that have reduced their dynamic range due to conflicting tonal context.

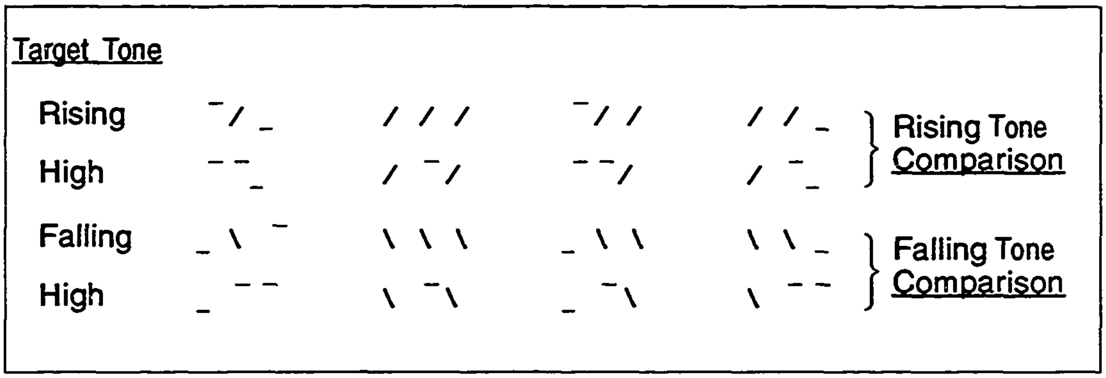


Figure 1.11. Tone patterns for tri-syllabic words and phrases used in Experiment 2.

Thirty-two words and phrases were used so that for each of the 16 tone patterns shown in Figure 1.11 there were two corresponding words or phrases. All of them were repeated five times and put in a randomized list. A male speaker of Beijing Mandarin who also served as one of the speakers for the acoustic analysis recorded the list with each word in the same carrier sentence used in Experiment 1.

As in Experiment 1, the recorded words were edited so that their phonetic structures were altered while their tonal characteristics were kept intact. The edited nonsense strings were then recorded back on tape in random order, each occurring twice, and each with its original context. Unlike Experiment 1, however, for most of the words, both the first syllable and the last syllable were edited. This was to further guarantee that the original words or phrases would not be recognized lexically.

Subjects

The subjects were ten native speakers of Beijing Mandarin who were students at The University of Connecticut.

Procedure

As in Experiment 1, an answer sheet was prepared on which, corresponding to each token, there were four choices each having the same initial character and the same final character but different characters for the second syllable. Each of the four characters for the second syllable represented the same syllable structure with a different tone. (See Appendix 2 for a complete list of the target words and their corresponding nonsense strings used as response choices.).

Subjects listened to the recorded tapes, and indicated for each token what the tone of the second syllable was by marking one of the four choices on the answer sheet.

Results

Figure 1.12 shows the tone identification responses for the target tones in this experiment. Each of the four bars represents the distribution of tone responses for a given target tone in a given context. The original high tone was heard as high tone more than 90% of the time. The original falling tone was heard as falling tone 87% of the time. The original rising tone, although heard as high tone 34% of the time, was heard as rising tone most of the time (63%). As a matter of fact, only three out of the eight words with a rising target tone had identification scores near the chance level of 25%. The rest all had fairly high accuracy of identification, all well above chance, ranging from 74% to 84%.

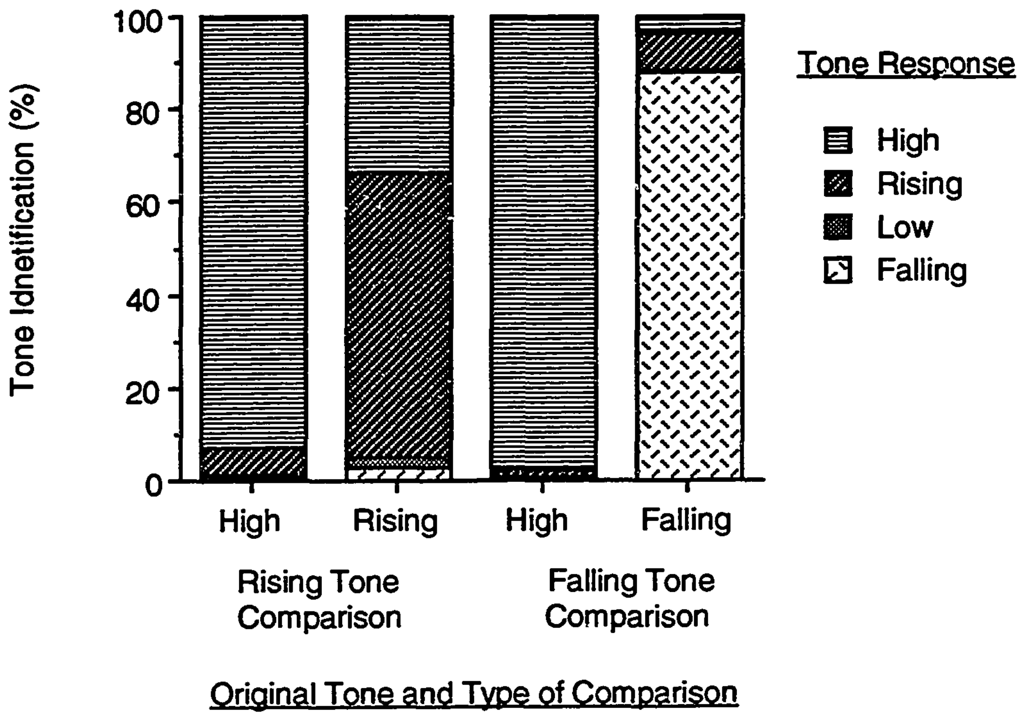


Figure 1.12. Tone responses for the dynamic (rising and falling) and the static (high) tones in Experiment 2.

Perceptual Experiment 3

These results in general confirm the finding in Experiment 1 that dynamic tones produced in conflicting context can still be identified with fairly high accuracy. However, with the presence of original high tones in the stimuli, the rising tones in some of the words were often confused with the high tone. Nevertheless, fewer than half of the words tested had their tonal categories wrongly identified. Also the original high tone having the same tonal context did not show any sign of tonal confusion with the rising tone, which is very different from a case of true confusion of tone categories in Mandarin, where a true rising tone and a rising tone derived from a low tone when preceded by another low tone are confused in both directions (Wang & Li, 1967). Therefore, while the results of this experiment suggest that the rising tone in Mandarin, at least in some speakers' speech, approximates the high tone in some instances, they do not yet indicate a merger of tone categories in the tonal context examined. Instead, the confusion seems more like a failure to compensate for the coarticulatory variation due to the conflicting tonal context.

Stimuli

Although Experiments 1 and 2 found fairly high accuracy of identification for tones produced in both compatible and conflicting contexts, the results did not show unambiguously where the information enabling listeners to allow for coarticulation is located. To make sure that it was the tonal context that was providing some crucial information about the tonal identity of the second syllable, it was necessary to check what would happen to tone perception without the help of the original tonal context. To do this, target tones with their original context replaced by white noise were used as stimuli in this experiment.

The raw tokens used in this experiment were the same tokens used in Experiment 1, except that only words or phrases whose first and last tones were either rising or falling tones were used. For example, words with tone pattern of / \ / were used, whereas words with tone pattern of / \ _ were not. Also, only words or phrases produced with the carrier sentence were used.

For each token, the entire context for the syllable carrying the target tone, i.e., both the preceding and the following syllable and the carrier sentence, was replaced by white noise. The duration of white noise was 500 ms both preceding and following the target syllable.

Subjects

Ten native speakers of Beijing Mandarin who were students at The University of Connecticut participated in this experiment.

Procedure

The stimuli were recorded onto a magnetic tape in random order, each repeated twice in immediate succession. An answer sheet was prepared on which, corresponding to each syllable, there were four characters representing syllables with different tones including the target tone, but all having the same CV structure as the target syllable. No effort was made to avoid using characters identical to those in the original words, because isolated single characters would not remind the subjects of the original words.

The subjects' task was to identify the tone of each target syllable by marking the appropriate character on the answer sheet.

Results

Subjects' response patterns are shown in Figure 1.13. Each of the four bars represents distribution of tone responses for a given target tone in a given context. It can be seen that response patterns differ dramatically depending on whether a target tone was produced in compatible or conflicting context. When the context was compatible, tone responses mostly agreed with the intended tone. When the context was conflicting, the majority of the tokens were perceived as having high tone whether the original tone was rising (74%) or falling (54%). The correct identification of both rising tone (17%) and falling tone (18%) dropped well below the chance level of 25%.

Overall, the identification of the tones in a compatible context was significantly higher than that of the tones in a conflicting context, $F(1, 9) = 357.70$, $p < 0.001$. There was no difference between the identification of the two tones, and there was no interaction between the effect of tone and context.

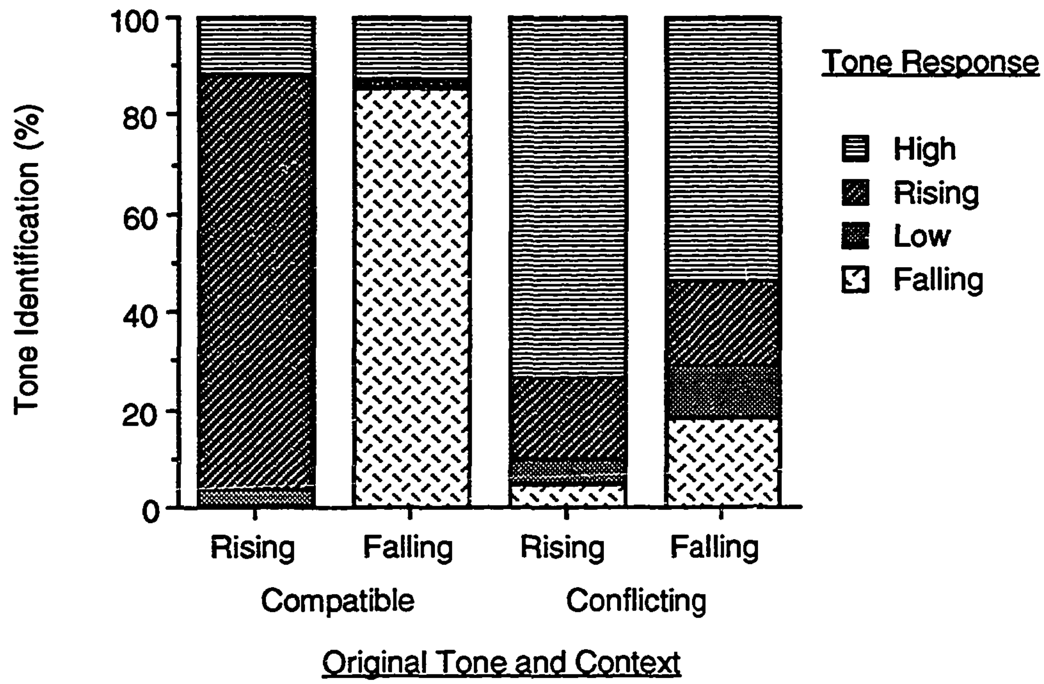


Figure 1.13. Tone response patterns for syllables whose original context was replaced by white noise in Experiment 3.

Discussion

The results of Experiment 3 demonstrate that when a tone is stripped of its original context, its identification varies dramatically depending on the nature of that context. When the context was compatible, identification scores remained high; when the context was conflicting, identification dropped below chance. This indicates that without the adjacent tones, listeners can only attribute the coarticulatory variation in tonal contour to the target tone itself. It also shows that a conflicting tonal context does effectively change the tonal contours to the extent that they resemble some other tone categories when heard without the original context. The fact that listeners heard as high tones most of the tones originally produced in the conflicting context indicates that the change of direction in the tonal contours observed in the F0 analysis, as discussed earlier, does not seem to be enough to make a listener hear a rising tone as a falling tone, or conversely. This result agrees with the finding by Whalen and Xu (1992) that Mandarin syllable excerpts with close to level F0 contours were heard mostly as the high tone, except when the absolute F0 is very low.

Perceptual Experiment 4

Experiments 1, 2 and 3 showed that context plays an important role in perception of tones. Presented with the original context, a tone can be identified correctly most of the time even if it has undergone severe distortion due to the context. Without the original context, identification of the same tone drops dramatically if it has been severely distorted by the context. The results of the first three experiments do not show definitely, however, what exactly in the context is responsible for the accurate tone identification in Experiment 1 and 2, and for the low accuracy of tone identification in Experiment 3. To confirm that it is the tonal information in the context, and not some other factor, that is responsible for the difference in tone identification across the previous experiments, changes should be introduced into the *tonal* context to test if tone identification is affected in the direction that can be predicted by a coarticulatory account of tone perception.

Stimuli

The first and last syllables in the trisyllabic words and phrases used in Experiment 3 were swapped in preparing the stimuli for this experiment. For each word, the position of the first and last syllables were exchanged by excising both of them from the original context in the acoustic signal and putting each into the original position of the other. Since lexical identity was destroyed when the first and last syllables were swapped, there was no need for any further phonetic modification on the initial consonant of the first or the last syllable.

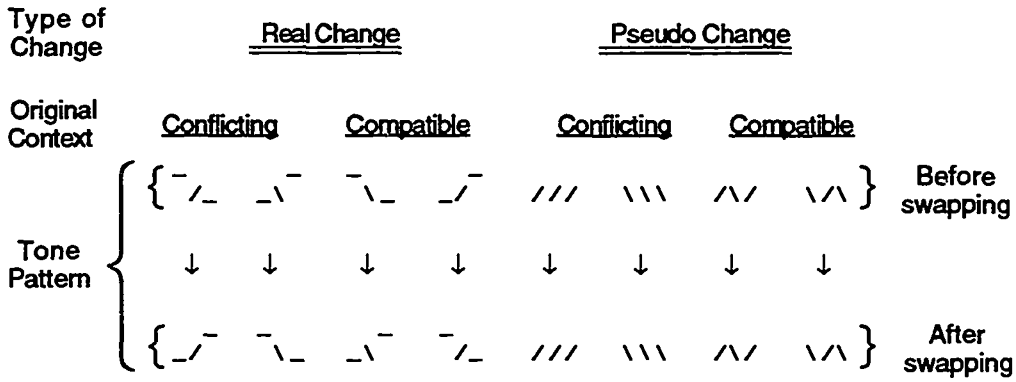


Figure 1.14. Swapping design for Experiment 4.

As a result of this simple swapping, two kinds of changes were introduced into the tonal context for the target tone, as shown in Figure 1.14. The first kind of change, referred to as real change, occurs in words whose first and last tones are different, for instance, words with the tonal pattern of ¯ / _ . The second kind of change, referred to as pseudo change, occurs in trisyllabic sequences whose first and last tones are the same, for instance, words with the tonal pattern of / / / . In both conditions, dramatic changes are introduced at the segmental level, because in both cases the first syllable and the last syllable are different in their CV structures. However, the pseudo change condition does not introduce significant alteration of the tonal environment for the second syllable, since the tonal values adjacent to the target tone remain the same, whereas the real change condition shifts the tonal values adjacent to the target tone to the opposite extreme, i.e., from high to low, or from low to high, thus changing an originally compatible context into a conflicting context, and vice versa.

As in Experiment 3, only tokens produced in the carrier sentence were used in this experiment, and all of them underwent the swapping operation described above.

All the modified utterances were recorded, together with the carrier sentences, onto magnetic tapes in random order with 10 repetitions each. For each tape, an answer sheet was prepared on which the test words were written in Chinese. For each trial, there were four characters printed as choices for the second syllable, each having a different tone in Mandarin, but all with the same CV structure as the original syllable. The first and the last characters were printed only once for each trial. They both represent the original syllables except that their positions are exchanged. Most of them are different characters from the original ones except in cases where no alternative characters could be found for the particular syllables. In no case were all the three characters the same as those in the original word.

Subjects

Ten Chinese students at The University of Connecticut, all native speakers of Beijing Mandarin, participated in the identification test.

Procedure

The subjects listened to the test tape through headphones, and decided, for each trial, which of the four choices printed on the answer sheet was the one they had just heard, and marked their choices on the answer sheet accordingly.

Predictions

The two types of original tonal context, namely, compatible and conflicting, were expected to have different contributions to the perception of the target tones. For the pseudo change condition, since no essential changes were introduced into the tonal environment, perception of the target tone was expected to be similar to that of Experiment 1. For the real change condition, identification of the target tone was expected to differ dramatically.

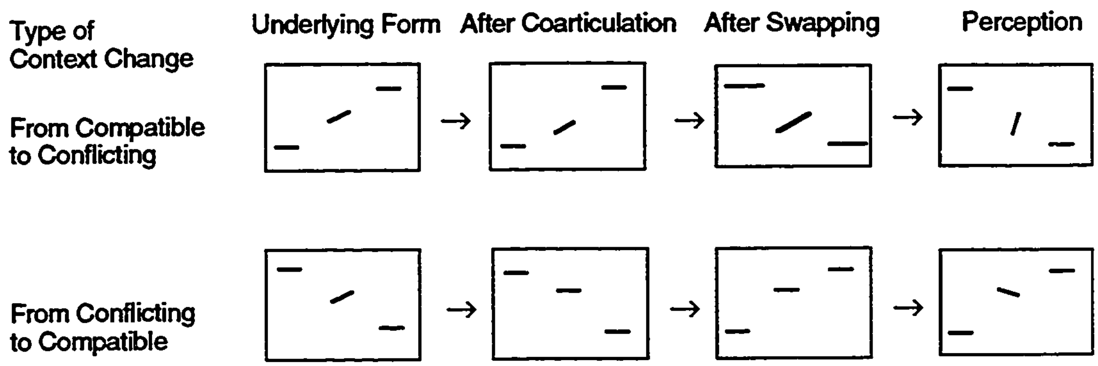


Figure 1.15. Predicted effect of swapping upon perception of coarticulated tones for syllables produced in the conflicting tonal context.

Figure 1.15 presents a schematic prediction for what would happen in the real change condition. The swapping of the first tone and the last not only results in significant changes in tonal environment for the tone of the second syllable, since both the beginning and the ending F0 values of the two tones differ significantly, but also completely alters the original tonal context which is responsible for the F0 variation in the second syllable. If listeners hold the altered context partially responsible for the flattened contour of the target tone, they would compensate for the influence of the altered context as if it had been there originally. The effect of this compensation process is to shift the perceived value of the target tone at a given point away from the value of the adjacent tone. When the original tonal context is compatible (See the upper flow chart), coarticulation of the target tone with the adjacent tones preserves its underlying tonal contour. After swapping, the context becomes conflicting. The compensation process would shift the values of a target tone away from the values of the adjacent tones in the altered context, as mentioned above. This shift would further confirm the current tonal contour in the target tone, which already agrees with the original form. As a result, the perceived tone would just be the original tone. When the original context is conflicting (see the lower flow chart), the underlying dynamic tonal contour of a target tone becomes flattened due to coarticulation with adjacent tones. After swapping the first and last syllables, the tonal context becomes compatible. Again, the compensation process would shift the values of a target tone away from the adjacent values of the tones in the altered context. The consequence of the shifting in this case would be, however, to perceive a tone with an opposite dynamic contour, i.e., hearing a rising tone as a falling tone, and vice versa, or to perceive a tone with a flattened contour, i.e., the high tone.

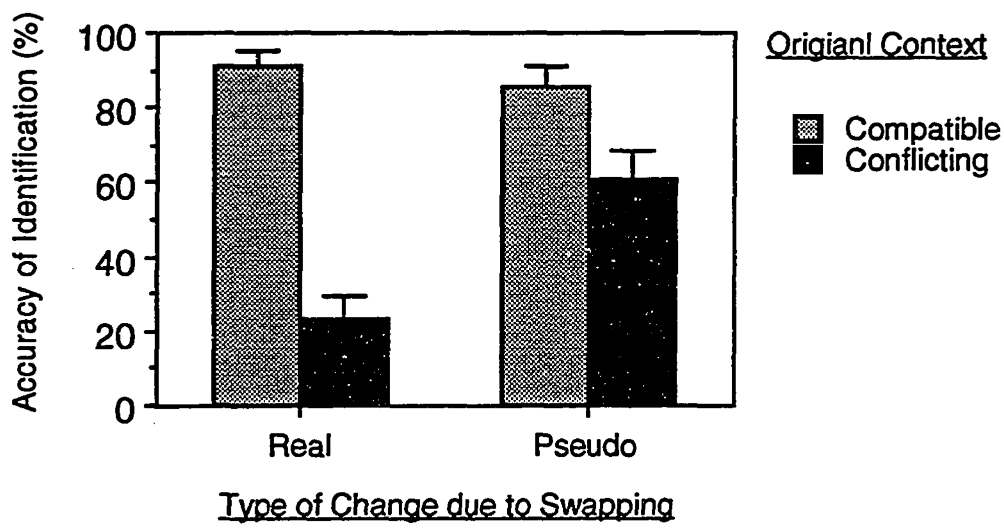


Figure 1.16. Accuracy of tone identification under the influence of original tonal context and type of change due to swapping in Experiment 4. The horizontal bars indicate standard error.

Results

Figure 1.16 displays the accuracy of tone identification under the effect of context and type of change. While both the main effect of context ($F(1,9) = 29.87$, $p < 0.001$) and the effect of change ($F(1,9) = 11.99$, $p < 0.01$) are significant, the most interesting is the significant interaction between the two factors, $F(1,9) = 65.07$, $p < 0.001$. The effect of original context is much greater when swapping produced real change than when swapping produced only pseudo change.

Figures 1.17 and 1.18 break up the results shown in Figure 1.16 into responses for the original rising and falling tones separately. In each of the two figures, we can see the proportion of all the tone responses for the rising or falling tone originally having compatible or conflicting contexts, and having undergone real or pseudo change as a result of swapping operation.

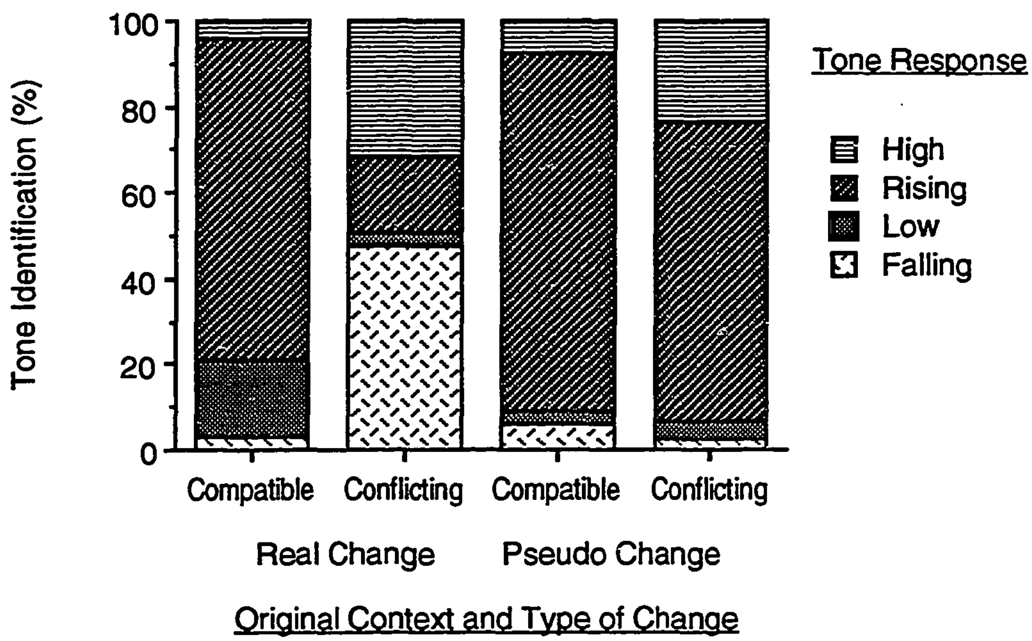


Figure 1.17. Tone responses for the rising tone after swapping in Experiment 4.

In Figure 1.17, it can be seen that under the pseudo change condition, a rising tone was mostly heard as rising tone whether the original context was compatible or conflicting, although as in Experiment 1, the compatible context was a little better than the conflicting context. For the real change condition, as predicted, there were dramatic differences between compatible and conflicting contexts. Most tokens originally in compatible contexts were heard as the rising tone (75%). Furthermore, the 18% of low tone responses should be considered as ambiguous between the low and the rising tone percepts due to the tone sandhi mentioned earlier. With the addition of this portion, the total rising tone responses would be increased to 93%. For the rising tone originally produced in conflicting contexts, the largest proportion of responses, as a matter of fact, almost half (49.7%) of them, were falling tone responses. The second largest proportion, about 31%, were high tone responses. The combined proportion of the falling and high tone responses adds up to 79% of the total number of responses, while the rising tone responses add up to only 18%. This pattern of responses demonstrates that the subject indeed compensated for the influence of the altered context introduced by the swapping operation as if it had been there originally, thus confirming the predictions shown in Figure 1.16.

Figure 1.18 shows tone response patterns for the falling tone. For the pseudo change condition, correct tone identification was 84% for the compatible context and 49% for the conflicting context. The moderate identification score for the conflicting context is probably due to listeners' compensation for the reversed declination of overall F0 contour due to swapping. The swapping operation, though not changing the tonal categories in the tonal context for the second falling tone in the tonal sequences of \ \ \ (which is the tonal sequence having conflicting context and subject to pseudo change by the swapping operation) reverses the F0 declination, which has been found in many languages including Mandarin (Gårding, 1987; Maeda, 1976; Pierrehumbert, 1979; 't Hart & Cohen, 1973), in the original signal. This reversed declination in the stimuli is compounded by listeners' expectation for the normal declination. In other words, the perceived overall pitch of the second falling tone in a tonal sequence of \ \ \ could have been the result of two raises, first by the swapping operation per se, and then by listeners' compensation for the normal declination as if it had been partially responsible for the actual F0 values in the stimuli. This double raising of perceived overall pitch plus the flattened F0 contour may have been the major causes for the high proportion of high tone responses for the second falling tone in the tonal sequence of \ \ \. Yet another possible additional cause is the carryover coarticulation between the second and the third syllable, which have been seen to be more prominent than the anticipatory coarticulation in our F0 analysis described earlier. This carryover coarticulation would lower the starting as well as the overall F0 value of the third syllable due to the low ending value of the falling tone in the second syllable. This would then raise the F0 value of the second syllable relative to that of the first syllable after swapping, thus increasing the number of high tone responses for the second syllable after the swapping operation.

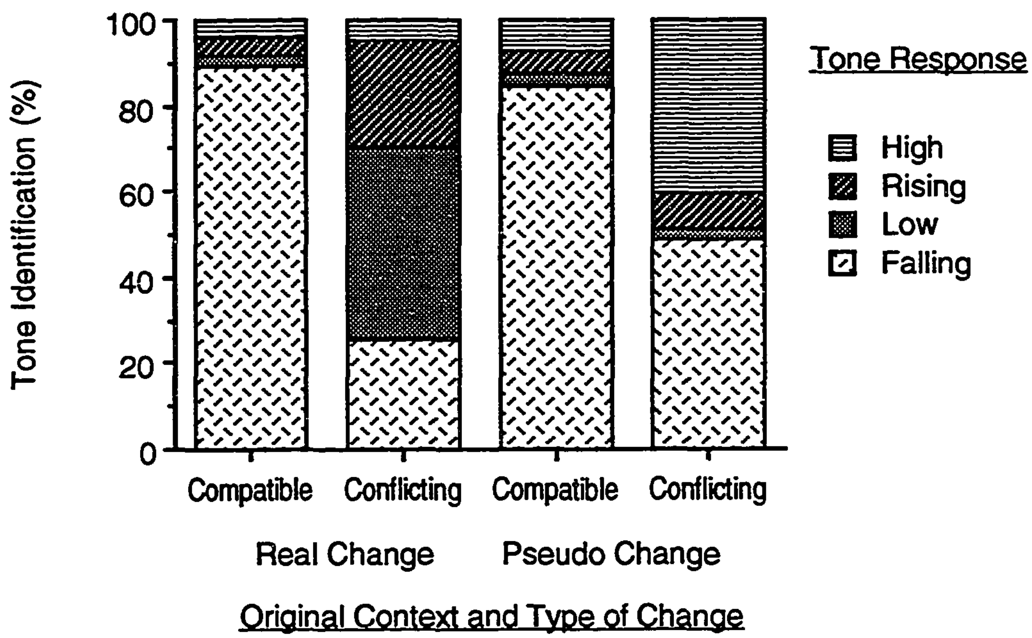


Figure 1.18. Tone responses for the falling tone after swapping in Experiment 4.

For the real change condition, again, there is striking difference between the compatible (89%) and the conflicting (26%) contexts. Also, although only 25% of the tokens were directly heard as rising tone for the original conflicting context, 45% were heard as low tone. Again in this condition, a low tone percept should be considered as nondistinct from a rising tone percept, because the target tone was followed by a low tone which, had it been there originally, might have caused the previous tone to change into a rising tone. So the combined percentage of rising tone percepts would then be 69%. Again, this pattern of responses confirms the prediction that subjects would compensate for the influence of the new context introduced by the swapping operation as if it had been there originally.

Discussion

The results of Experiment 4 clearly demonstrate that perception of tones depends to a great extent on the surrounding tonal context. The subjects' response patterns in this experiment indicate that listeners compensate for the influence of the tonal context by shifting the perceived pitch values of the target tones away from the F0 values of their adjacent tones. The clearest demonstration of this is that in the cases where the F0 contours of the target tone have been flattened by the originally conflicting tonal context during production, a newly introduced compatible context did, as predicted, shift the perceived pitch contours of the target tones further away from its underlying form, resulting in subjects' perceiving tones with the opposite dynamic directions from those of the underlying tones.

These results further indicate that in Experiments 1 and 2, it must have been the presence of the tonal information, not anything else in the context, that helped the subjects to achieve their high accuracy of identification of the target tones.

General Discussion for Chapter 1

At the beginning of this chapter, it was hypothesized that the rising tone variation in trisyllabic sequences is the result of tonal coarticulation. Four predictions based on this hypothesis were also made, as repeated here.

1) Not only does the rising tone undergo perturbation, but also the falling tone is subject to perturbation under comparable conditions;

2) Other things being equal, the magnitude of the contextual tonal perturbation is positively related to speaking rate;

3) Listeners will treat the contextual perturbation as coarticulation, and will hence compensate for it with the help of the tonal context present with the target tone;

4) Listeners' performance in identifying the perturbed tones will not be directly proportional to different magnitude of the perturbation due to different phonetic contexts or speaking rates.

Evidence from acoustic analyses

Prediction (1) states that both the rising and the falling tones are sensitive to contextual variations. By comparing tonal variation of the rising and the falling tones produced in either *compatible* or *conflicting* contexts, it was found that both tones in Mandarin deviated more from their underlying forms when produced in a conflicting context than in a compatible context. More specifically, a conflicting context was found to be able to distort a dynamic tone to the extent that the direction of its contour was sometimes reversed. When the tonal context was conflicting, the rising tone was often produced with a slightly falling contour and other times with nearly level contours, and the falling tone was produced most of the time with a falling contour with very small negative slope, and sometimes with a slightly rising contour. In contrast, when the tonal context was compatible, the rising tone had a clear rising contour and the falling tone had a clear falling contour.

Although the falling tone did not reverse its F0 contour in conflicting context as often as did the rising tone, the difference between the falling tones produced in compatible and conflicting contexts was found to be greater than the same difference for the rising tone, indicating that the falling tone had actually undergone more deviation from its canonical form than the rising tone. It was probably only because the slope of the contour in the canonical form of the falling tone was so steep that even the conflicting context was not enough to reverse its contour direction in most of the cases.

Prediction (2) says that the amount of contextual tonal variations is positively related to speaking rate. This prediction was also confirmed by the results of the acoustical analysis. It was found that the effects of the two kinds of contexts, i.e., compatible and conflicting ones, were greater when the syllables carrying the target tones were produced at a faster speaking rate (with a carrier sentence) than at a slower rate (without carrier sentence). As discussed earlier, both the coproduction model and the target-based model can predict this outcome. Further considerations will be needed in deciding which of those two models or some other model can best account for this particular finding as well as many other findings about coarticulation.

Evidence from perceptual experiments

Predictions (3) and (4) have to do with perception of contextual tone variations. Prediction (4) states that listeners will compensate for contextual perturbation as coarticulation with the help of the tonal context present with the target tone. The four perceptual experiments described above demonstrate that the perception of a tone is heavily dependent on its tonal context. Thanks to the phonetic modification through waveform editing in Experiments 1 and 2, it was possible to examine purely phonetic perception of fluently produced tones. When presented with their original tonal context in those two experiments, most tones were correctly identified even when there was severe distortion of tonal contours due to a conflicting context. When the tones were stripped of their original tonal context in Experiment 3, their identification fell below chance, and the tones perceived were mainly those whose canonical forms resembled the tonal contours actually produced. Since the target tones originally produced in conflicting context mostly had flattened tonal contours, they were mostly perceived as the high tone whose canonical contour is high and level. When the original context of a tone was altered in Experiment 4, subjects compensated for the altered tonal context as if it had been there originally. Most interestingly, when a tone originally produced in a conflicting context was presented in a compatible context, instead of helping to identify the underlying tone, the wrong context caused the flattened tone to be heard more often than not as a tone with the opposite dynamic direction.

These results of the perceptual experiments agree with previous findings about perception of coarticulated segmental phonemes. It has been shown that perception of one phoneme depends on the adjacent phonemes (Cooper, et al., 1952; Liberman, et al., 1952; Mann, 1980; Mann & Repp, 1980; Fowler, 1984; Fowler & Smith, 1986), indicating that listeners take coarticulation between adjacent phonemes into consideration when perceiving speech. In this study, listeners were found to parse tonal variation along coarticulatory lines: referring to contextual tonal information when interpreting a given stretch of surface tonal contour. As a result, variation in the surface contour is treated depending on the particular tonal context that surrounds it. When the surrounding context provides correct coarticulatory information in Experiments 1 and 2, the intended tones were mostly correctly identified. When the tonal context is different from the original one in Experiment 4, compensating for the newly introduced coarticulatory information resulted in much variation depending on the particular new relation between the target tone and the new tonal context.

Prediction (4) says that listeners' performance in identifying the perturbed tones will not be directly proportional to the magnitude of the perturbation. The F0 analyses found that the amount of tonal perturbation varied depending both on the type of tonal context and on speaking rate. In Perceptual Experiment 1, however, the difference in tone identification between faster and slower speaking rates was found to be insignificant. The perceptual difference between the compatible and conflicting contexts in Experiments 1 and 2, although significant, is far out of proportion to the acoustical difference found in the F0 analysis. As a contrast, the amount of perceptual difference between the two types of contexts in Experiment 3, where the target tones were presented without the original contexts, was much more comparable to the actual difference in F0. Therefore, prediction (4) is largely confirmed.

The confirmation of predictions (1) through (4) naturally leads to support for the coarticulation hypothesis: that the rising tone variation in trisyllabic sequence described by Chao (1948, 1968) is coarticulatory in nature. Chao's rule states that a rising tone changes into a high tone when preceded by a high or rising tone and followed by any tone other than the neutral tone. Although described by Chao as only applicable to conversational speech, this rule is generally taken by Chinese linguists as a phonological tone sandhi rule similar to the low tone sandhi rule by which a low tone changes into a rising tone when followed by another low tone. However, as mentioned earlier, the low tone sandhi has been shown to change a low tone into a tone that is perceptually indistinguishable from the rising tone (Wang & Li, 1967). Perceptual Experiments 1 and 2 showed that most of the rising tones produced with flattened F0 contours in conflicting context were still perceived as rising tone. This indicates that the intended tones as recovered by the listeners were still the underlying tones. In other words, as heard by the listeners, the speaker did not produce a tone that is different from the underlying one despite the distorted surface value due to conflicting context.

Other evidence

Other support for the coarticulation hypothesis comes from a recent study by Shih and Sproat (1992), who argue that the rising tone variation described by Chao is the result of phonetic implementation. On the one hand, they find that, when preceded and followed by the high tone (a condition also included in Chao's rule but not examined in the present study), the rising tone still has different F0 contours from the high tone. On the other hand, because of their examination of the rising tone variation in both trisyllabic and quadrasyllabic sequences, they were able to observe that the amount of pitch drop to the low target of a rising tone "varies with the prosodic strength of the syllable. As a result, the F0 contour of an extremely weak rising tone in high tone context approaches the shape of a high level tone (p. 193)."

The relationship between the rising tone variation and stress found by Shih and Sproat is comparable to the relationship between the rising tone variation and speaking rate found in this study. Both show that the contextual variation described by Chao is gradient rather than categorical. So, the results of both Shih and Sproat (1992) and of the present study demonstrate that the contextual rising tone variation is due to coarticulation rather than a phonological process that changes the tone category from one to another.

Further Discussion

Anticipatory vs. carryover effects

It is of special interest to notice that in Chao's rule (see page 15), the ending pitch of the tone preceding the rising tone must be high, whereas the following tone may have any starting pitch. What this implies is that if the variation is due to tonal coarticulation, there must be more carryover than anticipatory coarticulation. In Shih and Sproat (1992), when both the preceding and the following pitch is high, the mean F0 height of the rising tone in the middle approaches that of the following high tone in conversational speech. In the present study, the rising tone preceded by a high pitch and followed by a low or mid pitch had only slightly negative slope. In addition, its mean F0 is raised by the preceding high ending pitch, but not lowered by the following low starting pitch.

These findings seem to disagree with the conclusion by Shen (1990) that there is a general symmetry in Mandarin tonal coarticulation. Two of the possible sources of the disagreement are differences in the reading lists and in the instructions to the speakers. In Shen's study, nonsense strings of syllables were used as reading list, whereas in the present study as well as in Shih and Sproat (1992), real words and phrases were used. Also in Shen's study, speakers were instructed to stress the three syllables in each utterance evenly, whereas in the present study speakers were instructed to produce the utterances as naturally as possible. Shih and Sproat examined different styles of speech, and the greatest perturbation was found in the conversational style. Other sources of disagreement will be further discussed in Chapter 2.

Perceptual limit of coarticulation

In the perceptual Experiments 1 and 2, although nearly all the target tones had fairly high identifiability, significant differences were found between the same tone in a compatible context and in a conflicting context, with tones in the compatible context being identified better than those in the conflicting context. This finding was possible because efforts were made in this study to examine perception of phonetic units with different amounts of coarticulatory variation. Another study that examined perception of different amounts of coarticulation is Wright and Kerswill (1989). In that study, they looked at the perceptual effect of different degrees of velar assimilation of the English /d/ followed by a velar stop. Their data indicate that even with the original context, tokens of /d/ produced with less alveolar contact (determined palatographically) were identified as alveolar at a lower percentage than tokens with more alveolar contact. That result is similar to the findings of this study in that both show that an underlying phonetic unit is not always fully recoverable when the surface form deviates too much from the canonical form due to coarticulation.

It is not yet quite clear either from the results of the present study or from those of Wright and Kerswill (1989), however, whether the less than perfect recovery of an underlying form is due to the listener's incomplete recovery of the speaker's intention or rather due to the speaker's somewhat ambiguous production of the intended form. In other words, either the speaker has intended to fully realize an underlying form by means of coarticulation, and it is the listener who sometimes fails to fully recover the coarticulated underlying form; or it is the speaker who has reduced his effort in trying to fully realize the underlying form in fluent speech, especially when the phonetic context for that particular underlying form is conflicting, and the listener's seemingly imperfect recovery is actually a full recovery of a form that has been intended to be less clear for some reason, because listeners actually perceive what talkers do (Fowler, 1987). In either case, the limit of purely phonetic perception needs to be recognized. It seems that listener's ability to perceive fluent speech (even as casual as the particular utterance of the word *advertisement* cited by Stevens (Catford, Jusczyk, Klatt, Liberman, Remez, & Stevens, 1991), cannot be explained fully in terms of bottom-up processes of recovering either the phonetic segments or the articulatory gestures.

Chapter 2 — Asymmetry in Contextual Tonal Variation

Background

Chapter 1 found that the contextual tonal variation in the tone of the second syllable in a trisyllabic Mandarin word or phrase was mostly coarticulatory. More specifically, the tonal variation was such that the F0 values of that syllable approximated those of the neighboring tones. In other words, the tonal variation found there was largely assimilatory.

It was also found in Chapter 1, however, that the tone of the second syllable assimilated more to the preceding tone than to the following tone. In the case of the rising tone, it is even hard to say if there was any assimilation to the onset pitch of the following tone at all (cf. Figure 1.3).

Also as discussed in Chapter 1, Chao's (1948, 1968) rising tone sandhi rule (that a rising tone becomes a high tone when preceded by a high or rising tone and followed by any other stressed tones) seems to indicate that the rising tone is influenced more by the preceding tone than by the following tone, because the rule specifies that the preceding tone must be either the high or the rising tone, both of which have high ending pitch, but the following tone can be any of the four lexical tones. The acoustic aspect of Chao's rule was largely supported by Chapter 1 of this study as well as by Shih and Sproat (1992). Both studies found the rising tone to have a near flat F0 contour in a trisyllabic sequence when produced in highly fluent speech. However, the rising tone in the present study was followed by a low starting pitch, whereas in Shih and Sproat (1992) it was followed by a high starting pitch. Therefore, the following tone seems to have played only a minor role as far as the assimilation is concerned.

Those findings seem to be consistent with the patterns of contextual tonal variation in Vietnamese described by Han and Kim (1974), who examined tonal variation in disyllabic Vietnamese words and concluded that tonal variation was greater in the second syllable than in the first syllable.

An extensive study of tonal coarticulation in Mandarin by Shen (1990), however, did not find the same asymmetry. After examining both anticipatory and carryover coarticulation in trisyllabic sequences in Mandarin, she concluded that "Mandarin tones are affected by both carryover and anticipatory coarticulations, and the bi-directional effects are symmetric (p. 293)."

To make the picture more complicated, there has been evidence that the contextual effect on tones may not always be assimilatory. Shih (1986, p. 10, question mark in the original), for example, noticed:

> A tone 4 is higher (310-238) when tone 3 follows. The final H of tone 2 is also higher (281) in the same environment. It seems that a (prosodically strong?) tone 3 raises a H target in the preceding tone. The value of the final H of tone 1 is quite high, but not higher than that in other contexts. It is not clear whether it is under the influence of a following tone 3.

Gandour (1992b, p. 121) also observes that in Thai,

> The F0 contour of the preceding falling tone is generally higher and steeper when occurring before the low and rising tones, both of which have a low F0 onset. But when occurring before the falling tone, which has a high F0 onset, it is lower and shallower. ... Instead of a lowering effect, the low and rising tones induced a raising of F0 in the preceding falling tone.

Even Shen (1990), who concluded that contextual tonal effect in Mandarin is symmetrical, also noticed that "Tones 1 and 2 have the highest overall tonal values when preceding the two mid-onset tones, that is, Tones 2 and 3 (p. 285)."

In a study of Beijing Mandarin stage speech, Kratochvil (1984) observes that in sequences of rising tone + falling tone and falling tone + falling tone, the F0 range of the first syllable becomes narrower than the average F0 range of the particular tone group it belongs to, whereas in all the other cases, the F0 range of the first syllable becomes wider than the average F0 range of their respective tone groups. In other words, a falling tone, which has a high starting pitch, reduces the F0 range of the preceding rising or falling tone.

It thus seems likely that contextual tonal variation is more complicated than one would think. It probably involves both assimilation and dissimilation. In this chapter, acoustic analyses of disyllabic tonal sequences in Mandarin are conducted, and both anticipatory and carryover tonal variations are examined.

Another question that concerns this chapter has to do with the scope of tonal coarticulation. While segmental coarticulation has been known to extend over several segments, tonal coarticulation has never been clearly shown to extend beyond the adjacent syllables. In fact, it has been claimed that tonal coarticulation is strictly limited to the immediately adjacent tones in Mandarin (M. Lin & Yan, 1991; Shen, 1990). In both studies, however, the contextual effects on all the three (Shen, 1990) or four (M. Lin & Yan, 1991) syllables in the target sequences were treated collectively rather than separately in their analyses. It is possible that this method might have obscured the picture. To further explore the possibility of cross-syllabic tonal influence, tones in different prosodic positions are examined separately to assess their variation due to cross-syllabic tonal influences.

F0 Measurement

Material

Two main reading lists were used for production of the Mandarin tones. List A consists of the syllable /ma/ with the four Mandarin lexical tones:

List A

Pinyin:	mā	má	mǎ	mà
Character:	妈	麻	马	骂
Gloss:	'mother'	'hemp'	'horse'	'scold'

List B consists of sixteen disyllabic sequences of /mama/ that make up all the possible bitonal combinations of the four lexical tones of Mandarin. Among the sequences in List B, only 妈妈, 'mother', is a word in Mandarin. The others are either totally nonsense combinations, or sequences that, though could be considered grammatical phrases, are unnatural combinations in Mandarin. Unlike the real words used in Chapter 1, the disyllabic sequences used here would probably not induce maximal coarticulation, but they do provide systematic comparisons among different tonal combinations.

List B

māmā 妈妈	māmá 妈麻	māmǎ 妈马	māmà 妈骂
mámā 麻妈	mámá 麻麻	mámǎ 麻马	mámà 麻骂
mǎmā 马妈	mǎmá 马麻	mǎmǎ 马马	mǎmà 马骂
màmā 骂妈	màmá 骂麻	màmǎ 骂马	màmà 骂骂

The initial nasal consonant /m/ was used to insure an unbroken F0 contour throughout each syllable. At the same time, nasal-vowel boundaries are relatively easy to determine in the wave form, since there is abrupt change of wave form shapes as well as amplitude levels at the boundary.

The monosyllables were to be produced in isolation, and the /mama/ sequences were to be produced in different carrier sentences as well as in isolation. The four carrier sentences are as follows:

Carrier 1:	我教________联络	'I teach ___ to communicate.'
	Wǒ jiaō ___ liańluò.	
Carrier 2:	我教________练习	'I teach ___ to practice.'
	Wǒ jiaō ___ liaǹxí.	
Carrier 3:	我叫________联络	'I tell ___ to communicate.'
	Wǒ jiaò ___ liańluò.	
Carrier 4:	我叫________练习	'I tell ___ to practice.'
	Wǒ jiaò ___ liaǹxí.	

Notice that these carriers were designed in such a way that their phonetic as well as syntactic structures are very similar, except for their tones. There were two different pre-target syllables, /jiaō/ and /jiaò/, thus the tone before the target unit has either a high ending pitch or a low ending pitch. Likewise, the tones on the two post-target syllables start either at a low pitch, as in /liań/, or at a high pitch, as in /liaǹ/ and the rest of the syllable structures in the two syllables are the same. Except for the word 妈妈, 'mother', which makes perfect sense in all the four carrier sentences, none of the disyllabic combinations makes much sense in the carrier sentences.

Recording

Four native speakers of Mandarin, all of them male, produced those sentences. Three of the speakers were born and raised in Beijing, and so were native speakers of Beijing Mandarin. The fourth speaker (myself) was a native speaker of Standard Chinese, the official dialect that has the same phonetic system as Beijing Mandarin.

The /mama/ sequences were printed in Chinese in two different orders: a) the same order as in the list shown above; b) the reverse order. The carrier sentences were printed in Chinese on each of the two reading sheets below those /mama/ sequences.

A pre-recorded pacing tape was used to control the speaking rate of the speakers. On the tape were groups of six beeps with intervals of three seconds. The beginning of each group was signaled by a double beep, and the end of each group by an extra long beep.

The speakers first produced the monosyllable /ma/ with the four lexical tones in isolation. They then produced all the 16 /mama/ sequences without any carrier. Finally, they produced all the /mama/ sequences in each of the four carrier sentences. The speakers repeated each item six times, each repetition following a beep played from the pacing tape. The speakers were asked to make all their productions as natural as possible. As a result, they produced both syllables in each /mama/ sequence with roughly the same stress. However, for the word *māmā*, 'mother', they had to be told to say it with the same even stress pattern as the other sequences instead of the usual trochee stress pattern for the word for "mother." It turned out that they had no difficulty in following this instruction.

F0 extraction

The utterances were digitized with the Haskins PCM system (Whalen, et al., 1990) at a sampling rate of 20 KHz. F0 extraction was carried out manually. The wave form of an utterance was displayed on the computer screen, and was stretched so that each glottal cycle is shown clearly. The onset of each glottal cycle in the wave form of a /mama/ sequence (no measurements were obtained for the carrier sentences) was marked with a label. The distances between adjacent labels were recorded in a file. Those files were processed by a separate computer program written for the purpose. The program transforms the intervals between successive labels into F0 values; and the F0 curves obtained were smoothed using a simple window function incorporated in the program that eliminated any bumps or sharp edges in the F0 contour greater than 2 Hz.

Each F0 curve was then time-normalized, with the same program, within each of the four segments in the /mama/ sequence, namely, the first nasal, the first vowel, the second nasal, and the second vowel. Ten F0 frames were obtained for each of the two nasal segments; and 20 F0 frames were obtained for each of the two vocalic segments.

Figure 2.1 shows the F0 contour for the syllable /ma/ in four different tones averaged over all the tokens produced by the four speakers. In this figure and in all the subsequent figures of F0 tracing, the ordinate is F0 value in Hertz and the abscissa is frame numbers. The distances between frames are in normalized time. A nasal segment in a syllable consists of 10 frames, and a vocalic segment consists of 20 frames.

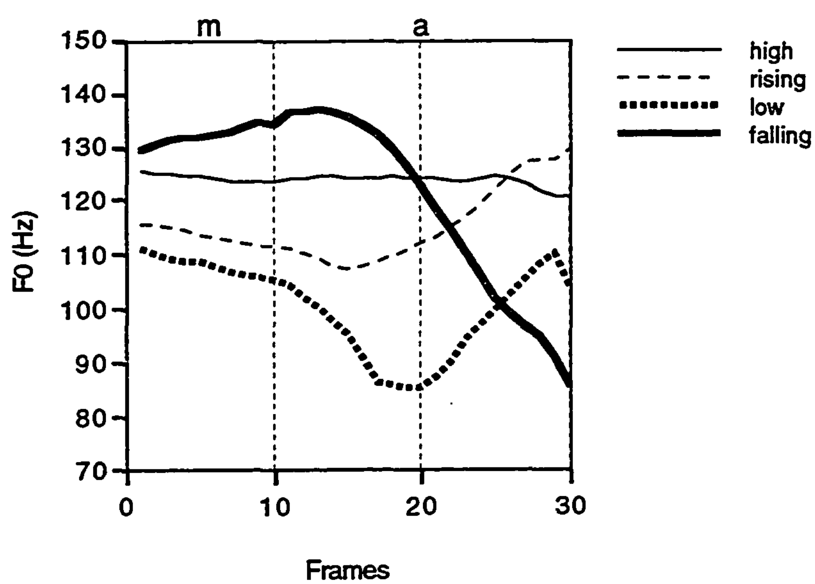


Figure 2.1. Mean F0 contours of four Mandarin tones in the syllable /ma/ produced in isolation.

From Figure 2.1, it can be seen that the high tone and the falling tone start with a higher F0 value than the rising tone and the low tone. It is interesting to notice that during the initial nasal segment, although each of the four tones has a different starting value, all of them, especially the extreme ones, are pointing back to some central value in the F0 range. This is probably one form of coarticulation, namely, the coarticulation between the rest position and the first pitch target in a tone. It is also interesting to notice that the contours do not reach their maximum or minimum values until some time after the beginning of the vocalic segment, and, it seems that the higher the starting pitch of a tone, the earlier the first F0 extreme is reached. For the falling tone, the first extreme is reached after about a fifth of the duration of the vocalic segment. The rising tone reaches its first extreme after the first quarter of the vocalic segment. For the low tone, the extreme is not reached until the middle of the vocalic segment. Presumably, this probably indicates that a lower pitch register is more difficult to reach than a higher pitch register.

Carryover Effects

Isolated disyllabic sequences

Figures 2.2 and 2.3 shows variations of F0 contours in the tones of the second syllable due to the influence of the preceding tones in the /mama/ sequences produced in isolation. Each panel in those figures plots the same tone in the second syllable being preceded by four different tones. To bring out the effects more clearly, when the tone of the first syllable has a high ending value, the curve is drawn with a thin line; when it has a low ending value, the curve is drawn with a thick line.

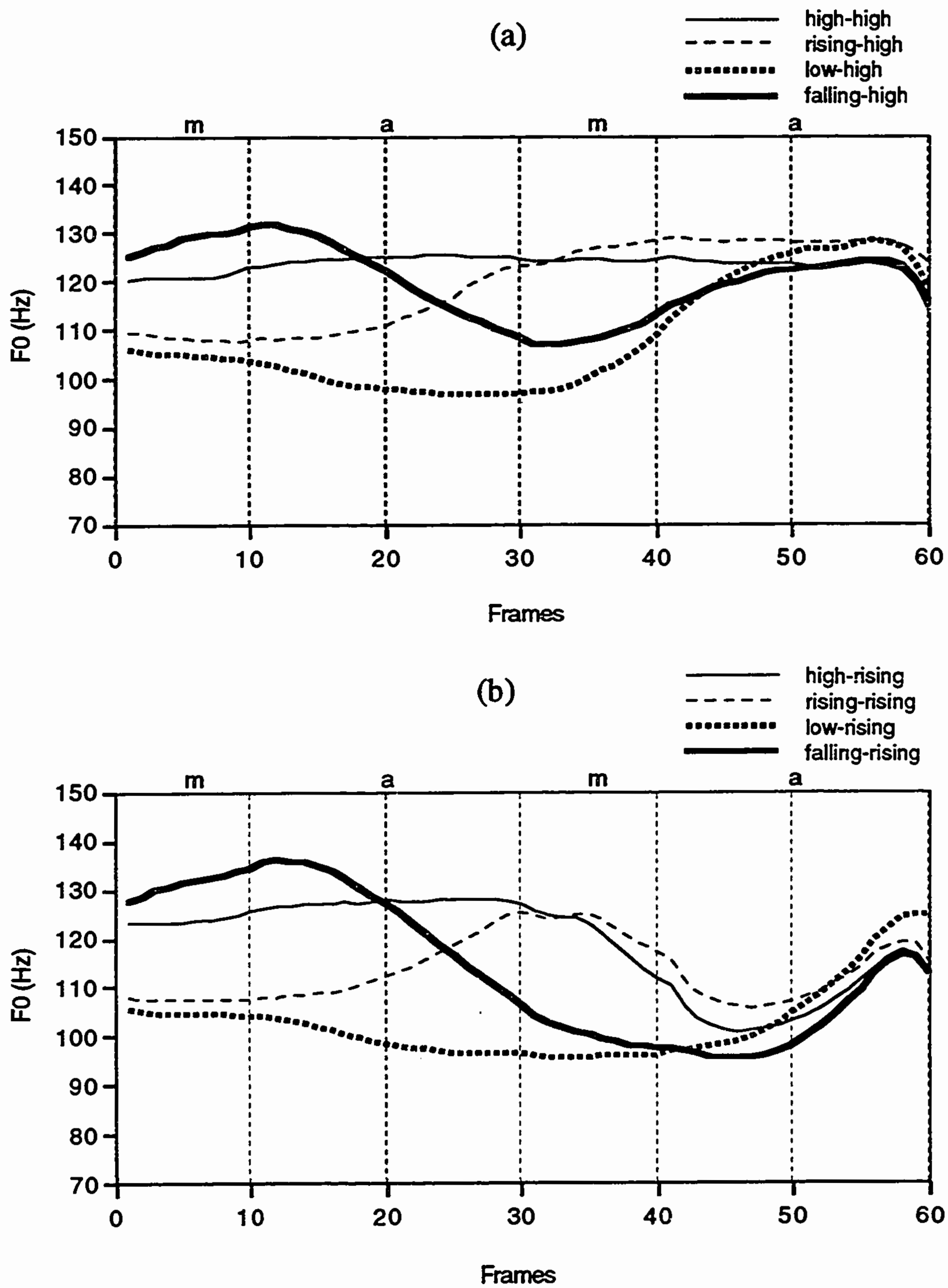
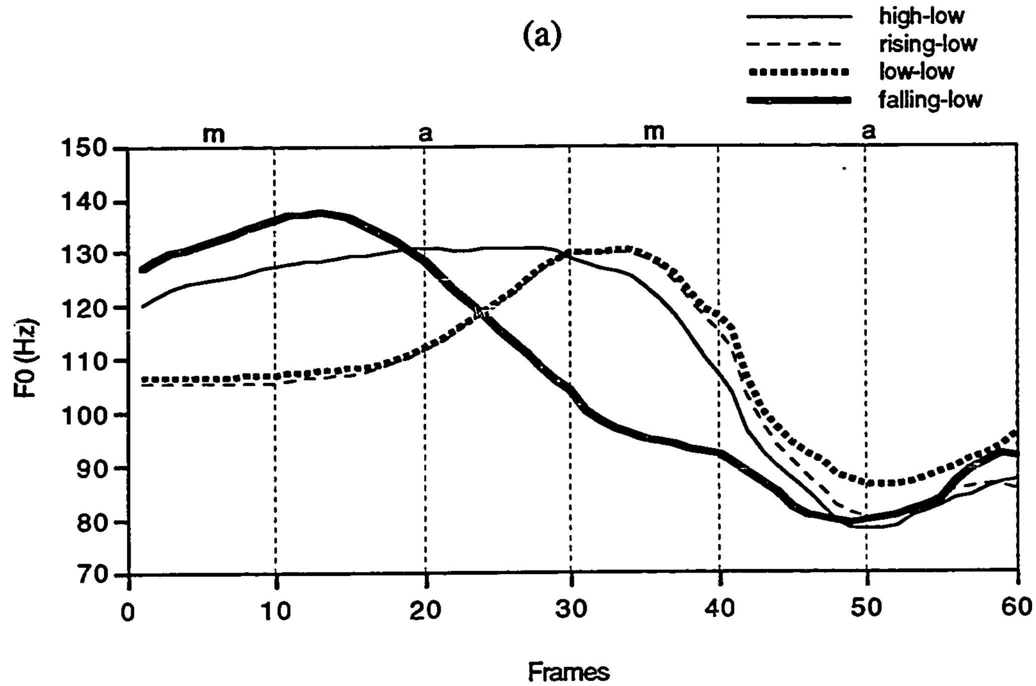


Figure 2.2. Mandarin high (a) and rising (b) tones under carryover tonal influence in the /mama/ sequences produced in isolation.



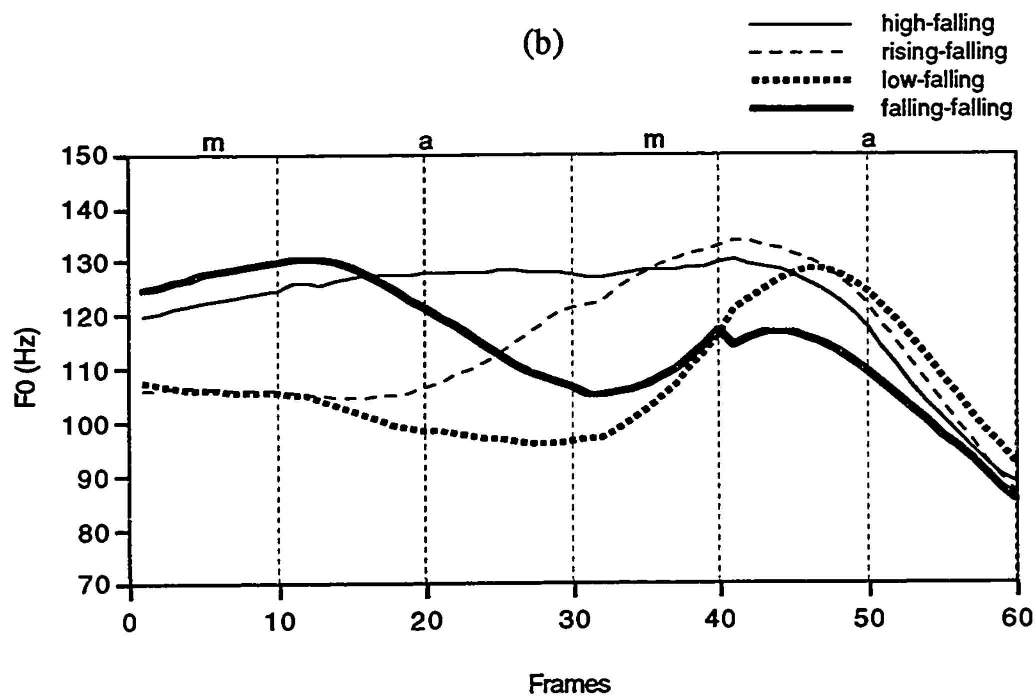


Figure 2.3. Mandarin low (a) and falling (b) tones under carryover tonal influence in the /mama/ sequences produced in isolation.

In Figures 2.2 and 2.3, the carryover effect can be clearly seen. The effect is apparently assimilatory. The starting F0 value of the second syllable is higher when the preceding tone has a high ending pitch (high or rising tone), but lower when the preceding tone has a low ending pitch (low or falling tone). This difference is rather large during the initial nasal consonant of the second syllable, and the difference persists for about a quarter of the vocalic segment of the second syllable in most of the cases. The only exception is when a low tone is preceded by another low tone. Because of the phonological sandhi discussed in Chapter 1, the first low tone assumes a F0 contour that resembles that of a rising tone.

The rising tone seems to have a little more raising power than the high tone in all the cases. Although consistent with Shen's (1990) observation that the overall F0 value of a tone is higher when following the rising tone than following the high tone, this phenomenon somehow disappears when the /mama/ sequences are produced in the carrier sentences, as will be shown later.

It might also be noticed that tones after a low tone do not always have the lowest maximum F0. The high tone following a low tone has almost the same maximum F0 as the high tone following another high tone. The rising tone following a low tone even has the highest maximum F0 among the four tones. However, for some reason, this phenomenon was not seen again for the /mama/ sequences produced with carrier sentences.

Statistical analysis for isolated disyllabic sequences

For each of the four segments, namely, the first and second nasal segments, and the first and second vocalic segments, four F0 measurements were taken for further analysis: (1) Maxf0 — maximum F0; (2) Meanf0 — mean of all the F0 measurements in a segment; (3) Minf0 — minimum F0 in a segment; and (4) Diff0 — difference between the maximum and minimum F0 in a segment.

To test the statistical significance of the carryover effects, separate ANOVAs were conducted with the above four F0 measurements as dependent variables. The independent variables are, (1) Tone — the four Mandarin tones: high, rising, low, and falling; and (2) Pre-pitch — the ending pitch of the first syllable: H (high) for the high and the rising tones, and L (low) for the low and the falling tones. Specifically, what is being tested here is whether the F0 value of the second syllable is raised by a high ending pitch but lowered by a low ending pitch of the first syllable.

Table 2.1. ANOVA Results for carryover effects on F0 of the second nasal and second vocalic segments in the /mama/ sequences produced in isolation.

		Maxf0		Meanf0		Minf0		Diff0	
Segment	**Effect**	**F**	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Nasal	Pre-pitch	330.76	***.0001***	443.06	***.0001***	263.83	***.0001***	.38	.5455
	Tone	2.81	.0609	2.33	.0995	4.54	**.0117**	5.40	***.0055***
	Pre-pitch x Tone	8.45	***.0005***	1.77	.1792	1.0	.4107	10.78	***.0001***
Vocalic	Pre-pitch	6.48	**.0178**	3.13	.0896	2.96	.0985	2.13	.1573
	Tone	8.76	***.0004***	36.57	***.0001***	26.81	***.0001***	5.38	***.0056***
	Pre-pitch x Tone	3.13	**.0444**	.047	.9860	2.09	.1276	12.27	***.0001***

Total df = 24

Because of the special behavior of the low tone before another low tone due to phonological sandhi, the sequence of low-low is not included in the analyses described below and in later sections. To offset the imbalance created by the exclusion of the low-low sequence, the sequences high-low and rising-low are merged by taking the mean of the two for each of the F0 measurements.

Table 2.1 shows the ANOVA results for the nasal and vocalic segments. In this and all the subsequent tables of ANOVA results, effects significant at the 0.05 level are printed in bold face, and effects significant at the 0.01 level are printed both in bold face and in italic.

In Table 2.1, for the nasal segment, the effect of Pre-pitch is significant for all the F0 measurements except Diff0. All the significant effects are assimilatory, i.e., a H Pre-pitch raises the F0 measurement, and a L Pre-pitch lowers the F0 measurement. For the vocalic segment, however, only the effect on the Maxf0 is significant. The effect of Tone is significant in most of the cases, which is not surprising since it is already well known that the four Mandarin tones have distinct pitch contours.

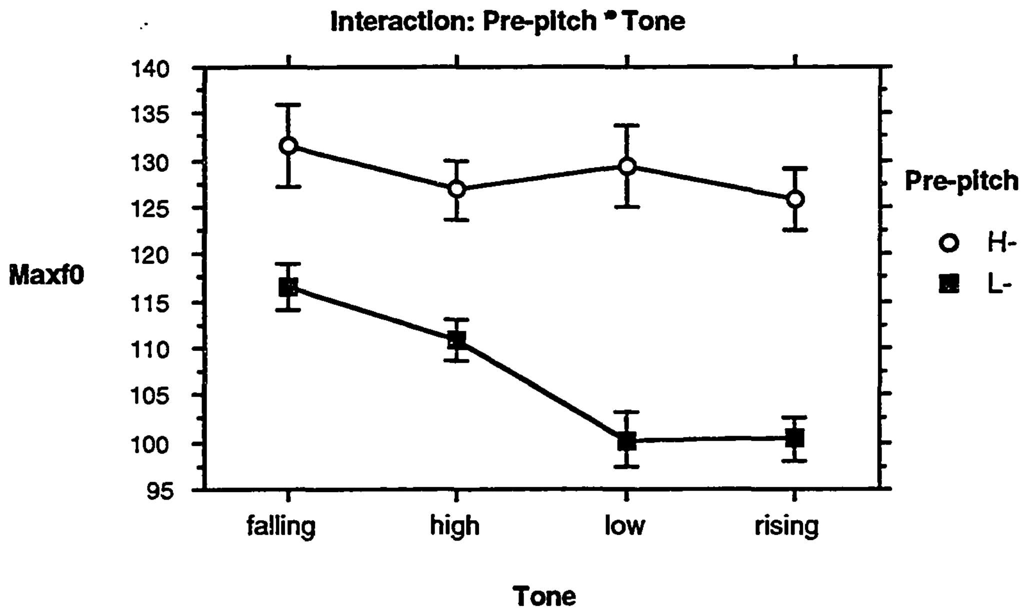


Figure 2.4. Interaction between Pre-pitch and Tone on Maxf0 for the second nasal segment in the /mama/ sequences. Here and in all the subsequent figures for carryover effects, the hyphen in the legend text indicates context, e.g., 'H-' means that the tone under examination is preceded by a high pitch.

The interaction between Pre-pitch and Tone is significant on Maxf0 and Diff0 for the nasal segment. Figures 2.4 and 2.5 show the interaction plot for the three cases. For Maxf0 (Figure 2.4), the interaction is apparently due to greater difference between the H and L Contexts for the low and rising tones than for the other two tones. And this greater difference is mainly because for the low and rising tones (cf. Figures 2.2 (b) and 2.3 (a)), the Maxf0 is taken from the beginning of the nasal segment, where the influence from the preceding tone is the greatest, whereas for the other two tones, the Maxf0 is taken from the end of the nasal segment, where the influence from the previous tone has already been reduced. For Diff0 (Figure 2.5), the interaction seems to be due to the reversed relation between Pre-pitch and Tone. For the high and falling tones, when the Pre-pitch is H, the Diff0 is smaller than when the Pre-pitch is L; for the rising and the low tones, the relation is just the reverse. This is easy to comprehend because in the former case, the high and the falling tones both have high starting pitch, so that when preceded by a tone with high ending value, there should not be much movement at the junction of the two tones. When those two tones are preceded by tones with lower ending pitch, it is necessary to glide from low to high, thus resulting in much F0 movement. For the cases of the rising and low tones, an analogous explanation applies. When there is discrepancy between the pitch values at the juncture of two tones, F0 movement would result.

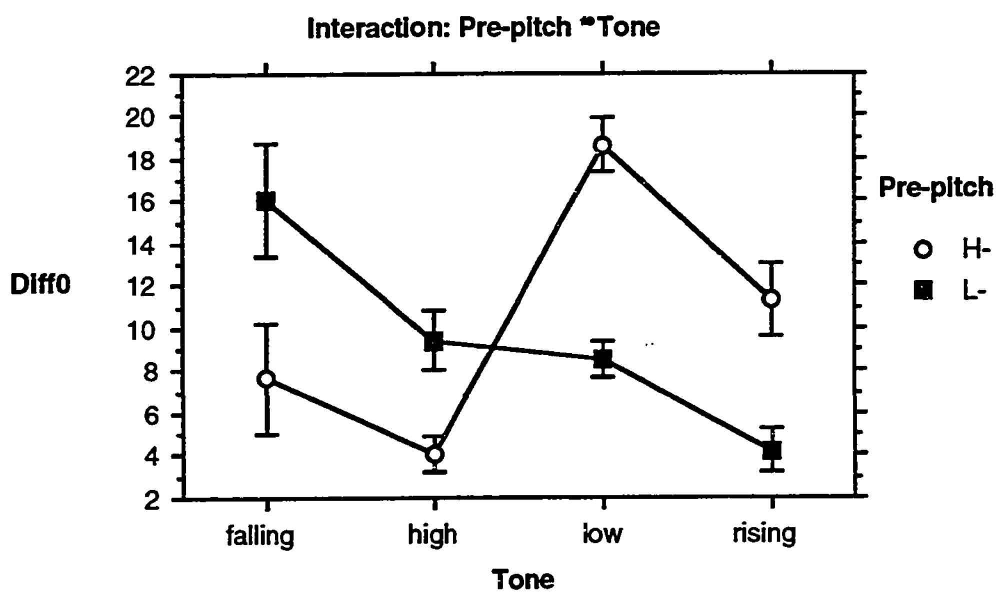


Figure 2.5. Interaction between Pre-pitch and Tone on Diff0 for the second nasal segment in the /mama/ sequences.

For the vocalic segment, there is significant interaction between Tone and Pre-pitch on Maxf0 and Diff0. The interaction on Maxf0 (Figure 2.6) is due to the greater carryover effect on the beginning of the vocalic segment for the low tone, where the Maxf0 is taken, than for other tones whose Maxf0 is taken later in the vocalic segment. The interaction on Diff0 (Figure 2.7) is mainly due to the difference in Diff0 between the low and the rising tones. As shown in Figures 2.2 (b) and 2.3 (a), the Diff0 for the low tone reflects the movement from the beginning of the vocalic segment to the lowest F0 near the middle of the vocalic segment. This movement is greater when the Pre-pitch is H than when it is L. The Diff0 for the rising tone, on the other hand, reflects the movement from the lowest F0, which is located between the first and the second quarters of the vocalic segment, to the highest F0 near the end of the vocalic segment. This movement, in contrast to that in the low tone, is greater when the Pre-pitch is L than when it is H. In both cases, the differences reflect the assimilatory influence of the preceding tones.

Disyllabic sequences produced in carrier sentences

Figures 2.8 and 2.9 show variations of F0 contours in the tones of the second syllable due to the influence of the preceding tones in the /mama/ sequences produced with carrier sentences. The carryover effect in this context is even more apparent than in isolation (Figures 2.2 and 2.3). Three observations can be easily made from those two figures: 1) In the first syllable, different tones have distinct F0 values till the very end of the first syllable. 2) At the boundary between the two syllables, the starting F0 value of a given tone in the second syllable vary enormously depending on the identity of the tone in the first syllable. 3) The difference due to preceding syllable decreases rather gradually over time: It remains quite large at the onset of the vowel in the second syllable; the differences can still be clearly seen even at the end of the second syllable for the high and the rising tones; because of this gradual reduction of carryover effect, the entire F0 contour for the high tone appears to have a rising contour when the preceding tone is either the low or the falling tone, both of which have a low ending pitch.

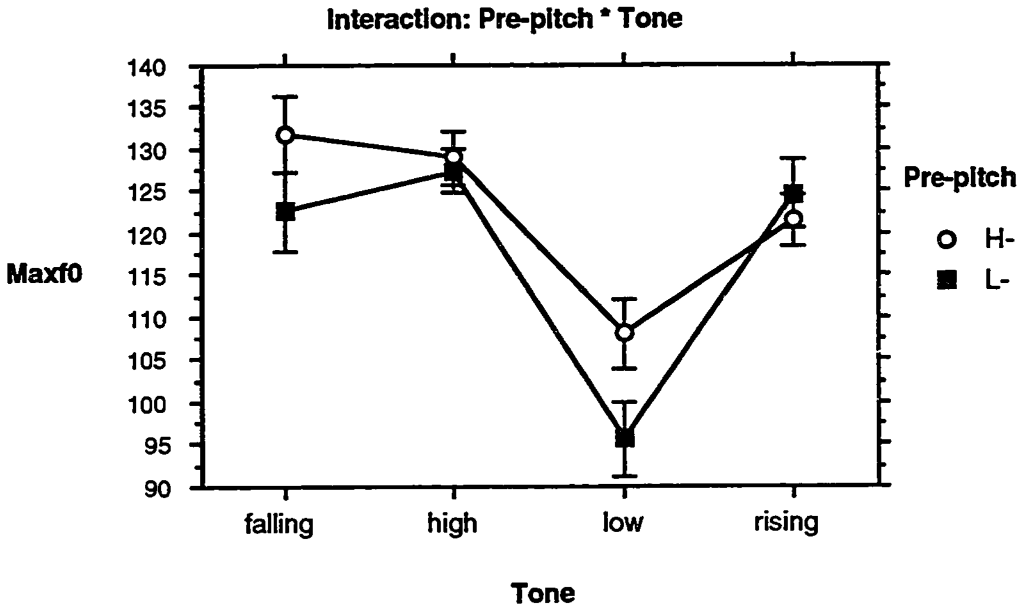


Figure 2.6. Interaction between Pre-pitch and Tone on Maxf0 for the second vocalic segment in the /mama/ sequences.

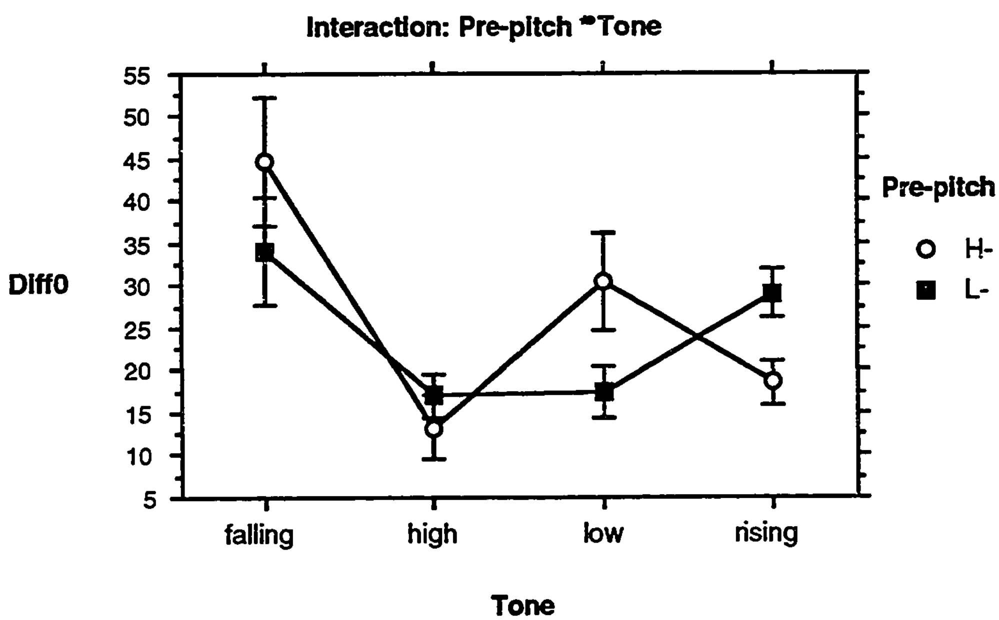


Figure 2.7. Interaction between Pre-pitch and Tone on Diff0 for the second vocalic segment in the /mama/ sequences.

for the high tone appears to have a rising contour when the preceding tone is either the low or the falling tone, both of which have a low ending pitch.

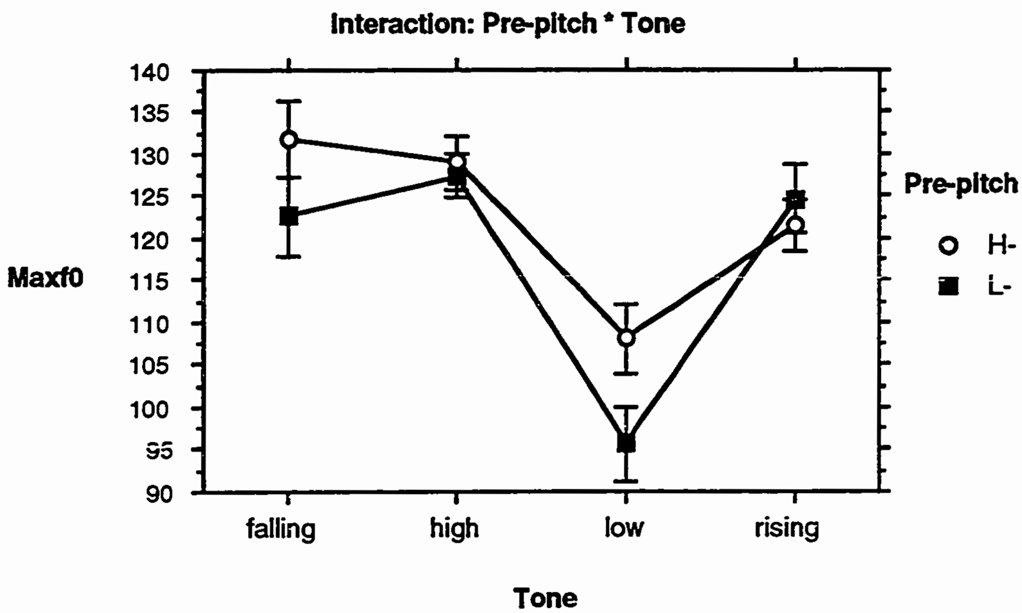


Figure 2.6. Interaction between Pre-pitch and Tone on Maxf0 for the second vocalic segment in the /mama/ sequences.

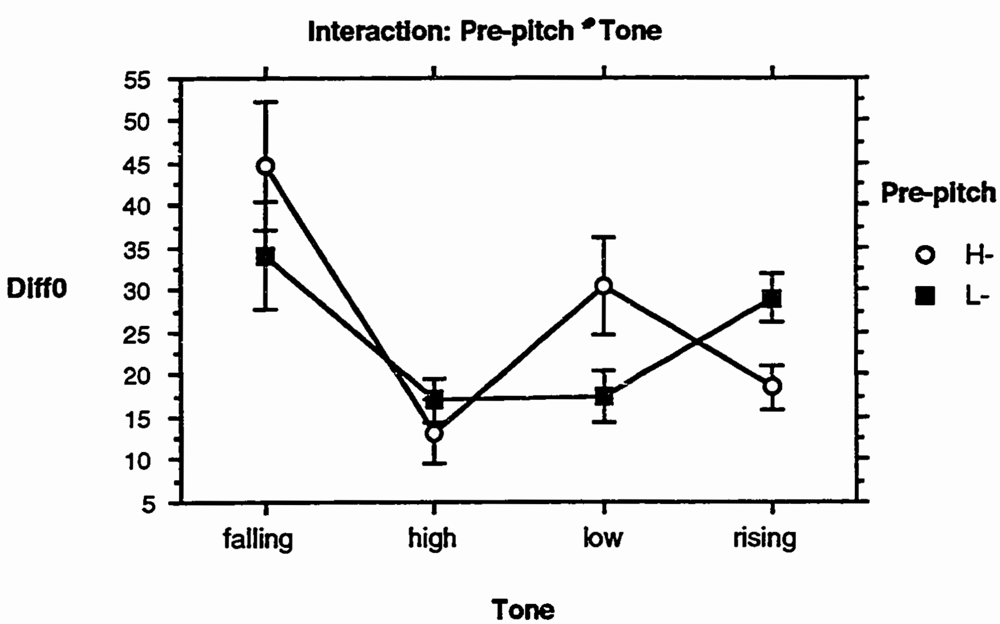


Figure 2.7. Interaction between Pre-pitch and Tone on Diff0 for the second vocalic segment in the /mama/ sequences.

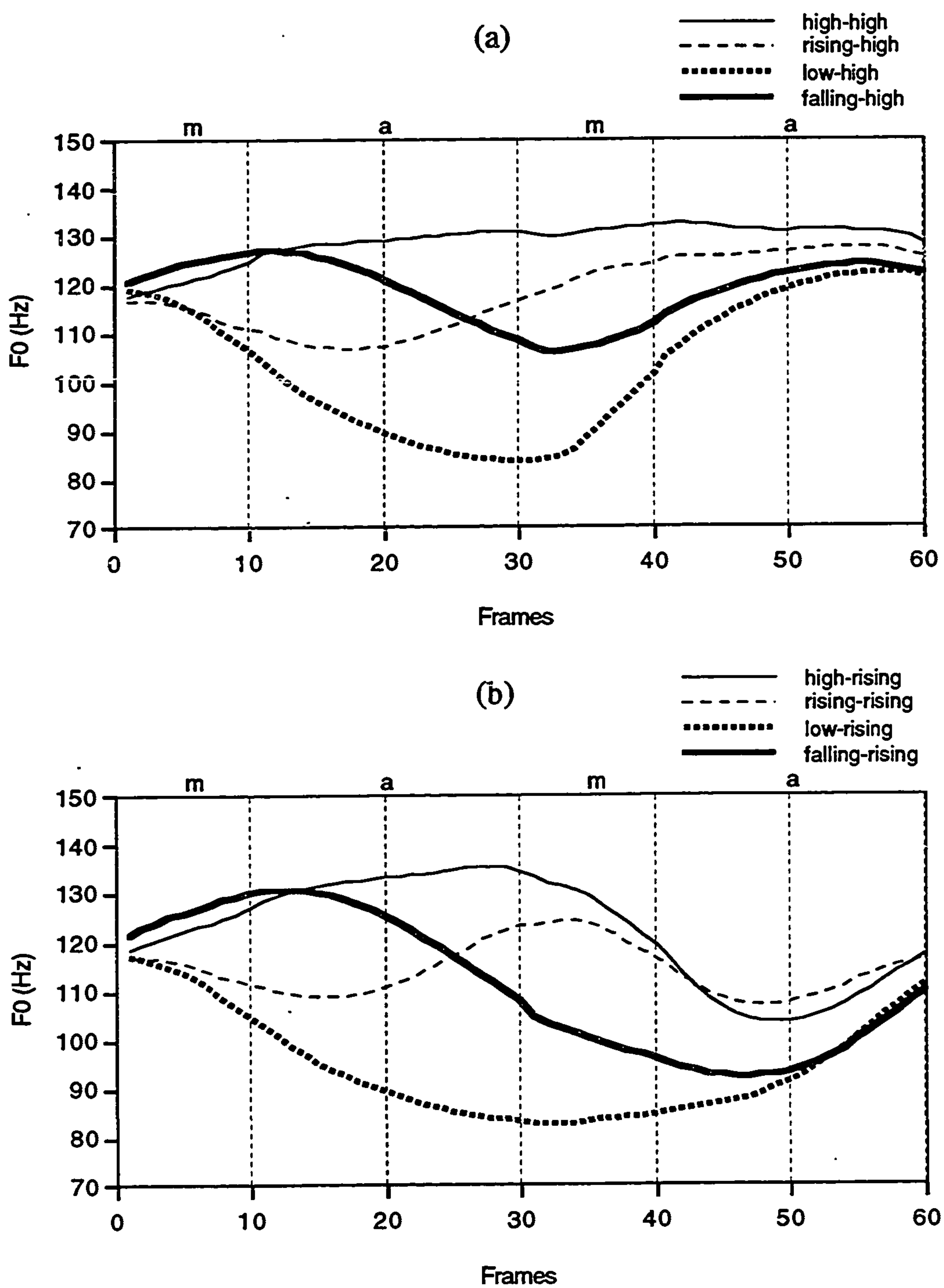


Figure 2.8. Mandarin high (a) and rising (b) tones under carryover tonal influence in /mama/ sequences produced with the carrier sentences.

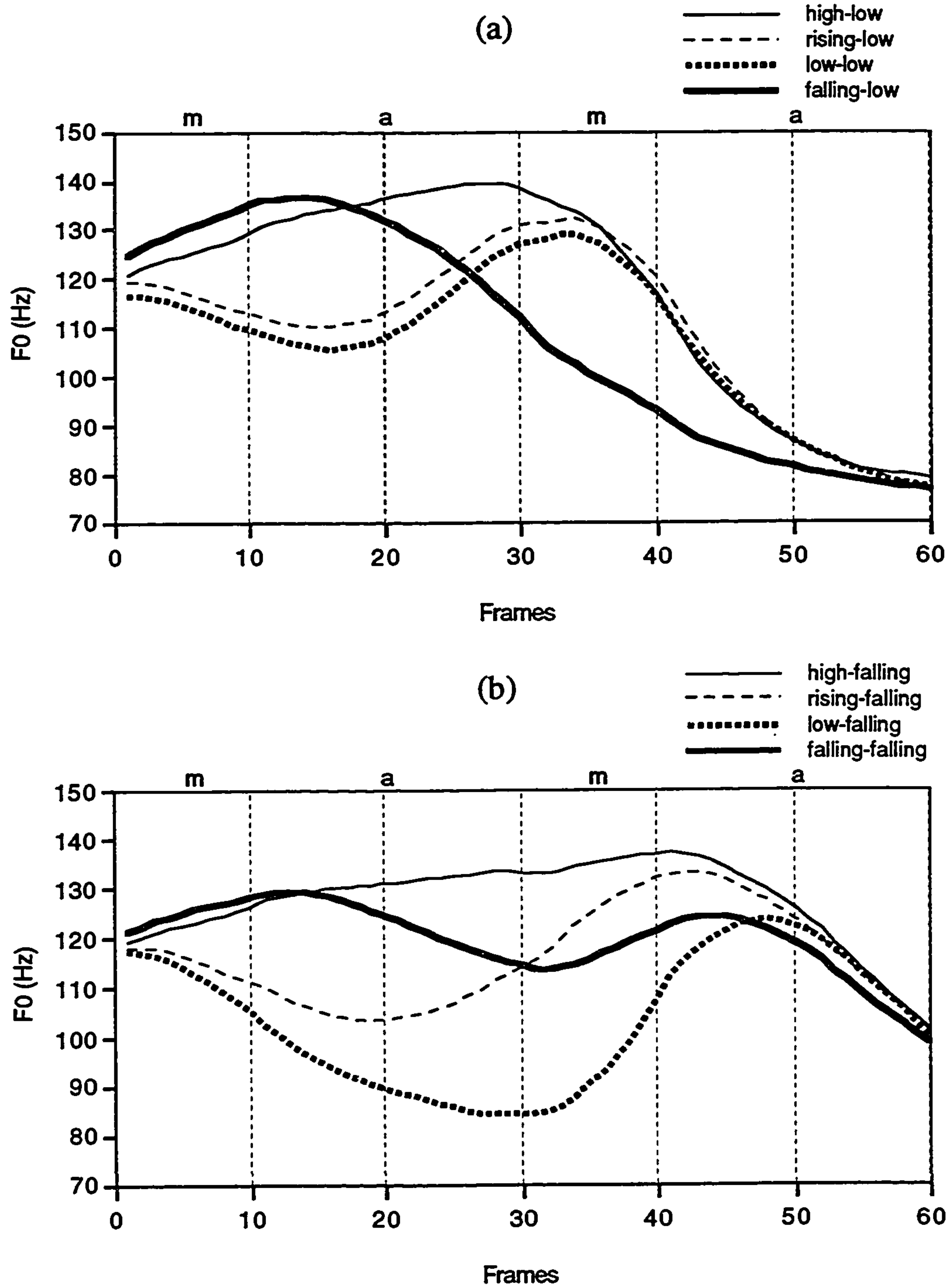


Figure 2.9. Mandarin low (a) and falling (b) tones under carryover tonal influence in /mama/ sequences produced with the carrier sentences.

Statistical analysis for disyllabic sequences produced in carrier sentences

Table 2.2 shows the ANOVA results for the nasal and vocalic segments in the /mama/ sequences produced with the carrier sentences. The effect shown in the table is Pre-pitch — the ending pitch of the first syllable: H (high) for the high and the rising tones, and L (low) for the low and the falling tones. Again, what is being tested here is whether the F0 value of the second syllable is raised by a high ending pitch but lowered by a low ending pitch of the first syllable.

Table 2.2. ANOVA Results for carryover effects on the second nasal and vocalic segments.

		Maxf0		Meanf0		Minf0		Diff0	
Segment	Tone	F	P	F	P	F	P	F	P
Nasal	high	23.26	*.0002*	28.59	*.0001*	24.22	*.0002*	4.82	.0444
	rising	43.23	*.0001*	48.16	*.0001*	41.86	*.0001*	8.81	*.0096*
	low	81.03	*.0001*	96.26	*.0001*	54.50	*.0002*	1.02	.3453
	falling	18.08	*.0007*	19.33	*.0005*	13.70	*.0021*	1.53	.2349
Vocalic	high	5.41	.0345	11.52	*.0040*	22.94	*.0002*	54.42	*.0001*
	rising	7.02	.0182	30.00	*.0001*	38.75	*.0001*	28.93	*.0001*
	low	42.59	*.0003*	12.41	*.0097*	.06	.8102	43.19	*.0003*
	falling	11.45	*.0041*	6.67	.0208	3.20	.0937	12.75	*.0028*

df: 15 for the high, rising, and falling tones; 7 for the low tone.

In Table 2.2, most of the effects are highly significant. For the nasal segment of the second syllable, the carryover effect is not significant on the Diff0 of the low and the falling tones. The effect on the Diff0 of the high tone is significant only at the 0.05 level.

For the vocalic segment of the second syllable, the Minf0 of the low and the falling tones are not affected by carryover effects. The carryover effects on Maxf0 of the high and the rising tones, and on MeanfO of the falling tone, though significant at the 0.05 level, do not reach the 0.01 level.

Referring back to the comparison of the averaged curves shown in Figures 2.8 and 2.10, one can see clearly that the carryover effects are assimilatory, i.e., the higher the ending value of the preceding tone, the higher the value of the following tone. In general, although the basic contour of a tone is preserved in disyllabic sequences, both the onset F0 and the overall F0 height of the tone is raised by a preceding tone that has a high ending pitch, and lowered by a preceding tone that has a low ending pitch.

Cross-syllabic carryover effects

Figures 2.10-2.13 show the effect of the preceding pitch in the carrier sentence on the tones of the /mama/ sequences as well as that of the ending pitch of the first syllable on the tones of the second syllable in the /mama/ sequences. In those figures, the pitch height of the last tone in the carrier sentence before a /mama/ sequence is represented by the thickness of the plotting line: thin for high, and thick for low. The tone combination for the /mama/ sequence is represented by the line pattern of the plotted curves.

Three carryover effects can be seen in Figures 2.10-2.13. (1) the ending pitch of the carrier on the tone of the first syllable in the /mama/ sequence; (2) the ending pitch of the first syllable on the tone of the second syllable in the /mama/ sequence; (3) the ending pitch of the carrier on the tone of the second syllable in the /mama/ sequence. Effects (1) and (2) are on immediately adjacent tones, whereas effect (3) is on nonadjacent tones. Effect (2) has been discussed in earlier sections.

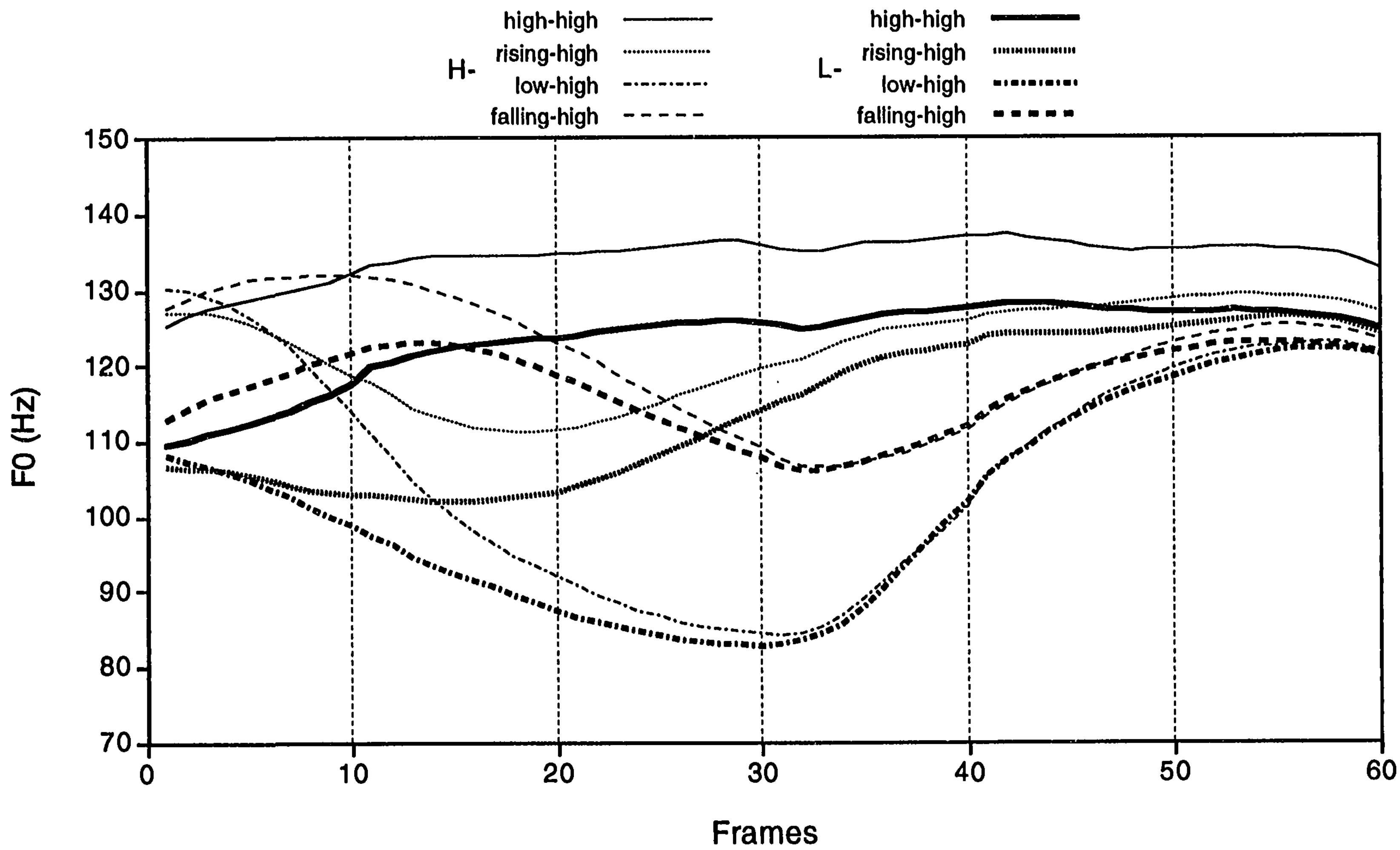


Figure 2.10. Mandarin high tone under immediate and cross-syllabic carryover tonal influence.

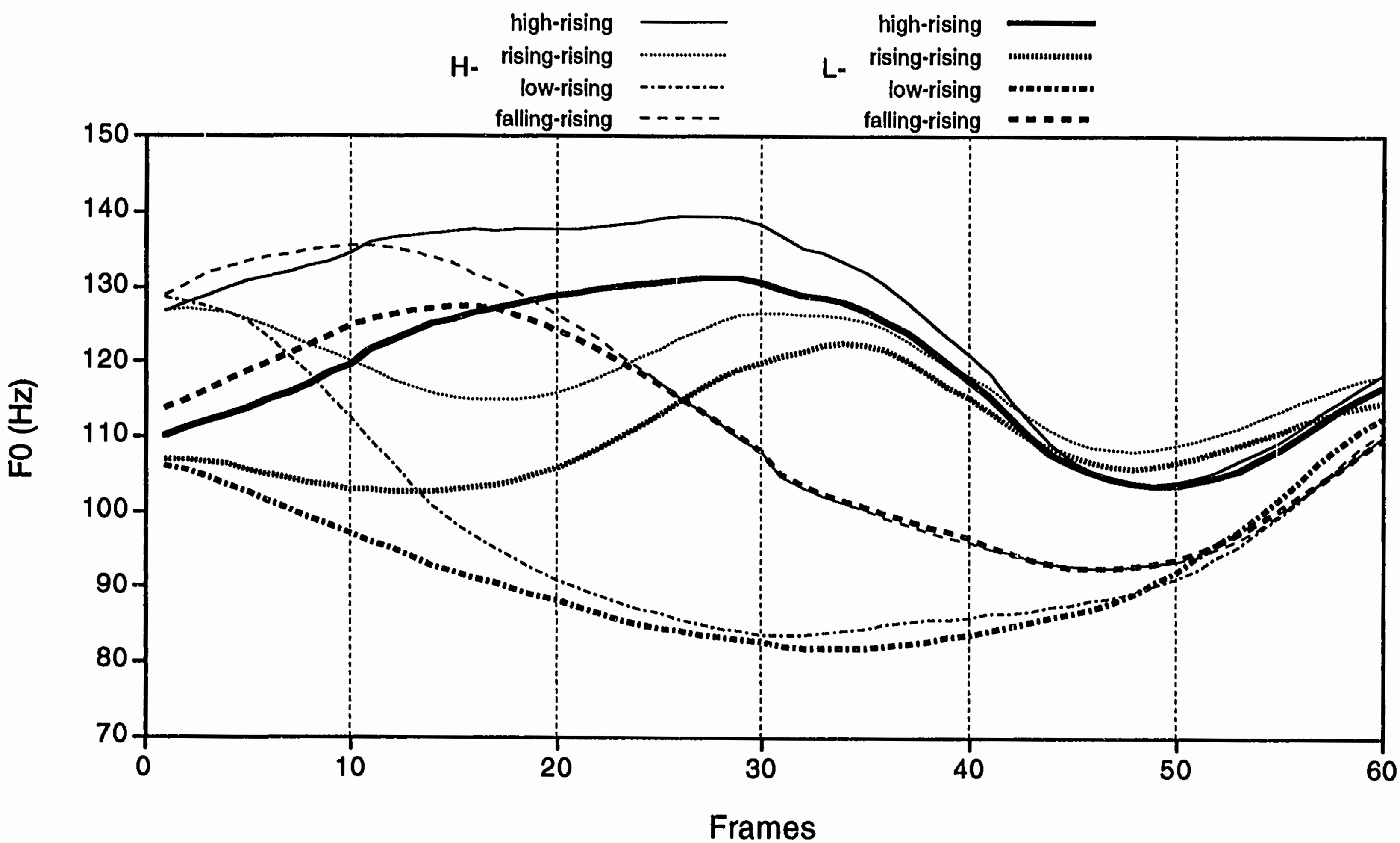


Figure 2.11. Mandarin rising tone under immediate and cross-syllabic carryover tonal influence.

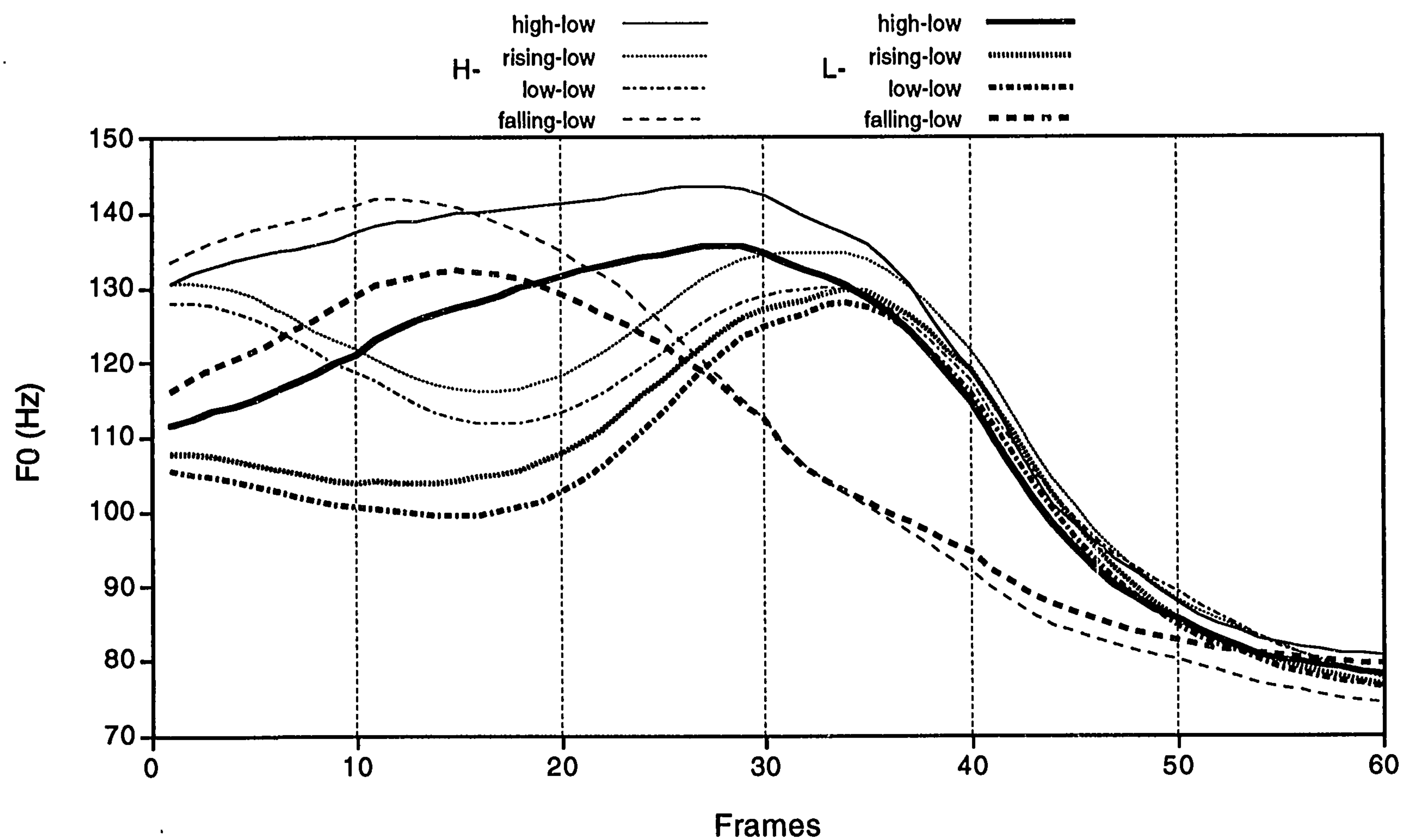


Figure 2.12. Mandarin low tone under immediate and cross-syllabic carryover tonal influence.

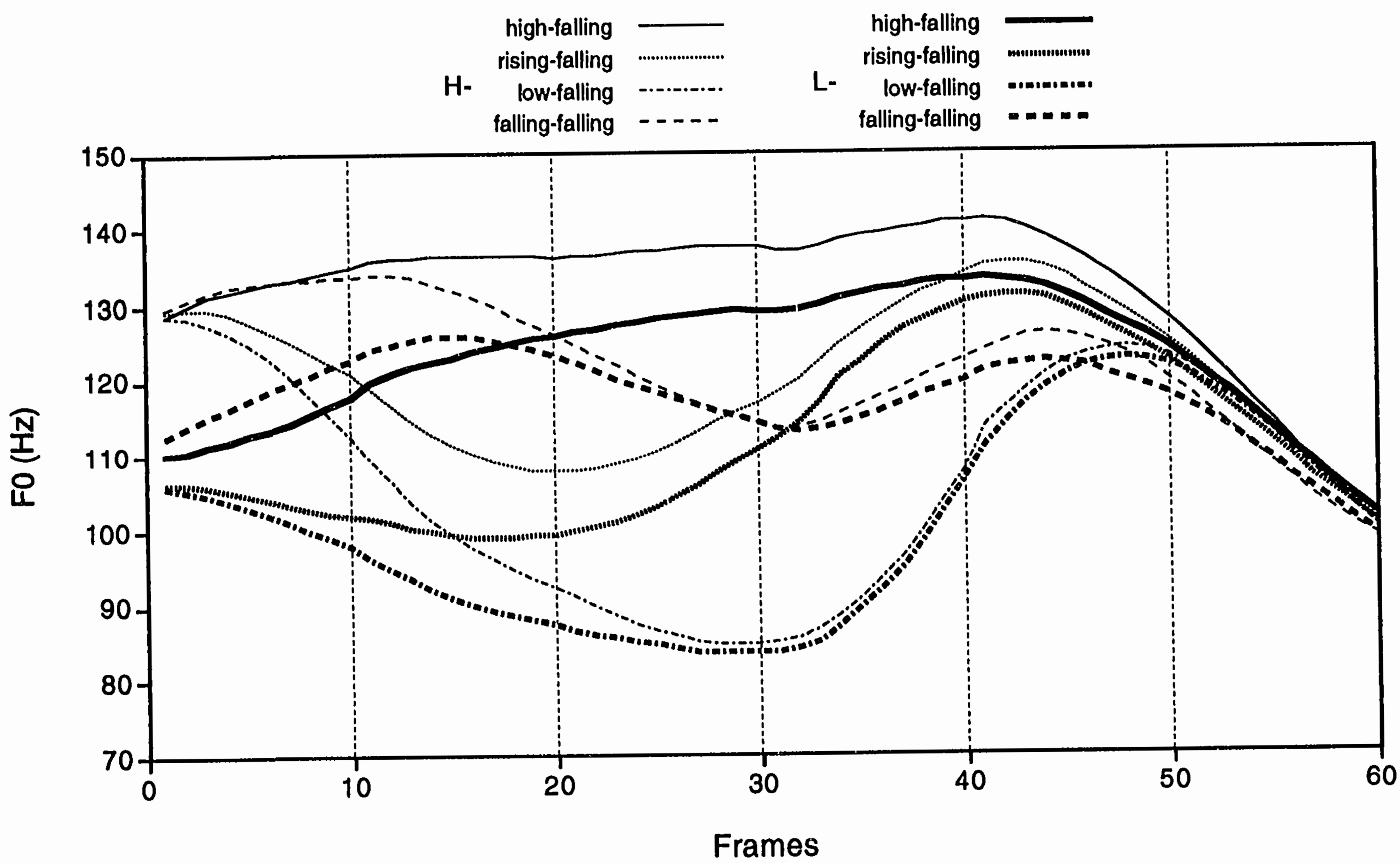


Figure 2.13. Mandarin falling tone under immediate and cross-syllabic carryover tonal influence.

Carryover effects (1) and (2) are both consistent and extensive. The F0 of a tone is higher when following a high-ending context than a lower-ending context, and with only a few exceptions (falling-rising, falling-low, falling-falling), the differences persist all the way through the tone of the first syllable. Notice that the exceptions are found only in case the tone being influenced is the falling tone. Also, as discussed in earlier sections for effect (2), the carryover effect on the falling tone usually fades away toward the end of the tonal contour.

Carryover effect (3), although not as consistent and extensive as effects (1) and (2), is apparent in many cases. The clearest case is for the high-high tone sequence, in which the difference remains as large as 6 Hz even at the end of the second high tone. In fact, the high tone, among the four tones, is the most transparent to the carryover effect by the preceding tone, letting it affect its own onset and the overall F0 height, as well as the onset and the overall F0 height of the next following tone. The rising tone is also rather transparent to the carryover effect, although somewhat less in degree than the high tone, letting it affect itself as well as the next following tone.

Because the carryover effect on the falling tone does not usually reach the end of the tone, the falling tone seems to often block the carryover effect on the next following tone. The low tone does not seem to be very transparent to the carryover effect either, leaving only a very small effect, if any at all, to be carried over to the next following tone.

Statistical analysis for cross-syllabic carryover effect

Several three-factor ANOVAs were conducted to examine the immediate as well as the cross-syllabic carryover effects. The independent variables are (1) Pre-word Pitch — the ending pitch of the syllable in the carrier sentence immediately preceding the /mama/ sequence (Hi or Lo); (2) Pre-pitch, the ending pitch of the first syllable in the /mama/

sequence (H or L); (3) Tone of the second syllable in the /mama/ sequence (high, rising, low, or falling). The dependent variables are Maxf0, Meanf0, Minf0, and Diff0.

Table 2.3 lists all the main effects and all the two-way interactions obtained in the ANOVAs. None of the three-way interactions is significant, and so they are not listed here.

Table 2.3. ANOVA results for cross-syllabic carryover effects

		Maxf0		Meanf0		Minf0		Diff0	
Segment	Effect	F	P	F	P	F	P	F	P
Nasal	Pre-word Pitch	8.09	*.0063*	7.82	*.0073*	6.50	.0138	4.79	.0332
	Pre-pitch	100.76	*.0001*	107.49	*.0001*	77.66	*.0001*	.721	.3997
	Tone	14.10	*.0001*	9.15	*.0001*	6.13	*.0012*	5.33	*.0028*
	Pre-word Pitch x Pre-pitch	10.68	*.0019*	9.62	*.0031*	11.20	*.0015*	.003	.9547
	Pre-word Pitch x Tone	.03	.9928	.12	.9512	.653	.5846	4.85	*.0048*
	Pre-pitch x Tone	1.83	.1528	.47	.7023	.13	.9429	3.36	.0255
Vocalic	Pre-word Pitch	4.13	.0472	2.10	.1538	.49	.4868	5.67	.0209
	Pre-pitch	51.53	*.0001*	45.78	*.0001*	43.59	*.0001*	9.76	*.0029*
	Tone	40.40	*.0001*	125.34	*.0001*	100.85	*.0001*	18.23	*.0001*
	Pre-word Pitch x Pre-pitch	12.65	*.0008*	14.15	*.0004*	8.41	*.0054*	1.18	.2825
	Pre-word Pitch x Tone	.38	.7688	.62	.6073	1.26	.2989	.74	.5348
	Pre-pitch x Tone	3.50	.0218	1.70	.1786	10.07	*.0001*	39.56	*.0001*

Total df = 52

As in the data we have already seen, the effect of Pre-pitch is highly significant on all the F0 measurements for both the nasal and the vocalic segments, except on Diff0 for the nasal segment.

For the nasal segment, the effect of Pre-word Pitch is significant on all the F0 measurements, although the effects on Minf0 and Diff0 are significant only at the 0.05

level. For the vocalic segment, the effect of Pre-word Pitch is significant only for Diff0. However, it is interesting to notice that the interaction between Pre-word Pitch and Pre-pitch is highly significant for both the nasal and the vocalic segments.

Figures 2.14 and 2.15 plot the interaction for the two segments. In the two figures, the H Pre-pitch is represented by darker bars, while the L Pre-pitch is represented by lighter bars; the Hi Pre-word Pitch is represented by filled bars, while the Lo Pre-word Pitch is represented by striped bars. It is clear from those two figures that the difference in F0 measurement due to Pre-pitch is greater when the Pre-word Pitch is Hi than when it is Lo. This indicates that the high and the rising tones in the first /ma/, both of which have high ending pitch, are more transparent to carryover effect of the preceding tone upon the next following tone.

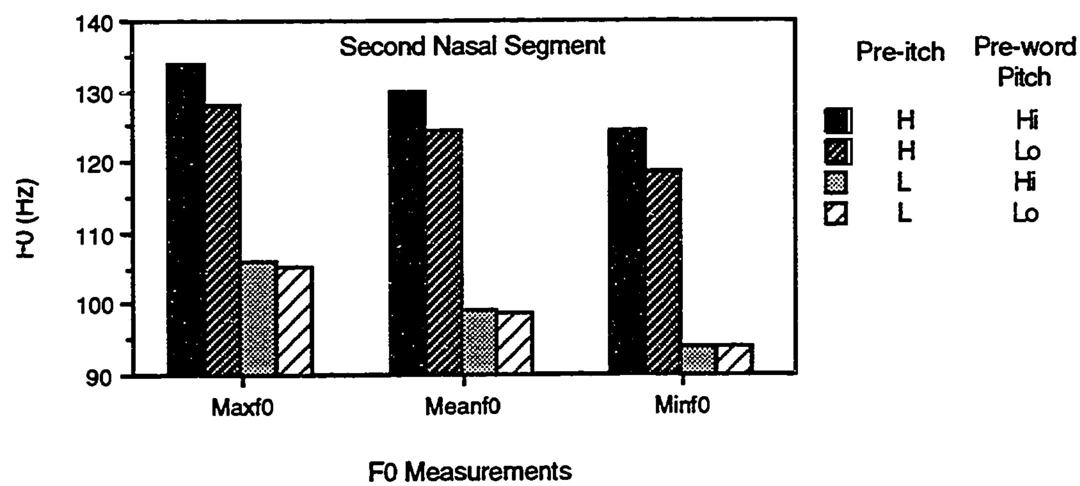


Figure 2.14. Interaction of Pre-word Pitch and Pre-pitch on Maxf0, Meanf0, and Minf0 for the second nasal segment in the /mama/ sequences.

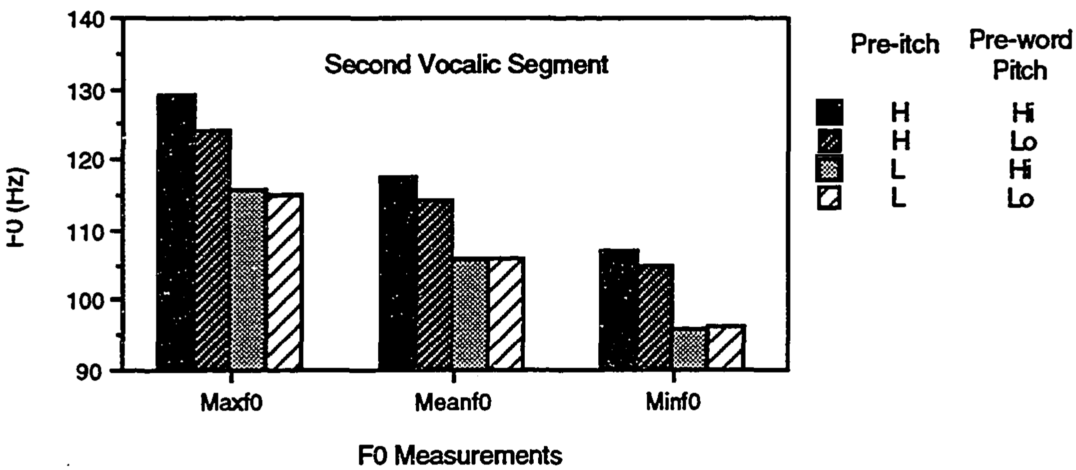


Figure 2.15. Interaction of Pre-word Pitch and Pre-pitch on Maxf0, Meanf0, and Minf0 for the second vocalic segment in the /mama/ sequences.

Several other significant interactions are also of interest to us. The interaction between Pre-pitch and Tone on Diff0 (see Figure 2.16) is a reflection of the fact that when the adjacent pitches of two syllables have very different values, i.e., in cases of "conflicting context" as defined in Chapter 1, there must be a more rapid glide between the two pitches, and the glide usually occurs during the initial sonorant consonant if there is one. As a result, when a high or falling tone is preceded by a tone with a high-ending pitch, the Diff0 is smaller than when it is preceded by a tone with a low-ending pitch. For the rising or the low tone, the reverse is the case. That is, when they are preceded by a tone with high-ending pitch, the Diff0 is larger than when they are followed by a tone with a low-ending pitch.

The highly significant interaction between Pre-word Pitch and Tone on Diff0 (Figure 2.17) is due to a similar mechanism, which makes the glides even steeper in cases of "conflicting context" than "compatible context."

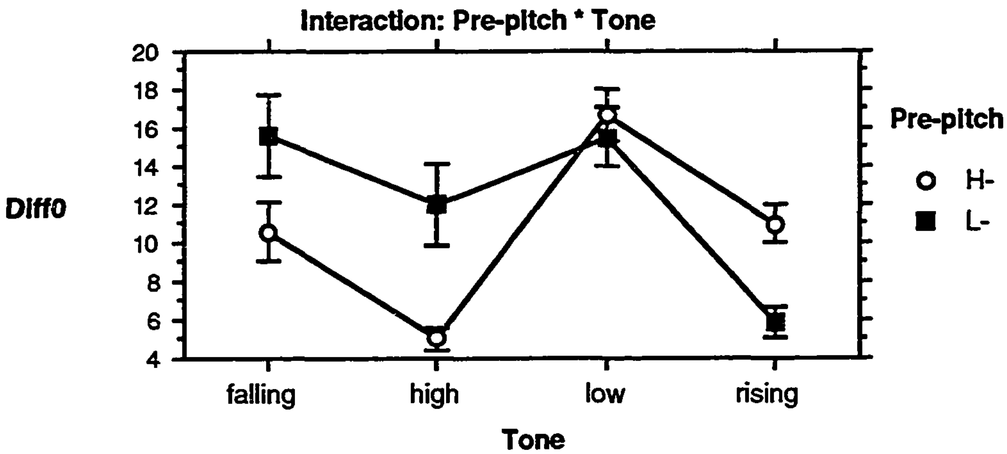


Figure 2.16. Interactions between Pre-pitch and Tone on Diff0 for the second nasal segment in the /mama/ sequences.

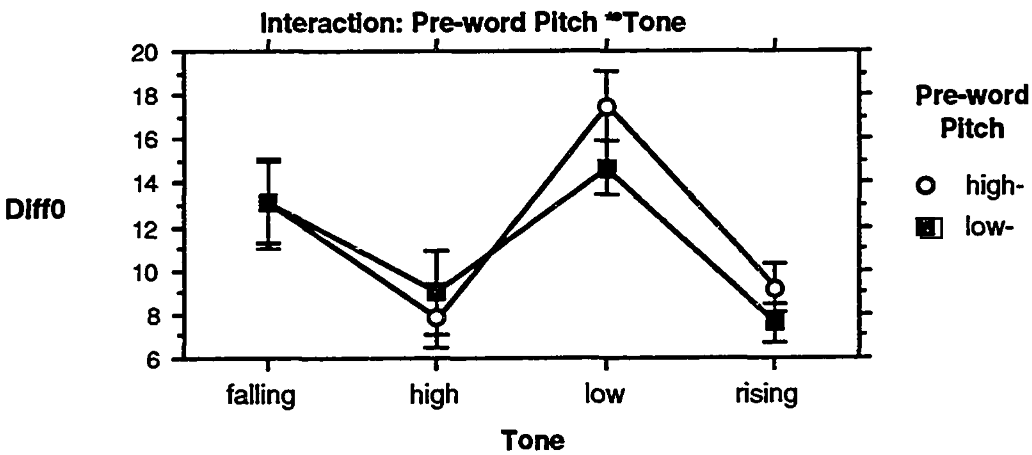


Figure 2.17. Interactions between Pre-word Pitch and Tone on Diff0 for the second nasal segment in the /mama/ sequences.

For the second vocalic segment, there is no interaction between Pre-word Pitch and Tone on any of the F0 measurements. However, the interaction between Pre-pitch and Tone is highly significant for Minf0 and Diff0, and moderately significant for Maxf0. The interaction on Maxf0, as seen in Figure 2.18, is mainly due to the fact that the Maxf0 for the low tone is taken from the beginning of the second vocalic segment, whereas for other tones the Maxf0 is taken much later when the carryover effect is reduced.

The interaction between Pre-pitch and Tone for MinfO, as shown in Figure 2.19, is a reflection of the fact that, while the carryover effect exerted by the previous syllable is extensive for the MinfO of the high tone and the falling tone, the effect is minimal for the low and falling tones.

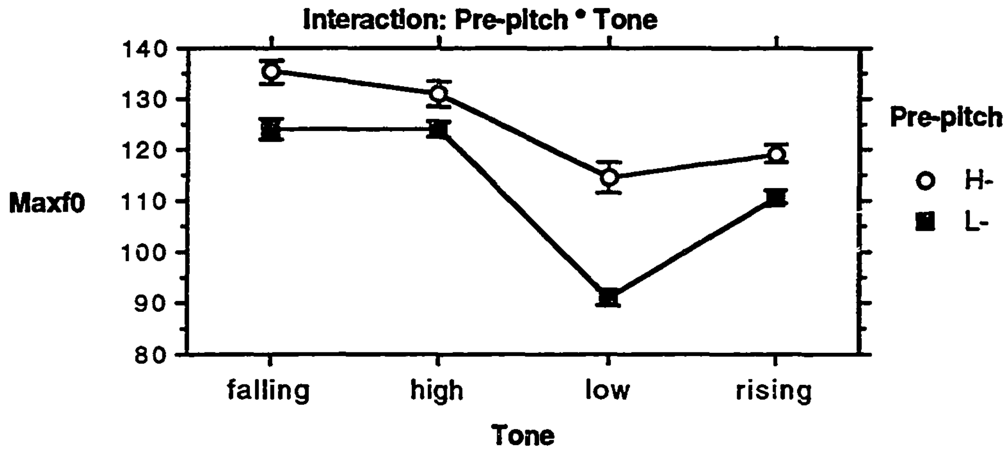


Figure 2.18. Interactions between Pre-pitch and Tone on MaxfO for the second vocalic segment in the /mama/ sequences.

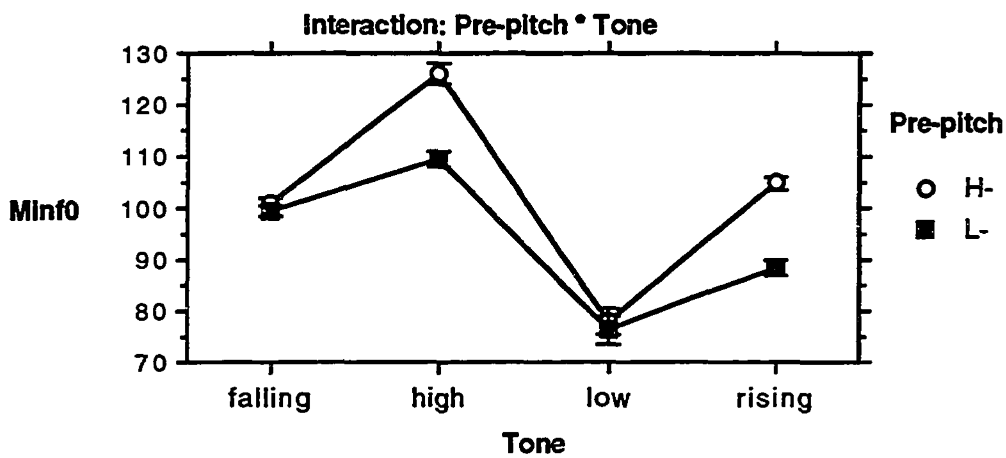


Figure 2.19. Interactions between Pre-pitch and Tone on MinfO for the second vocalic segment in the /mama/ sequences.

The highly significant interaction between Pre-pitch and Tone on Diff0 for the second vocalic segment (Figure 2.20) is also interesting. For the high and the rising tones, the Diff0 is greater when the Pre-pitch is low than when it is high. For the high tone, this is so because when the Pre-pitch is low, the whole F0 contour becomes rising rather than high level; for the rising tone, the greater Diff0 is due to the fact that when the Pre-pitch is low, the minimum F0, which usually occurs in the first half of the vocalic segment, can be lower than when the Pre-pitch is high, thus increasing the slope of the rising contour. On the other hand, both the low tone and the falling tone have greater Diff0 when the Pre-pitch is high than when it is low. In this case, the greater Diff0 associated with high Pre-pitch value is due to higher starting F0 both for the falling tone and for the low tone. What is common to all the four tones is that the carryover effect influences the first half of a following tone more than the last half of it.

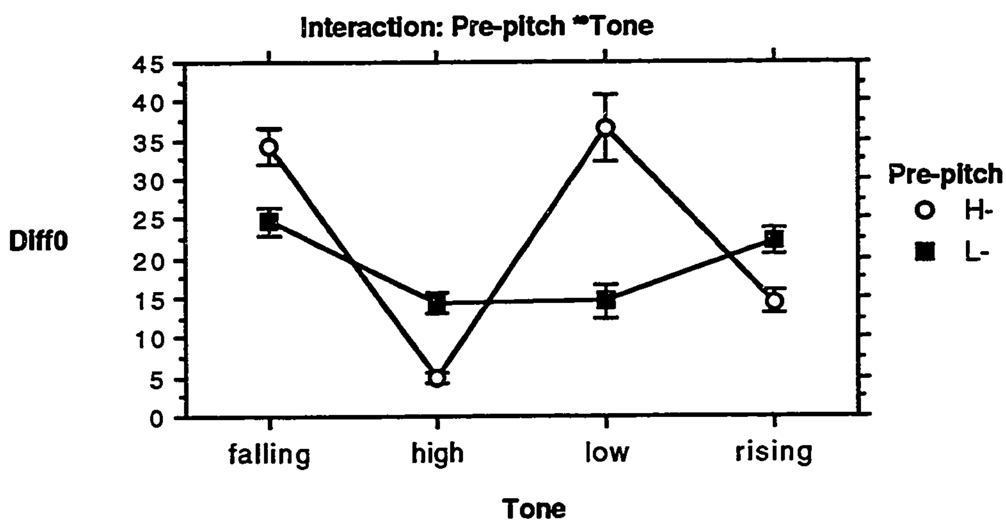


Figure 2.20. Interactions between Pre-pitch and Tone on Diff0 for the second vocalic segment in the /mama/ sequences.

Summary of findings about carryover effect

To sum up the findings so far in this chapter, it is observed that there is a strong carryover effect among Mandarin tones. The effect is assimilatory, that is, a tone with a low ending pitch lowers the F0 of the following tone, and a tone with a high ending pitch raises the F0 of the following tone. The high pitch region seems to be more susceptible to the influence of the carryover effect, and the lowest pitch region seems to have strong resistance to the effect. There is immediate as well as a cross-syllabic carryover effects. However, the effects do fade away gradually over time: during the initial nasal consonant, there are rapid F0 movements, which are larger when the adjacent pitches of two neighboring tones have opposing values than when they have similar values; the F0 differences due to carryover effects remain quite sizable during the vowel, though with reduced magnitude; in certain cases, the effect is still clearly visible even at the end of the second syllable, indicating that even the third syllable in a row may be affected. Since the low pitch region is not much affected by the carryover effects, tones having a rather low pitch value, especially those with a low final pitch (the low and falling tones), apparently block the carryover effect and prevent it from being carried over to the next following syllable. The nature of this blocking seems be similar to the coarticulatory resistance reported by Recasens (1984, 1985).

Because of their assimilatory nature, and because they are unlikely to be heard as change of tonal categories, the carryover effects found here meet the criterion discussed in the Introduction of this thesis, and thus should be considered instances of coarticulation. Carryover tonal coarticulation has been found in several studies (Chuang, et al., 1971; Han & Kim, 1974; M. Lin & Yan, 1991; Shen, 1990; Shih, 1986; Wu, 1984), however, because of the initial nasal consonants in the syllables being examined here, the effects are much clearer than have been seen before. More interestingly, cross-syllabic coarticulation has been clearly observed for tones. This kind of cross-syllabic coarticulation is comparable to the vowel-to-vowel coarticulation reported by Magen (1989) and Whalen (1990), because in both cases, the coarticulatory effects extend across a whole syllable. Together, they pose serious difficulties for various existing models of coarticulation. For the feature-spreading models (Benguerel & Cowan, 1974; Henke, 1966; Keating, 1985, 1988; Moll & Daniloff, 1971), the apparent gradual reduction over time of the carryover effect does not seem to result from spreading of the distinctive features. For the various overlapping models (Fowler, 1977, 1980; Browman & Goldstein, 1985, 1986; Liberman et al., 1967; Liberman & Mattingly, 1985), a tonal contour could be defined either as a constellation of several gestures, or as a single dynamic gesture. In either case, however, it is difficuit to conceive the functions of tonal gestures each having the duration of several syllables. The long distance coarticulation cannot be simply attributed to articulatory inertia either. After all, the distance covered by the initial movement toward the starting pitch target in a tone is much greater than the remaining difference at the end of the F0 contours caused by different preceding tones. Inertia should have been well overcome by then, given the momentum displayed in the beginning portion of the syllable.

Anticipatory Effects

Isolated disyllabic sequences

Figures 2.21 and 2.22 shows the F0 variations of the tone of the first syllable due to the influence of the second syllable in the /mama/ sequences produced in isolation. When the second /ma/ has a high starting F0 value, the curve is drawn with a solid line; when the second /ma/ has a low starting F0 value, the curve is drawn with a dashed line.

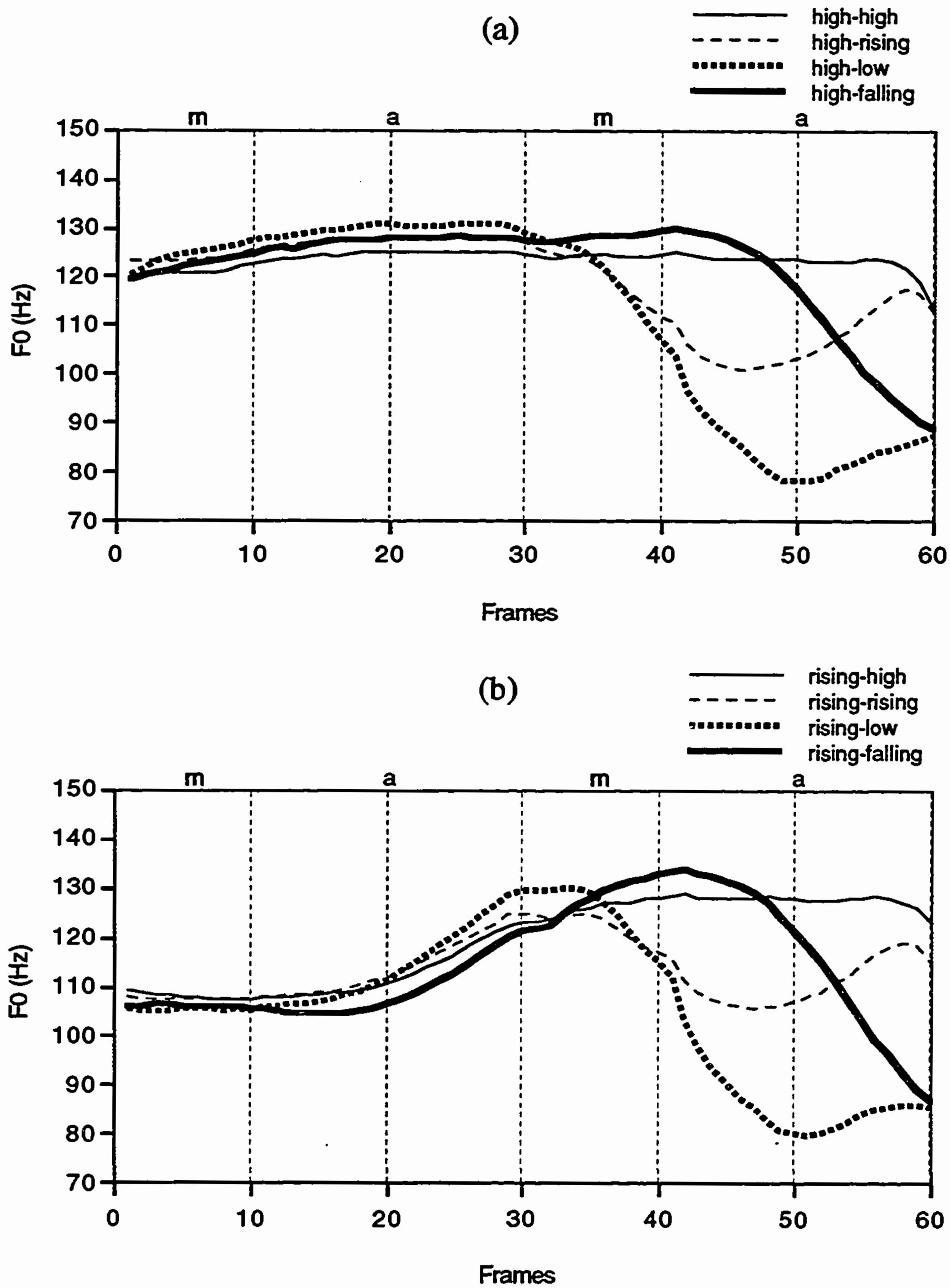


Figure 2.21. Mandarin high (a) and rising (b) tones under anticipatory tonal influence in /mama/ sequences produced in isolation.

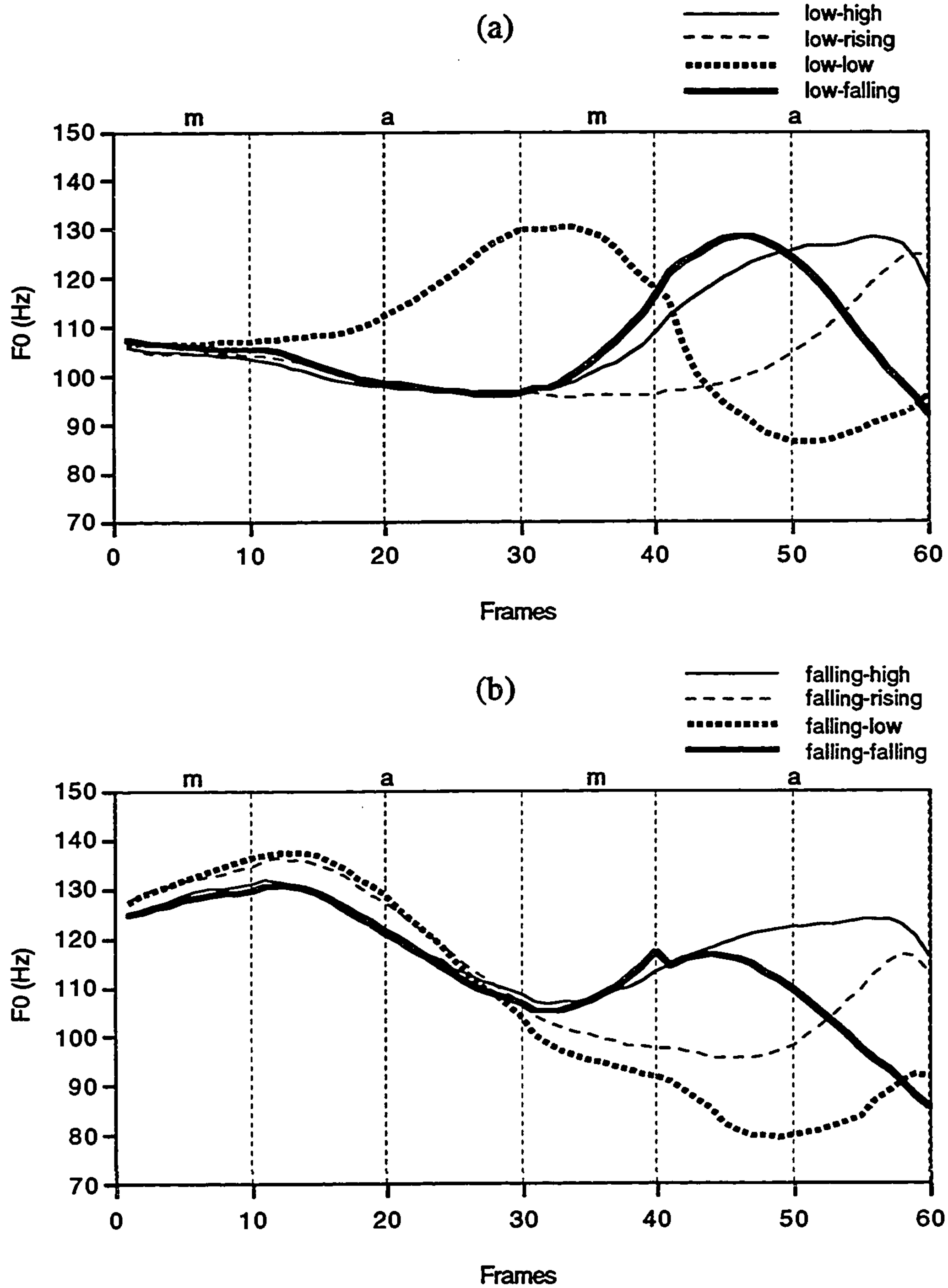


Figure 2.22. Mandarin low (a) and falling (b) tones under anticipatory tonal influence in /mama/ sequences produced in isolation.

Surprisingly, instead of tonal assimilation, in most of the cases, a dissimilatory effect can be seen. That is, when the first syllable is followed by a tone with low starting pitch, its F0 contour is somewhat higher than when it is followed by a tone with high starting pitch. When the first syllable has a high tone, the maximum F0 is highest when the second syllable has a low tone, and the second highest when the second syllable has a rising tone. The maximum F0 is lowest before a high tone, and second lowest before a falling tone. When the first syllable has a rising tone, its maximum F0 ranges from the highest to the lowest when followed by the low, the rising, the high, and the falling tones respectively. The highest F0 values in the vocalic segment occur at the end of the segment. However, it is interesting to observe that when followed by the low tone, the F0 remains on a plateau across the end of the first vocalic segment, and extends as far as about a third of the way into the second nasal segment. When the first syllable has a falling tone, its F0 peak is in the first quarter of the first vocalic segment. When it is followed by a rising or a low tone, the peak F0 value is about the same, and both of them are higher than when the second syllable has the high or the falling tone, in which cases the peak F0 values are almost identical. When the first syllable has a low tone, however, the F0 contour of the first syllable does not seem to be affected by the following tone, except by the low tone due to the phonological sandhi that changes the first low tone into a rising tone.

Statistical analysis for isolated disyllabic sequences

The dependent and independent variables used in the statistical analyses are the same as those used in examining the carryover effect. Table 2.4 shows the ANOVA results for the nasal and vocalic segments.

For the nasal segment, the effect of Post-pitch is significant only on Maxf0. The effect of Tone is highly significant for all the F0 measurements except Diff0. The interaction between Post-pitch and Tone is significant on two of the F0 measurements, Maxf0 and Meanf0, as shown in Figures 2.23 and 2.24. It can be seen in those figures that the interaction is due to greater influence of the context on the high tone and the falling tone than on the other two tones. It can also be seen that for the high and the falling tones, the mean Maxf0 and Meanf0 are higher when the following tone has a low starting pitch.

Table 2.4. ANOVA Results for anticipatory effects on F0 of the first nasal and first vocalic segments in /mama/ sequences produced in isolation.

		Maxf0		Meanf0		Minf0		Diff0	
Segment	Effect	F	P	F	P	F	P	F	P
Nasal	Post-pitch	5.51	*.0275*	3.04	.0942	.76	.3930	1.66	.2104
	Tone	13.30	*.0001*	20.62	*.0001*	17.99	*.0001*	.78	.5163
	Post-pitch x Tone	3.48	.0316	3.09	.0461	1.12	.3614	.33	.8074
Vocalic	Post-pitch	33.53	*.0001*	38.14	*.0001*	.58	.4555	24.46	*.0001*
	Tone	13.50	*.0001*	18.56	*.0001*	22.52	*.0001*	36.34	*.0001*
	Post-pitch x Tone	4.06	.0182	3.34	.0360	9.30	*.0003*	9.07	*.0003*

Total df = 24

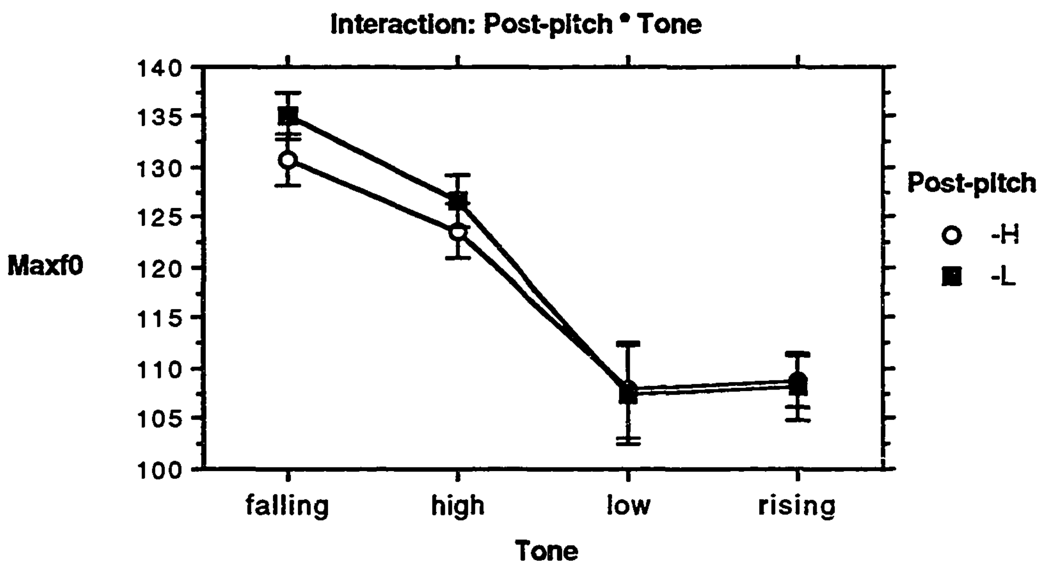


Figure 2.23. Interactions between Post-pitch and Tone on Maxf0 for the first nasal segment in the /mama/ sequences. Here and in all the subsequent figures for anticipatory effects, the hyphen in the legend text indicates context, e.g., 'H-' means that the tone under examination is preceded by a high pitch.

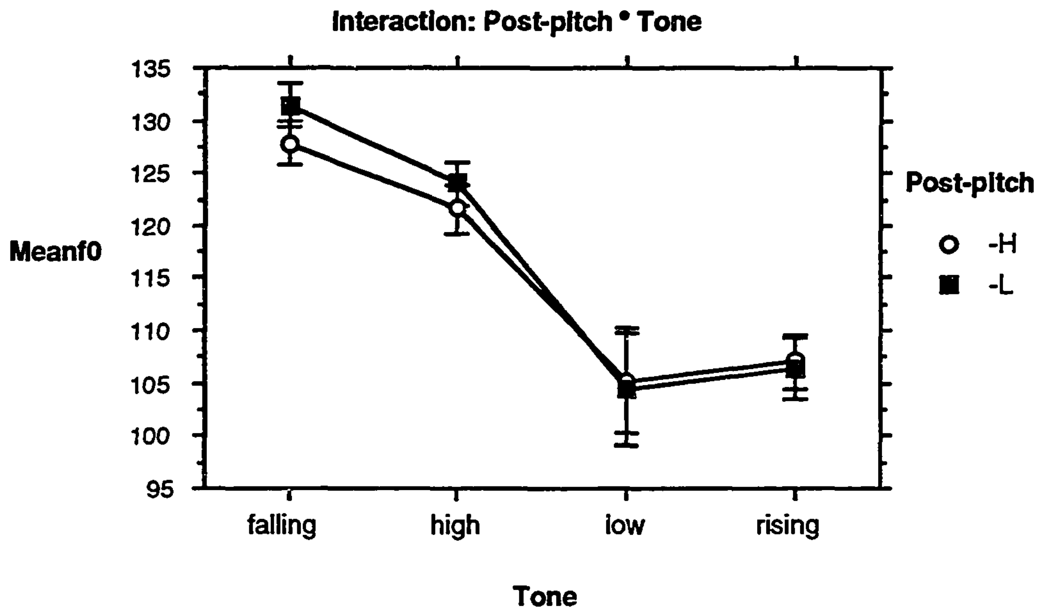


Figure 2.24. Interactions between Post-pitch and Tone on Meanf0 for the first nasal segment in the /mama/ sequences.

For the vocalic segment, the effect of Post-pitch is highly significant on all the F0 measurements except Minf0. The effect of Tone is highly significant for all the F0 measurements, which is not surprising. The interaction between Post-pitch and Tone is significant on all the F0 measurements. Figures 2.25-2.28 plots all the interactions between Tone and Post-pitch. It can be seen in Figures 2.25 and 2.26 that except for the low tone, the mean value for Maxf0 and Meanf0 are higher when the Post-pitch is low than when it is high. For Minf0 (Figure 2.27), the Post-pitch does not make much difference. For Diff0 (Figure 2.28), the effect of Post-pitch is greatest for the two more dynamic tones, namely, the falling tone and the rising tone.

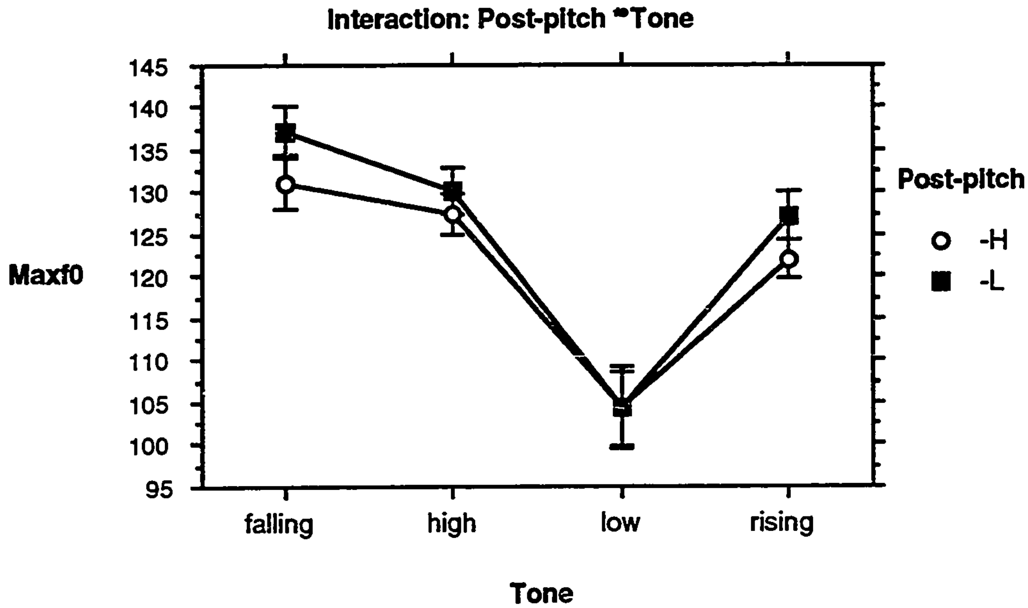


Figure 2.25. Interactions between Post-pitch and Tone on Maxf0 for the first vocalic segment in the /mama/ sequences.

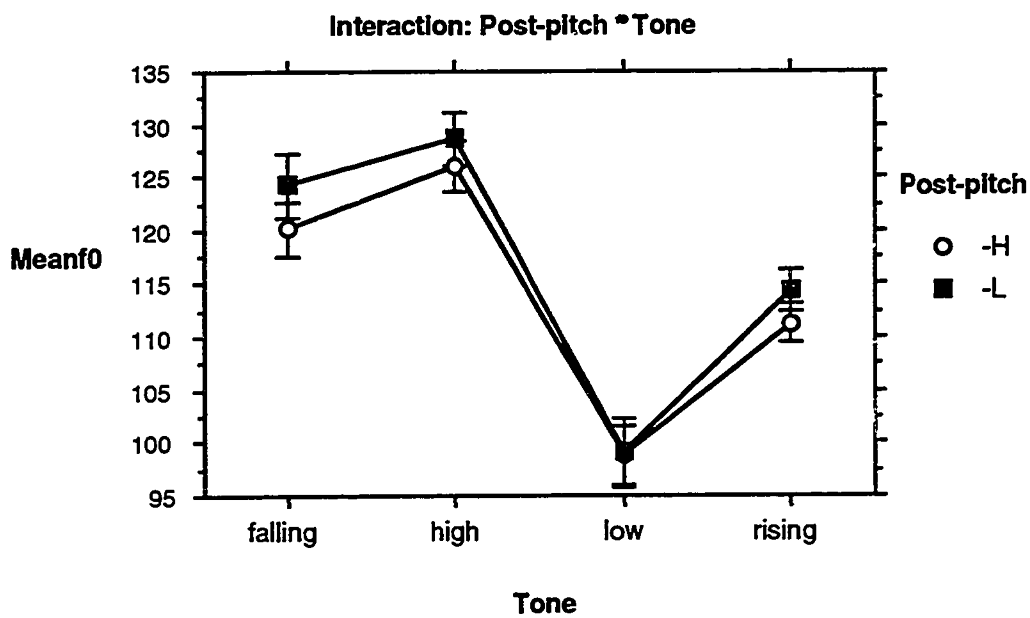


Figure 2.26. Interactions between Post-pitch and Tone on Meanf0 for the first vocalic segment in the /mama/ sequences.

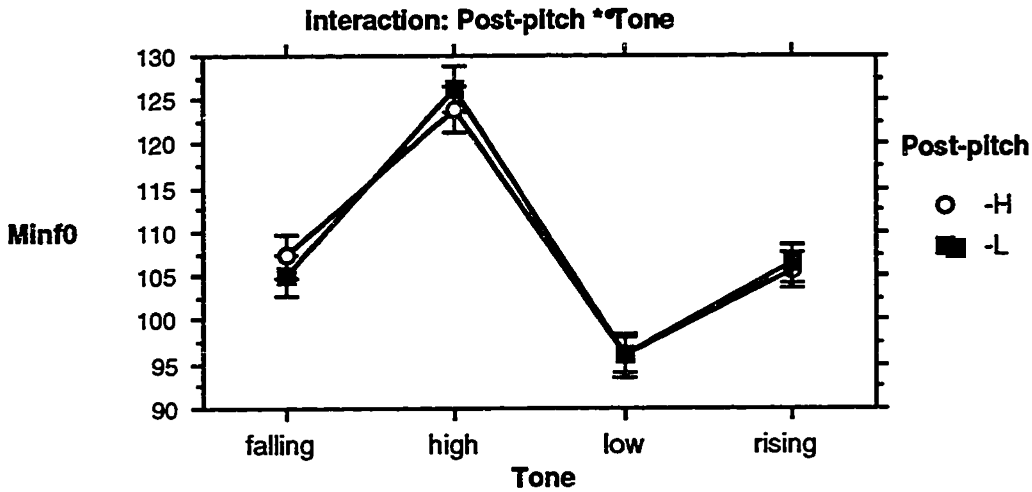


Figure 2.27. Interactions between Post-pitch and Tone on Minf0 for the first vocalic segment in the /mama/ sequences.

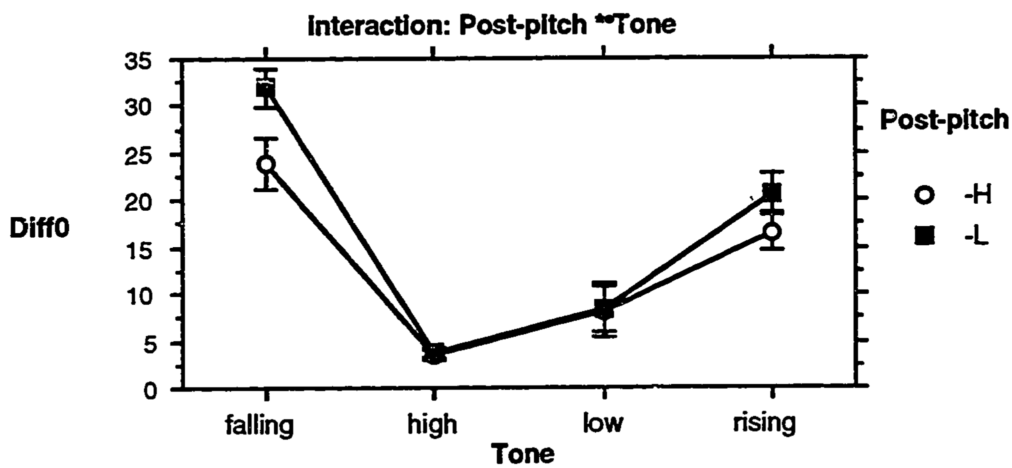
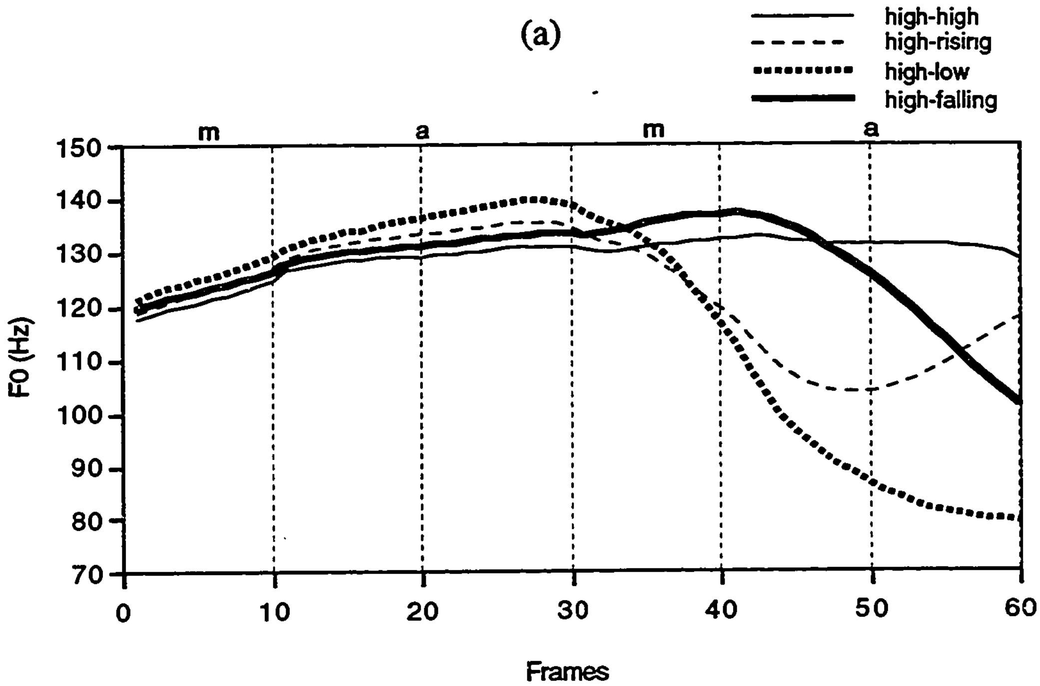


Figure 2.28. Interactions between Post-pitch and Tone on Diff0 for the first vocalic segment in the /mama/ sequences.

Disyllabic sequences produced in carrier sentences

Figures 2.29 and 2.30 shows the F0 variations of the tone of the first syllable due to the influence of the second syllable in the /mama/ sequences produced in carrier sentences. When the following tone has a high starting F0 value, the curve is drawn with a solid line; when the following tone has a low starting F0 value, the curve is drawn with a dashed line.



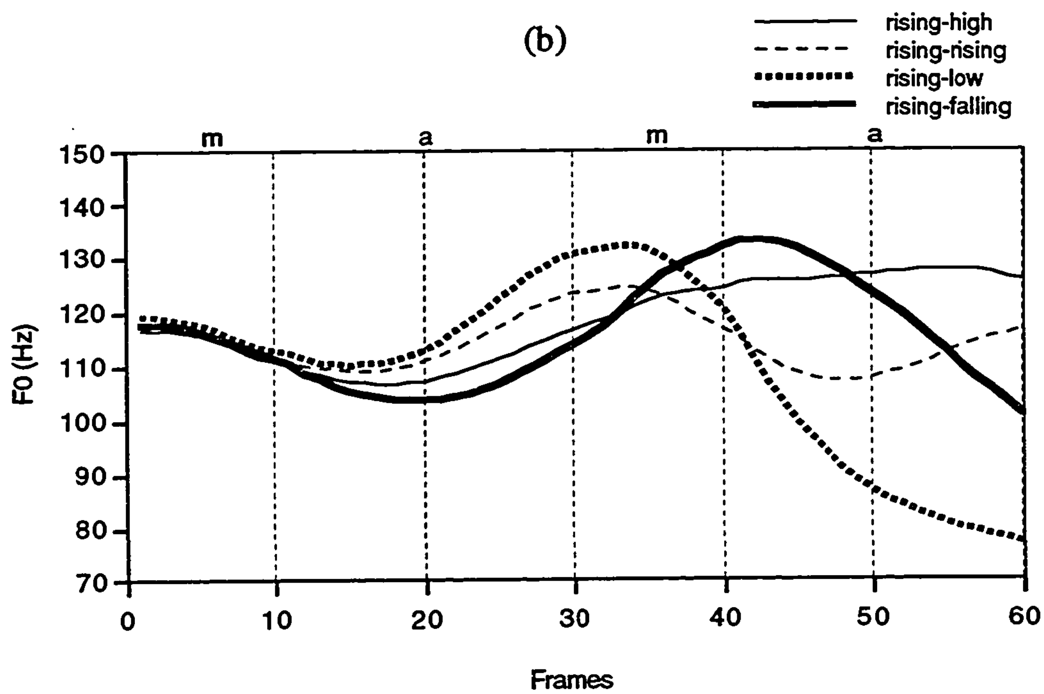


Figure 2.29. Mandarin high (a) and rising (b) tones under anticipatory tonal influence in /mama/ sequences produced in carrier sentences.

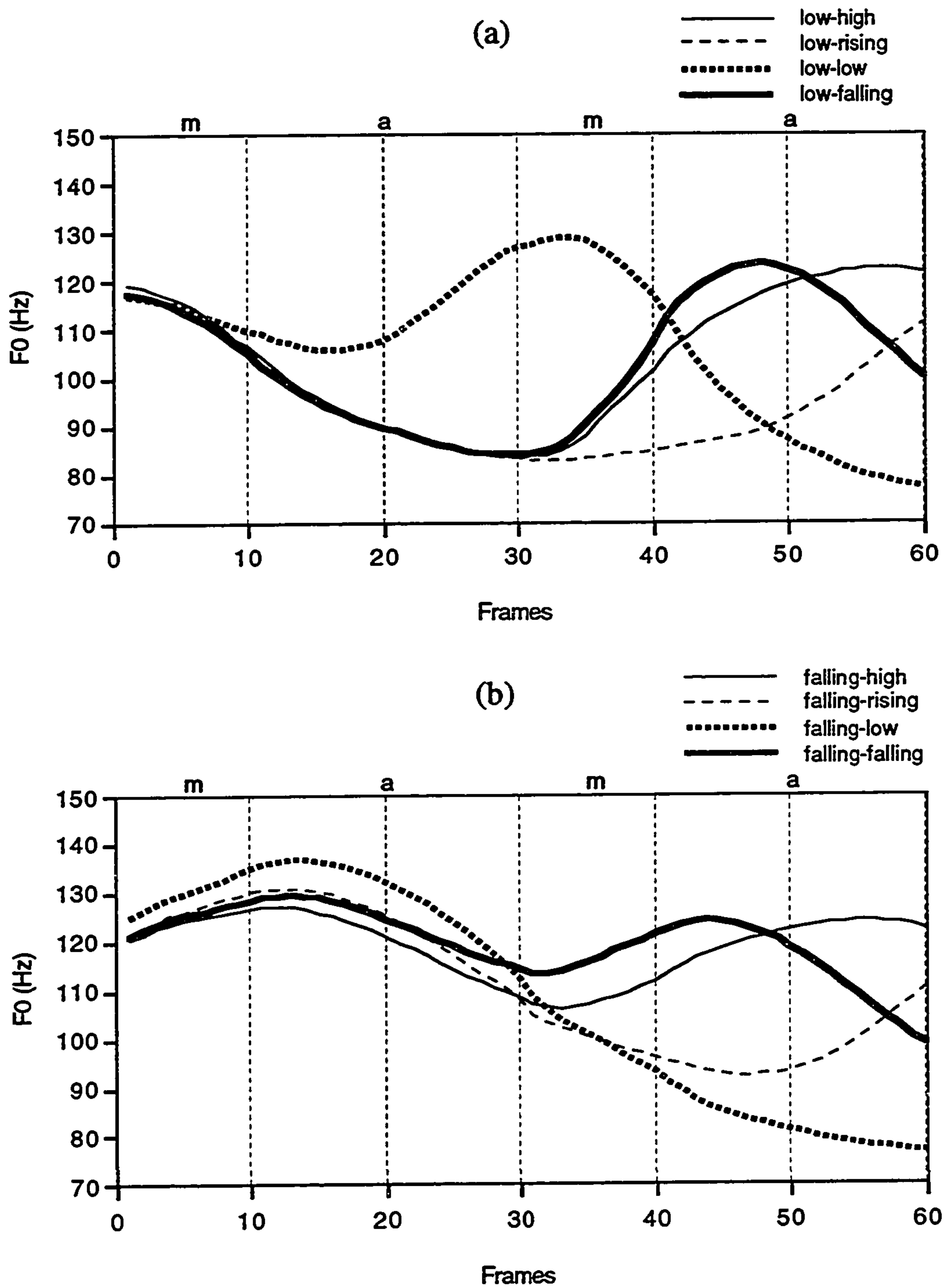


Figure 2.30. Mandarin low (a) and falling (b) tones under anticipatory tonal influence in /mama/ sequences produced in carrier sentences.

As in isolation, instead of anticipatory assimilation as one would expect, what we see here is anticipatory dissimilation. When the first syllable has the high tone, both dashed lines are higher than the solid lines in the first vocalic segment (Frames 11-30), and the differences remain throughout the entire vocalic segment. When the first syllable has the rising tone, the same trend can be seen, and the differences remain until almost half way through the second nasal segment (Frames 11-35). When the first syllable has the falling tone, a similar trend can be also seen. The differences are seen throughout the first vocalic segment when followed by the low tone, and about half way through the first vocalic segment when followed by the rising tone. For the low tone, however, no effect of the following tone can be seen except for the apparent tone sandhi when the following tone is also a low tone.

Statistical analysis for disyllabic sequences produced in carrier sentences

Table 2.5 shows the ANOVA results for the nasal and vocalic segments. The effect shown in the table is Post-pitch — the starting pitch of the first syllable: H (high) for the high and the falling tones, and L (low) for the rising and the low tones. What is being tested here is whether the F0 value of the first /ma/ is raised by a L starting pitch but lowered by a H starting pitch of the second /ma/.

As shown in Table 2.5, for the high tone, the anticipatory effect is significant on all the F0 measurements for both the nasal and the vocalic segment. For the rising tone, the anticipatory effect is significant for the vocalic segment on all the F0 measurements, but for the nasal segment, it is only marginally significant on MinfO. For the falling tone, the anticipatory effect is significant on all the F0 measurements for both nasal and vocalic segments except on the MinfO of the vocalic segment. There is no significant anticipatory effect for the low tone on any of the F0 measurements.

Table 2.5. ANOVA Results for anticipatory effects on the first nasal and vocalic segments.

		Maxf0		Meanf0		Minf0		Diff0	
Segment	Tone	F	P	F	P	F	P	F	P
Nasal	high	40.25	*.0001*	52.46	*.0001*	15.15	*.0014*	8.20	.0119
	rising	3.36	.0869	4.51	.0509	5.33	.0357	.67	.4260
	low	4.10	.0827	4.95	.0614	5.10	.0584	.04	.8467
	falling	20.38	*.0004*	14.08	*.0019*	7.78	.0137	17.10	*.0009*
Vocalic	high	33.77	*.0001*	51.44	*.0001*	40.44	*.0001*	6.56	.0217
	rising	50.24	*.0001*	37.17	*.0001*	17.29	*.0008*	38.11	*.0001*
	low	2.62	.1495	.60	.4649	.48	.5095	.13	.7318
	falling	29.66	*.0001*	18.88	*.0006*	1.70	.2116	38.56	*.0001*

df: 15 for the high, rising, and falling tones; 7 for the low tone.

The statistics demonstrate that the anticipatory dissimilation seen in Figures 2.29 and 2.30 is real: a tone that begins with a low pitch raises the Maxf0 and Meanf0 of the preceding tone, except for the low tone; a low beginning pitch of a tone also raises the Minf0 of the preceding high and rising tones. The F0 range of a tone as represented by Diff0, except for the low tone, is also widened by a low starting pitch of the following tone.

Cross-syllabic anticipatory effects

Figures 2.31-2.34 show F0 variation of the tones in the /mama/ sequences under immediate as well as cross-syllabic anticipatory influences. In those figures, the pitch height of the first tone in the carrier sentence after a /mama/ sequence is represented by the width of the plotting line: thin for high, and thick for low. The tone combination for the /mama/ sequence is represented by the line pattern of the plotted curves.

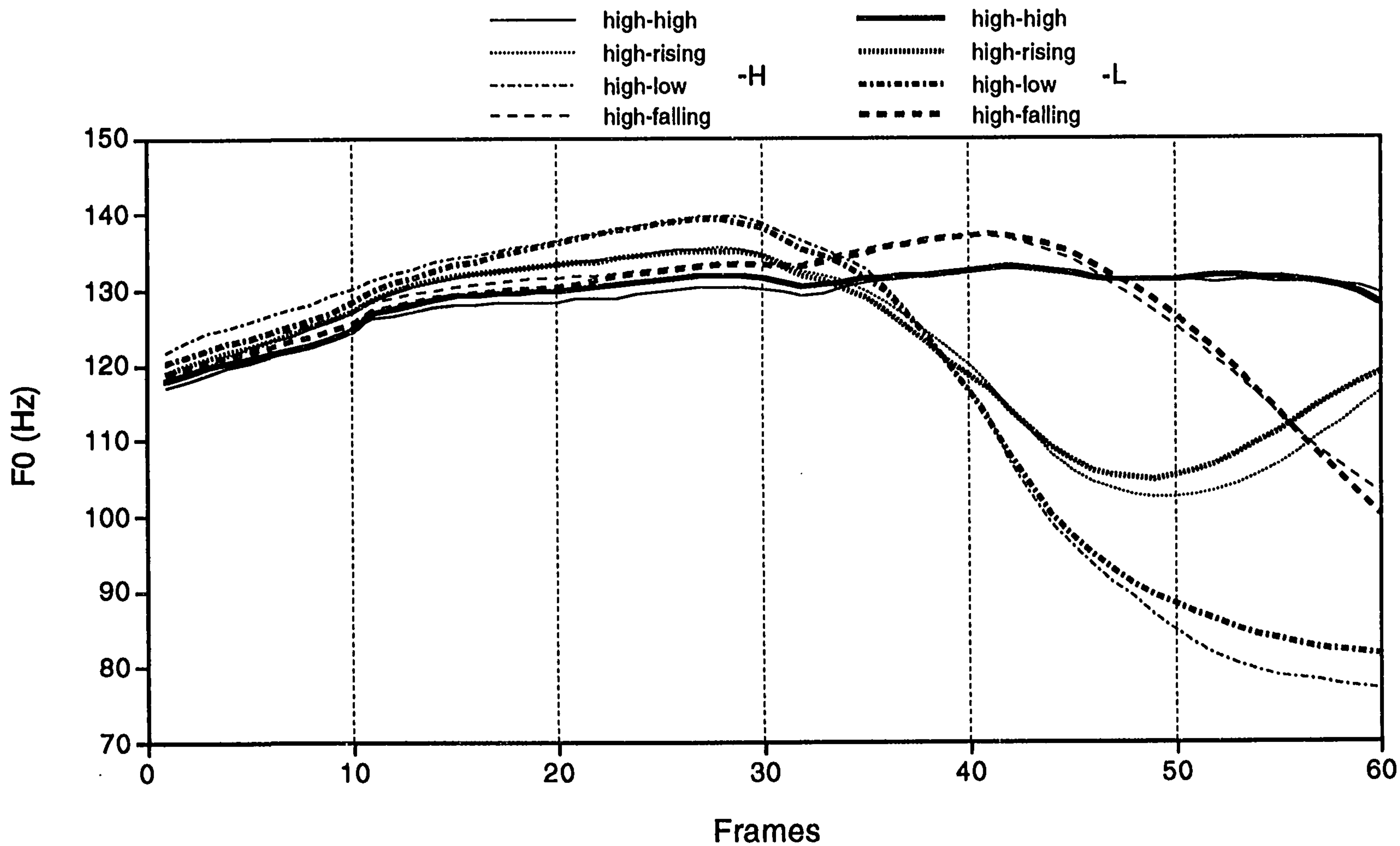


Figure 2.31. Mandarin high tone under immediate and cross-syllabic anticipatory tonal influence.

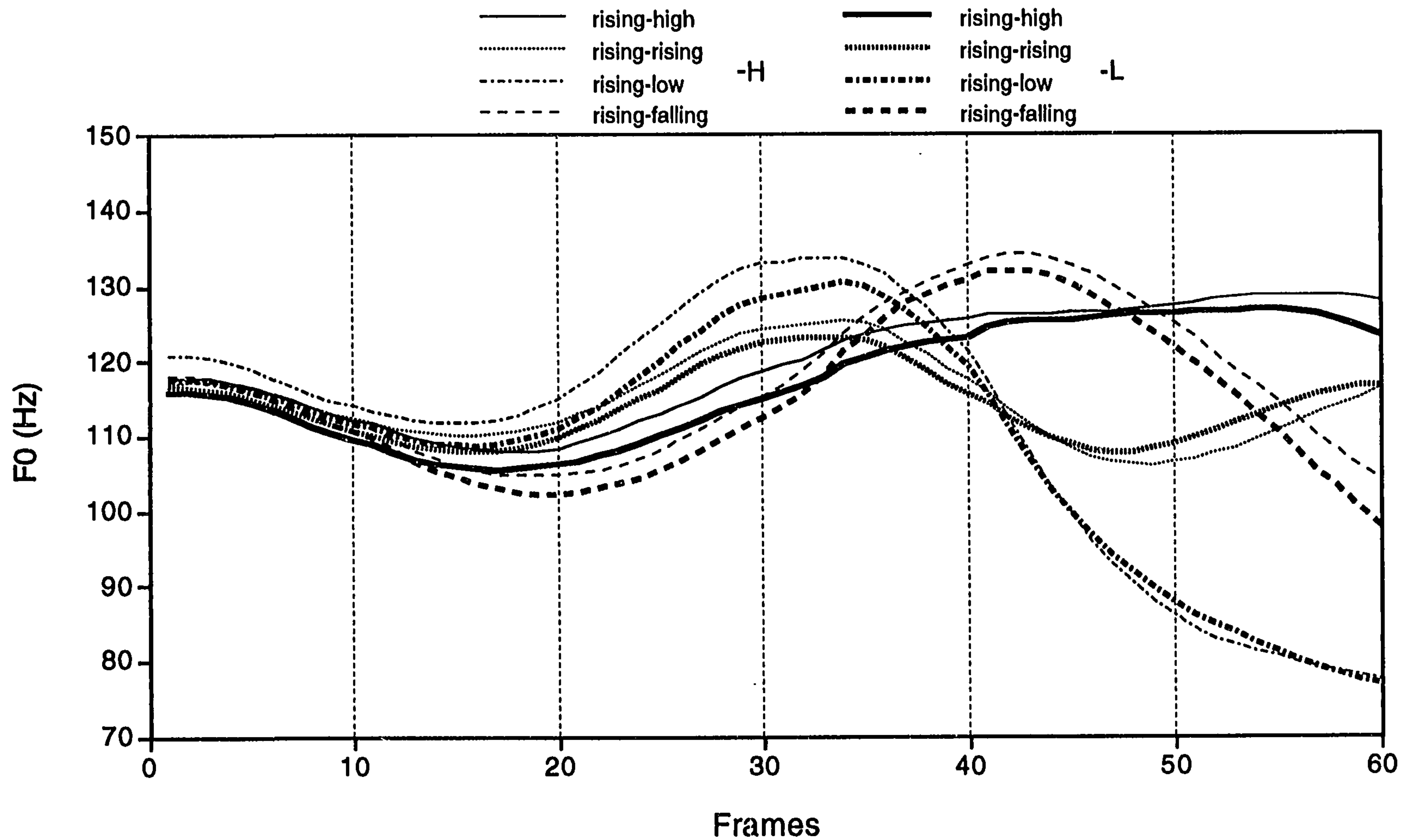


Figure 2.32. Mandarin rising tone under immediate and cross-syllabic anticipatory tonal influence.

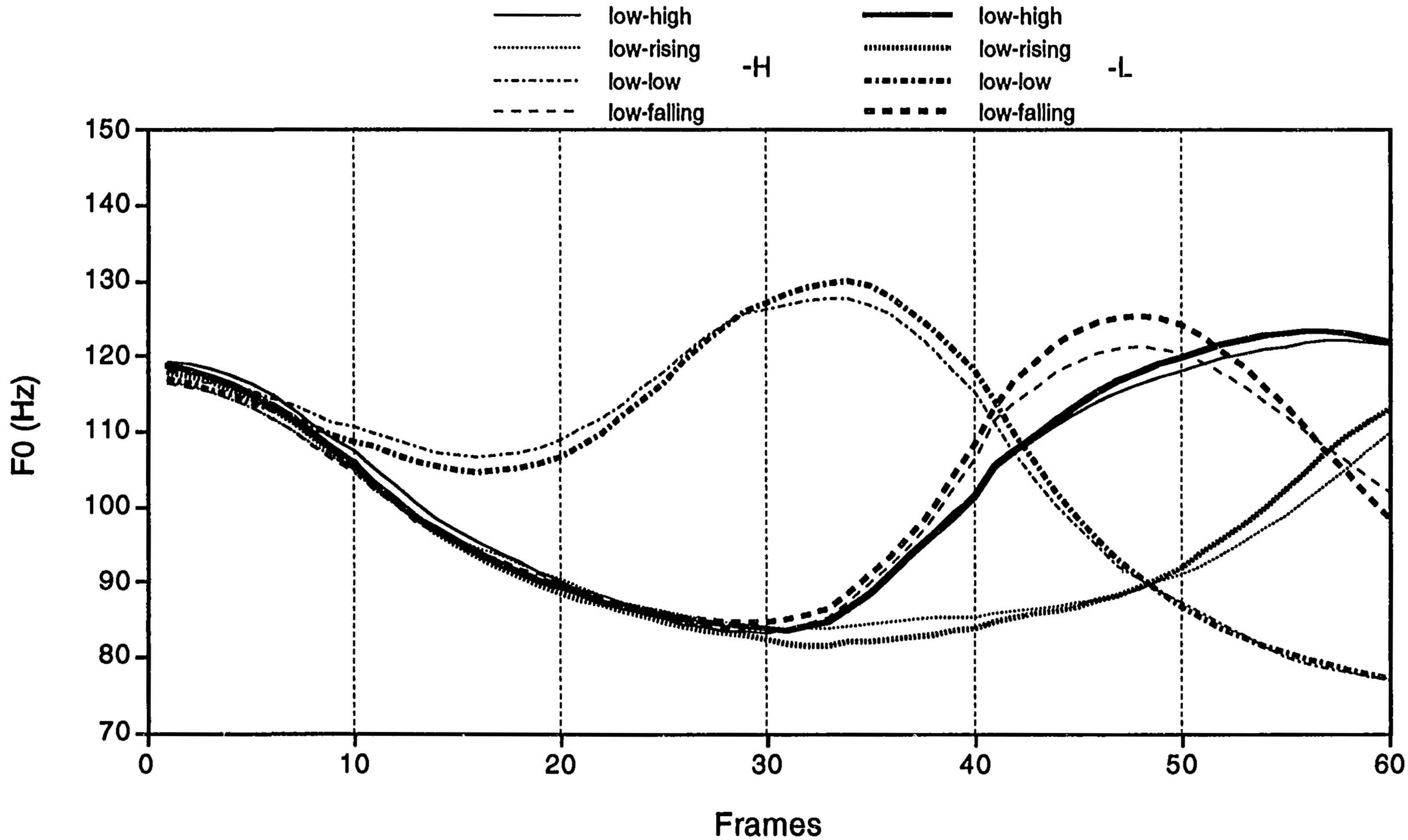


Figure 2.33. Mandarin low tone under immediate and cross-syllabic anticipatory tonal influence.

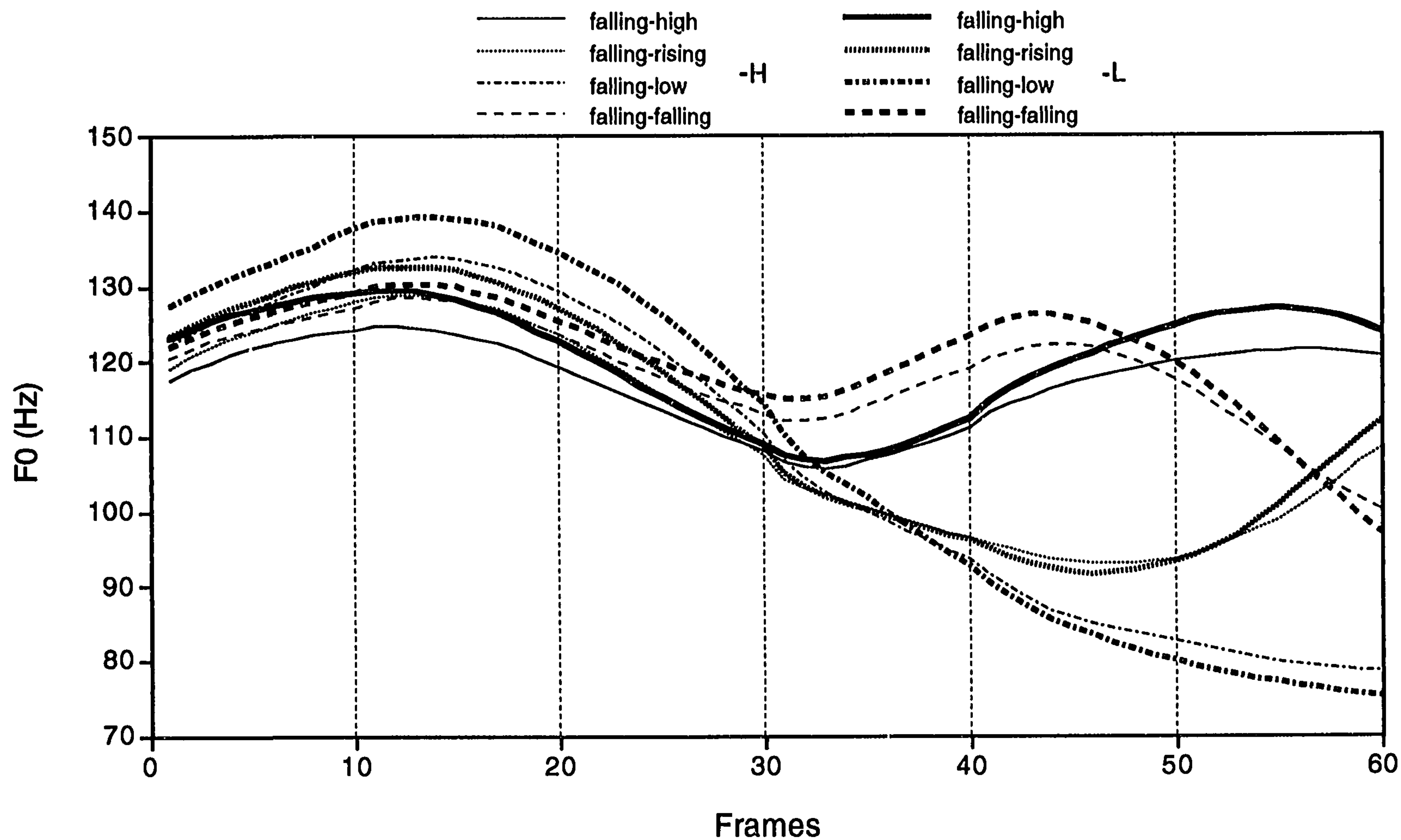


Figure 2.34. Mandarin falling tones under immediate and cross-syllabic anticipatory tonal influence.

From Figures 2.31-2.34, there does not seem to be a clear picture for the cross-syllabic anticipatory effect, or even a clear-cut picture of the immediate anticipatory effect exerted by the carrier on the second syllable in the /mama/ sequence. The largest anticipatory dissimilation is seen when the first syllable has a falling tone. Not only is the effect clearly seen in the second /ma/ for all the four tones, except for the low tone, but also the effect is extensive on the first /ma/. For each tone combination, the cross-syllabic effect of anticipatory dissimilation by the carrier is clearly seen on top of the dissimilatory effect exerted by the second syllable on the first syllable.

However, when the first /ma/ has a high tone, the effect is small and inconsistent. When the first /ma/ has a rising tone, the effect seems to be assimilatory rather than dissimilatory; when the second /ma/ has a low tone, an anticipatory dissimilation can be seen on that syllable, but not on the first /ma/.

Statistical analysis for cross-syllabic anticipatory effect

Several three-factor ANOVAs were conducted to examine the immediate as well as the cross-syllabic anticipatory effects. The independent variables are (1) Post-word Pitch — the starting pitch of the syllable in the carrier sentence immediately following the /mama/ sequence (Hi or Lo); (2) Post-pitch, the starting pitch of the second syllable in the /mama/ sequence (H or L); (3) Tone of the first syllable in the /mama/ sequence (high, rising, low, or falling). The dependent variables are Maxf0, Meanf0, Minf0, and Diff0.

Table 2.6 lists all the main effects and all the two-way interactions obtained in the ANOVAs except the interaction between Post-word Pitch and Post-pitch which is not significant for any of the F0 measurements. None of the three-way interactions are significant, and so are not listed here.

Table 2.6. ANOVA Results for cross-syllabic anticipatory effects.

		Maxf0		Meanf0		Minf0		Diff0	
Segment	**Effect**	**F**	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Nasal	Post-word Pitch	.62	.4347	.577	.4508	.50	.4822	.071	.7914
	Post-pitch	23.79	***.0001***	16.70	***.0002***	8.88	***.0044***	11.79	***.0012***
	Tone	3.35	**.0260**	3.64	**.0186**	5.25	***.0031***	3.13	**.0335**
	Post-word Pitch x Tone	3.50	**.0217**	3.87	**.0142**	3.54	**.0208**	.65	.5851
	Post-pitch x Tone	10.80	***.0001***	8.03	***.0002***	4.06	**.0115**	7.14	***.0004***
Vocalic	Post-word Pitch	.01	.9394	.01	.9340	.11	.7382	.15	.7005
	Post-pitch	70.90	***.0001***	60.15	***.0001***	10.06	***.0025***	50.78	***.0001***
	Tone	18.51	***.0001***	37.88	***.0001***	40.95	***.0001***	9.97	***.0001***
	Post-word Pitch x Tone	4.79	***.0051***	4.08	**.0112**	2.11	.1102	2.14	.1064
	Post-pitch x Tone	13.12	***.0001***	7.61	***.0003***	10.03	***.0001***	10.04	***.0001***

Total df = 52.

Table 2.6 shows that the effect of Post-pitch is significant on all the F0 measurements for both the nasal and vocalic segments, indicating a strong anticipatory effect on the tone of the first syllable, as was seen in the last section. Also as seen in the last section, the anticipatory effect mostly influences the part of a tone that has a high pitch. This is reflected in the significant interaction between Post-pitch and Tone for all the F0 measurements in both the nasal and vocalic segments shown in Table 2.6.

Figures 2.35-2.42 provide more details for the interactions. As seen in Figures 2.29 and 2.30, for the nasal segment, the anticipatory effect is greater on the falling tone and the high tone, both of which have a high pitch during the initial nasal segment, than the rising tone and the low tone, both of which have low pitch during the initial nasal segment.

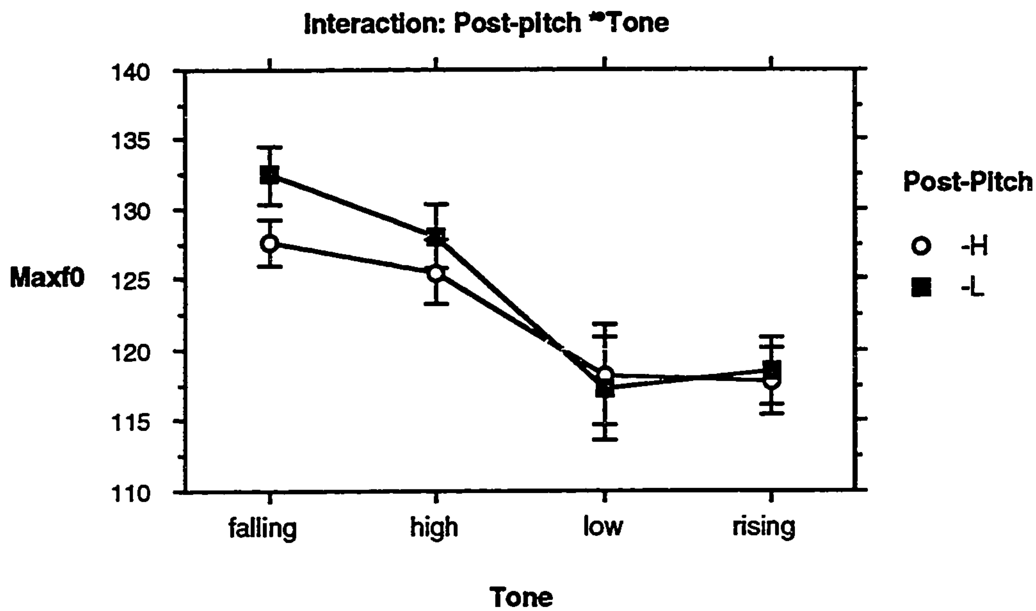


Figure 2.35. Interaction of Post-pitch and Tone for anticipatory effect on Maxf0 for the first nasal segment in the /mama/ sequences.

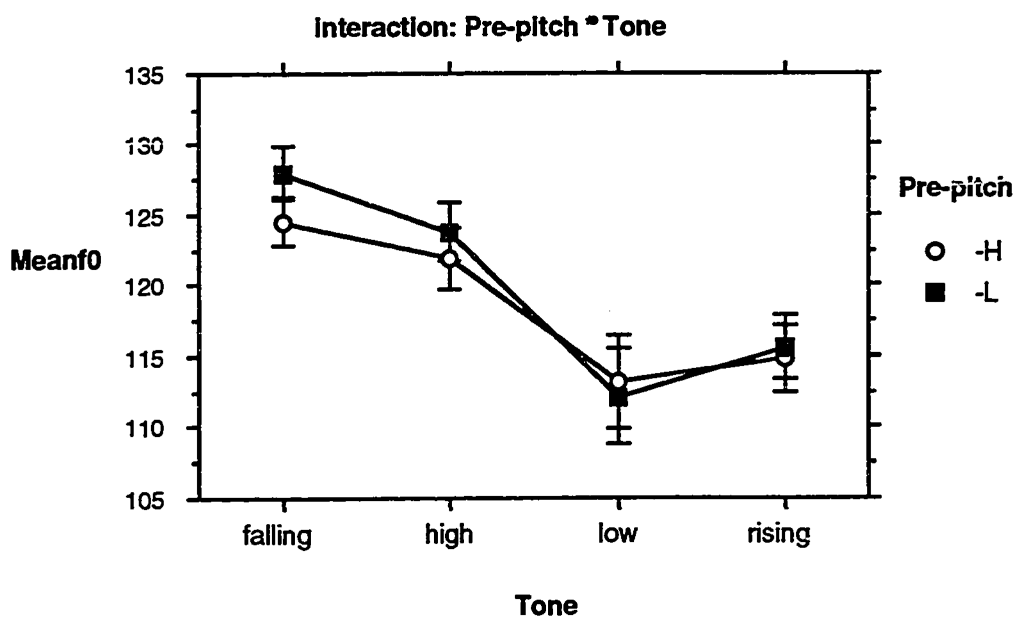


Figure 2.36. Interaction of Post-pitch and Tone for anticipatory effect on Meanf0 for the first nasal segment in the /mama/ sequences.

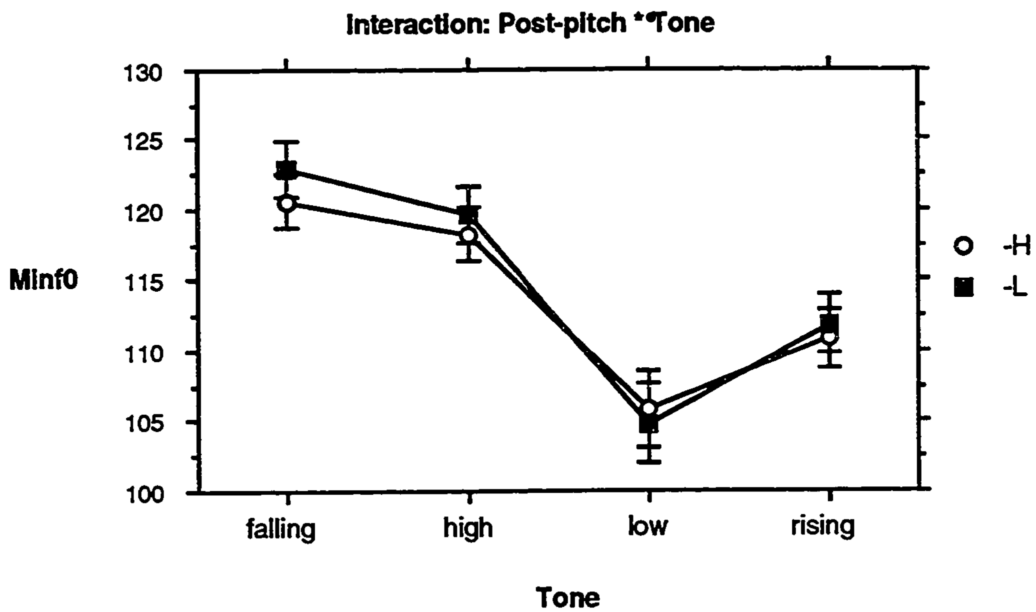


Figure 2.37. Interaction of Post-pitch and Tone for anticipatory effect on Minf0 for the first nasal segment in the /mama/ sequences.

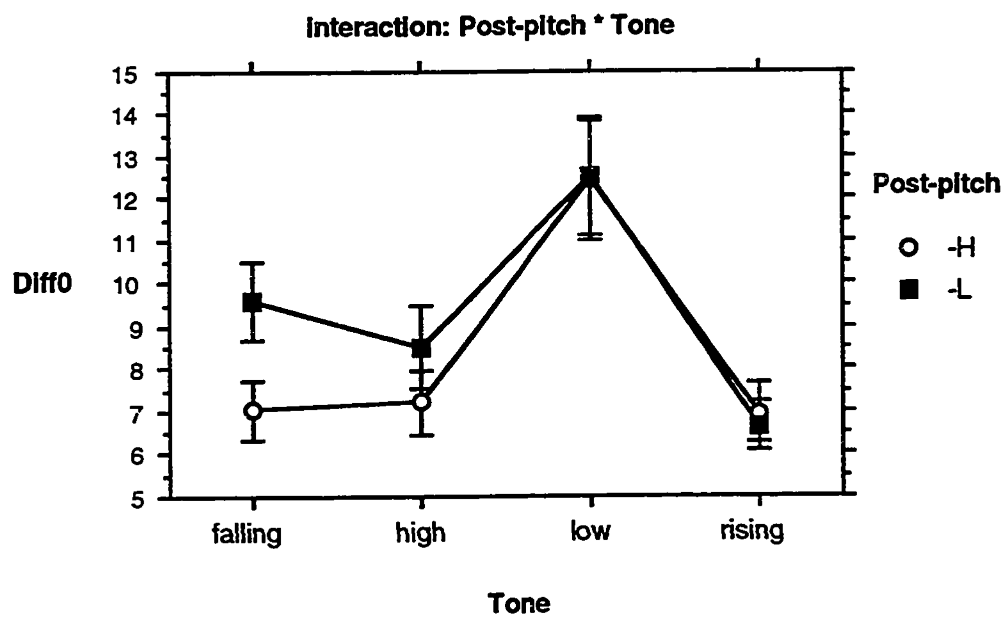


Figure 2.38. Interaction of Post-pitch and Tone for anticipatory effect on Diff0 for the first nasal segment in the /mama/ sequences.

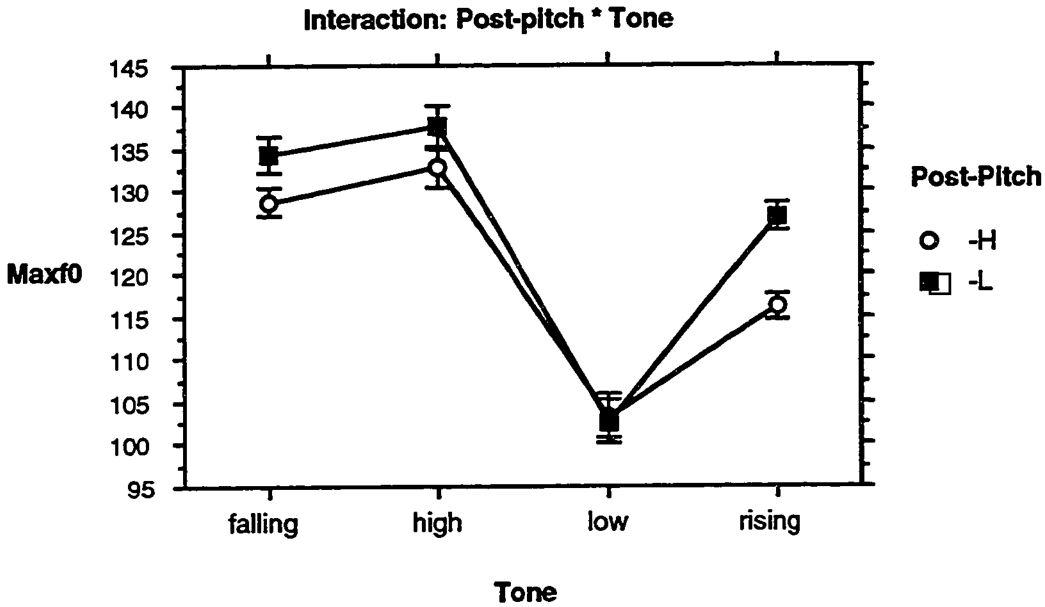


Figure 2.39. Interaction of Post-pitch and Tone for anticipatory effect on Maxf0 for the first vocalic segment in the /mama/ sequences.

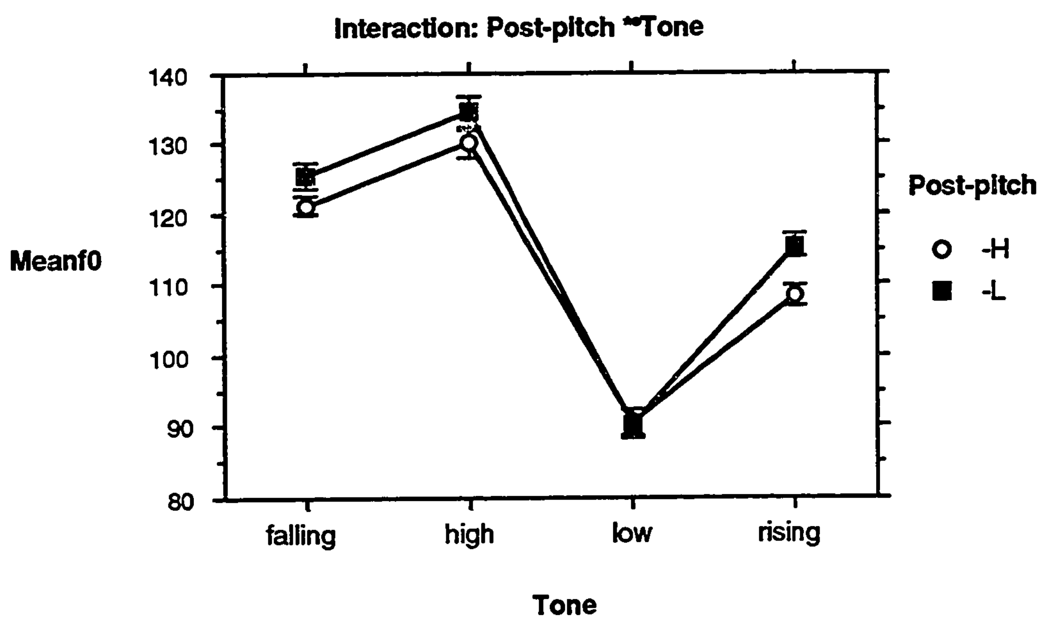


Figure 2.40. Interaction of Post-pitch and Tone for anticipatory effect on Meanf0 for the first vocalic segment in the /mama/ sequences.

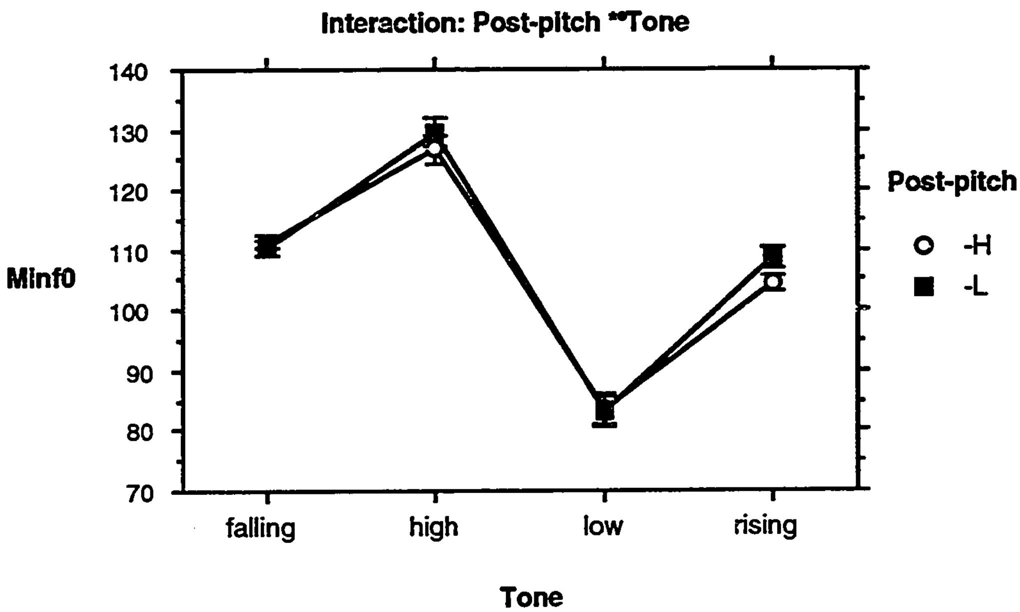


Figure 2.41. Interaction of Post-pitch and Tone for anticipatory effect on MinfO for the first vocalic segment in the /mama/ sequences.

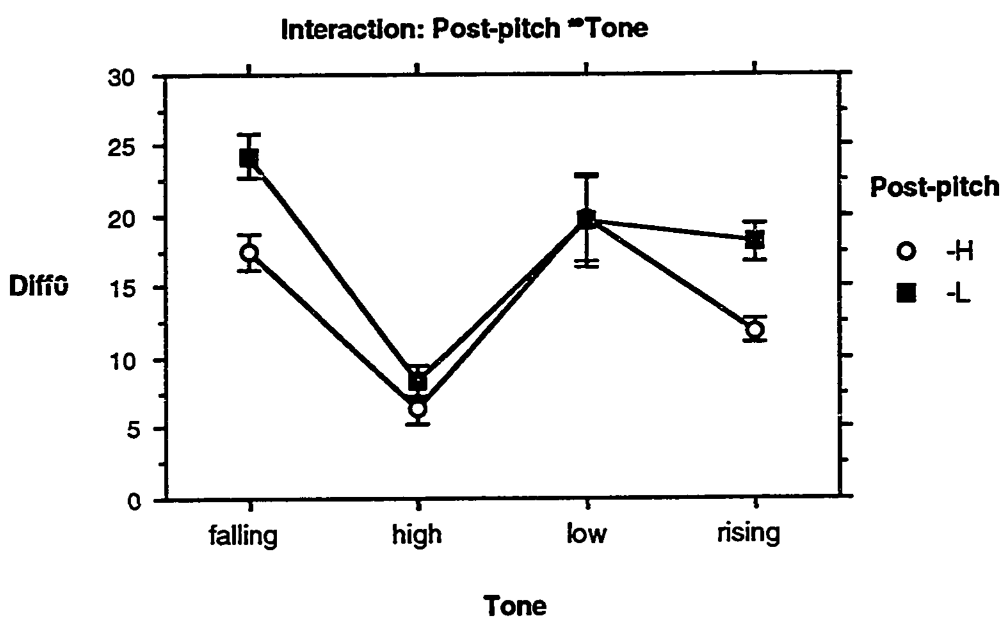


Figure 2.42. Interaction of Post-pitch and Tone for anticipatory effect on DiffO for the first vocalic segment in the /mama/ sequences.

For the vocalic segment, the dissimilatory effect can be clearly seen on Maxf0 and Meanf0 for the high, the rising and the falling tone, but not for the low tone. On the Minf0, Post-pitch makes some difference for the high and the rising tone, but little for the falling tone and the low tone, both of which have their lower extremes at the end of the vocalic segment. Post-pitch makes much difference on Diff0 for the falling tone and the rising tone, but very small difference for the high tone, and almost no difference for the low tone.

For the cross-syllabic anticipatory effect, Table 2.6 shows that the main effect of Post-word Pitch is not significant on any of the F0 measurements. However, there is significant interaction between Post-word Pitch and Tone on several F0 measurements. Figures 2.43 and 2.47 plot those significant interactions. In those figures, three patterns of effect can be seen for different tones. For the falling tone, there is strong dissimilatory effect; for the rising tone, there is moderate assimilatory effect; for the high and the low tones, however, there is almost no effect. It seems that the case of cross-syllabic anticipatory effect is not as clear-cut as that of the immediate anticipatory effect.

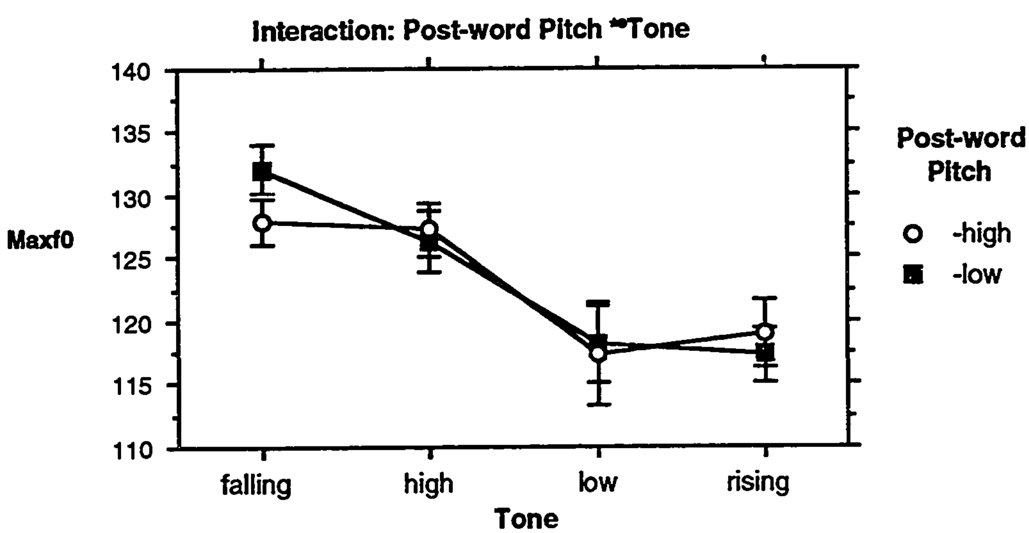


Figure 2.43. Interaction of Post-word Pitch and Tone for anticipatory effect on Maxf0 for the first nasal segment in the /mama/ sequences.

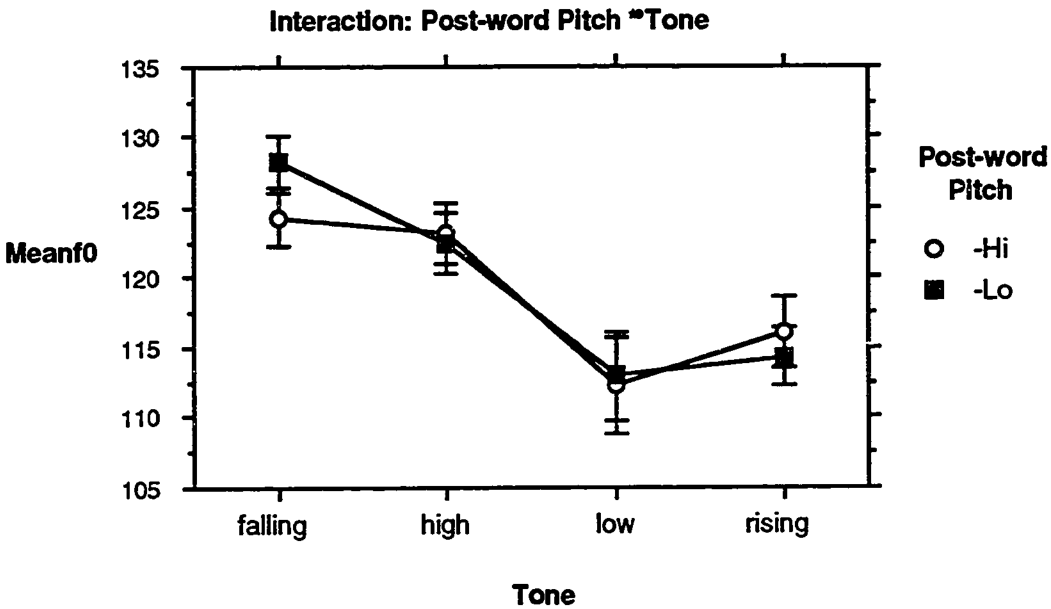


Figure 2.44. Interaction of Post-word Pitch and Tone for anticipatory effect on Meanf0 for the first nasal segment in the /mama/ sequences.

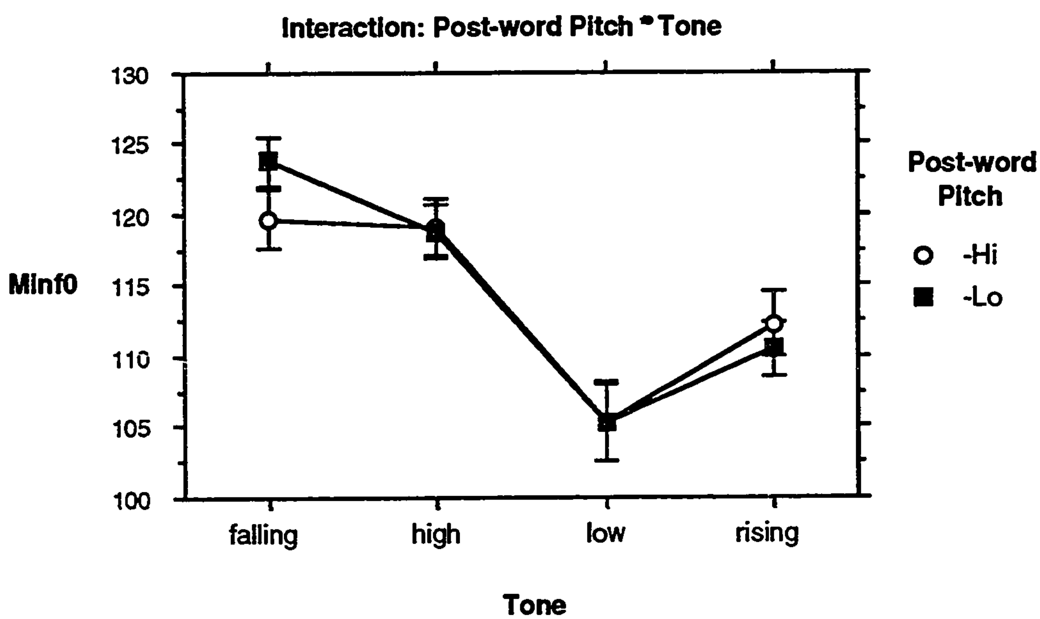


Figure 2.45. Interaction of Post-word Pitch and Tone for anticipatory effect on Minf0 for the first nasal segment in the /mama/ sequences.

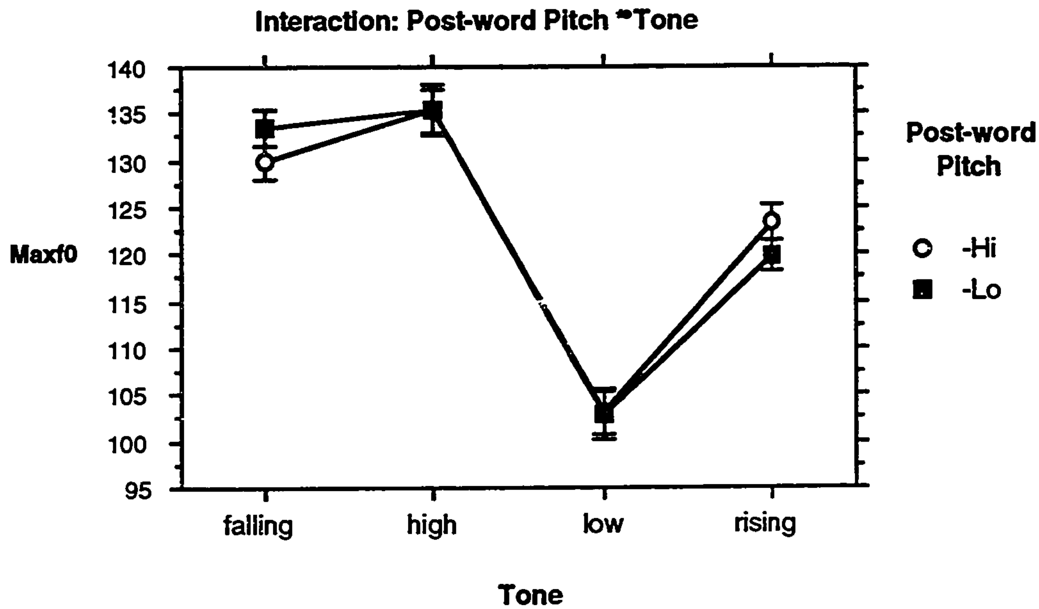


Figure 2.46. Interaction of Post-word Pitch and Tone for anticipatory effect on Maxf0 for the first vocalic segment in the /mama/ sequences.

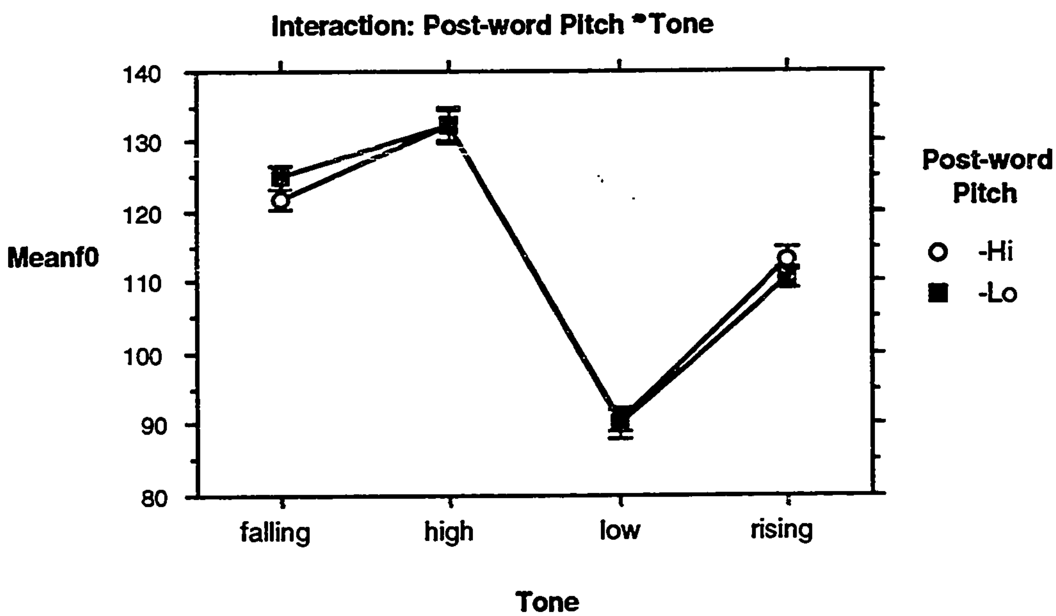


Figure 2.47. Interaction of Post-word Pitch and Tone for anticipatory effect on Meanf0 for the first vocalic segment in the /mama/ sequences.

Summary of findings about anticipatory effect

To sum up briefly the findings about anticipatory tonal variation, surprising as it may seem, it is quite convincing that the immediate anticipatory effect within the /mama/ sequences is dissimilatory. More precisely, a tone that has a lower starting pitch tends to raise the maximum as well as the overall F0 of the preceding tone, except in cases where the preceding tone is a low tone. For the cross-syllabic anticipatory effect, the case is not yet clear. From the data examined in the previous section, it seems that the effect is not very consistent: dissimilatory for the falling tone, assimilatory for the rising tone, and little effect for the high and the low tones. Further study is needed to clear this up.

Evidence from Other Studies for Anticipatory Dissimilation

The most interesting finding in the present study of contextual tonal variations is the anticipatory tonal dissimilation. This phenomenon has never been systematically reported before. However, as mentioned at the beginning of this chapter, there has been some evidence in the data collected by several studies. One of them is by Shih (1986). She measured the starting and ending F0 of the four Mandarin lexical tones in disyllabic words. The following table shows her measurements of the F0 values of the first syllable when followed by different tones.

Table 2.7. Starting and ending F0 values of the Mandarin syllable /fu/ in four tones when followed by different tones as measured by Shih (1986).

	fū	fú	fǔ	fù
Isolation	266-266	211-253	214-154	287-159
Before high tone	258-270	219-245	213-176	299-218
Before rising tone	274-291	216-257	223-168	300-227
Before low tone	273-290	223-281	——	310-238
Before falling tone	268-286	216-243	225-178	300-227

Although her speaker's Mandarin is that of Taiwan, the F0 values shown in this table are surprisingly similar to those obtained in the present study.

Shen (1990, p. 285) also noticed that the high and the rising tones "have the highest overall tonal values" when preceding the rising and the low tones.

In a recent study (Gandour, 1992b), anticipatory tonal dissimilation was found for the falling tone in Thai. Although no systematic conclusion was reached concerning anticipatory dissimilation because the falling tone was the only tone checked for anticipatory effect when followed by different tones, the trend is similar to what is seen in the present study.

More interestingly, the raising of F0 before a low tonal target seems to be not unique to the East Asian tone languages. Recently, a similar phenomenon was found in one of the African tone languages, Yoruba (Connell & Ladd, 1990; Laniran, 1992). It would be interesting to examine in further studies if those two phenomena found in East Asian languages and in African languages are due to similar principles.

Exploring the Nature of Anticipatory Dissimilation

Systematic tonal dissimilation has almost never been discussed as a phonetic process. The literature provides us with only occasional discussion over some aspects of the dissimilation process similar to those found in the present study. There are two existing accounts for some of the aspects of the anticipatory dissimilation.

a) *Phonological deletion.* This account is given by Shih (1986, p. 10), who states,

> when a H target follows, the rising slope of a tone 2 turns out to be the interpolation from its M target to the H of the next syllable, rather than to its own H. This phenomenon suggests that the final H of a tone 2 is deleted, or absorbed by the next H target. Tone 2 seems to end lower in this context only because the value at the syllable boundary does not represent the H target, but a point on the

rising slope. When the following target is L, the position of the final H is unaffected and the pitch value at the end of the syllable reflects the H value.

This deletion account seems to take into consideration only the anticipatory dissimilation in the rising tone. The anticipatory tonal dissimilation in Mandarin, however, seems to apply to all the tones that have at least one high pitch target: high, rising and falling tones. It would be hard to argue, from the data of the present study or even from the data in Shih (1986), that either of the two high targets in the high tone is deleted when followed by the rising or the low tone. It is even harder to attribute the F0 variation in the falling tone to deletion of the high target when it is followed by a high or a falling tone. Beside the apparent inadequacy in explaining dissimilation, the deletion account also has the potential of being too powerful, because it is not restrained by any implicit physiological or perceptual constraint.

b) *Wide swing.* This account is due to Gandour (1992b, p. 121), who states that when a falling tone is followed by a low tonal target in Thai,

> ... the transition is from a tone that traverses the high region of the voice range to tones that initially traverse the low region. This transition requires complex adjustments of the vocal folds. Because of vocal fold dynamics, one may speculate it is easier in some articulatory sense to move from an even higher F0 to an extremely low F0. This vocal fold adjustment is analogous to what happens when a semi-trailer swings wide to make a sharp right or left turn. The extra wide turn facilitates the movement from a street going in one direction to a street cutting off at a 90 degree angle. The anticipatory effects on the slope of the preceding falling tone are believed to follow as a consequence of the adjustments in height. From a given height to a fixed F0 onset, the slope must necessarily be steeper from a higher F0. Back to the semi-trailer analogy, the angle of the turn varies as a direct consequence of the wider swing around the corner.

This account compares transition between pitch registers to the turning of a semi-trailer. Following this analogy, a low-to-high F0 transition should be just the same as a high-to-low transition, which means that, not only should a high pitch target be raised by a following low pitch target, but also a low pitch target should be lowered by a following high pitch target. However, neither is the final F0 in the Mandarin falling and low tones affected by the starting pitch of the following tone (cf. Figure 2.30), nor is the final F0 of the Thai falling tone lowered by the high initial F0 of a following falling tone (Gandour, 1992b). The only support for the wide-swing account may be found in Figure 2.29 (b), in which the minimum F0 of the rising tone is seen lowered by the initial high target of the following tone. In general, however, the wide-swing account cannot explain why the high pitch range is affected more than the low pitch range by the following tone.

It was shown earlier that minimum F0 of the low and the falling tones was affected neither by the anticipatory effect nor by the carryover effect, suggesting that a low tonal target is more resistant than a high tonal target to contextual influences. It seems that an adequate account for anticipatory tonal dissimilation ought to be able to explain why carryover assimilation as well as anticipatory dissimilation affect the high pitch range more than the low pitch range. This taken into consideration, two alternative accounts are suggested as follows.

c) *Limit of low pitch register.* According to this account, it is more difficult to reach a low tonal target than a high tonal target due to articulatory constraints. As an alternative, the upper range in the preceding tone is extended. Supporting evidence for this account is in the finding that a low final pitch in a tone (as in a low or falling tone) is not affected by either the carryover or the anticipatory effect, as discussed earlier in this chapter, indicating that the low pitch range is much less flexible than the high pitch range. Physiological support for this assumption may be found in Erickson (1976). She observed that while the production of a low pitch involves the activities of the strap muscles (mainly, thyrohyoid, sternohyoid and sternothyroid), those muscles only contribute actively to lowering F0 when the pitch is to drop below a threshold level, usually near the midrange. This indicates that to reach the lower pitch range, extra effort by the speaker may be needed. Thus, the lower a pitch target, the more difficult it is to reach it. Since the low target for the low tone is lower than that of the rising tone, greater effort is needed to expand the upper range of the previous tone, and hence greater raising effect is exerted by the low tone than by the rising tone, as is evident in Figures 2.21, 2.22, 2.29, and 2.30.

A potential problem with this account is that the highest F0 in the falling tone rises only about as high as when it is produced in isolation. See Figure 2.48 for comparison of the falling tone produced in isolation and when followed by the low tone. It may be the case that it is not the low or rising tones that raise the F0 value of the preceding tone, but rather, the high starting pitch of the high or falling tones that lowers the F0 value of the preceding tone. Or it may be the case that both processes, i.e., the lowering effect by a following high starting pitch and the raising effect by a following low starting pitch, are at work simultaneously.

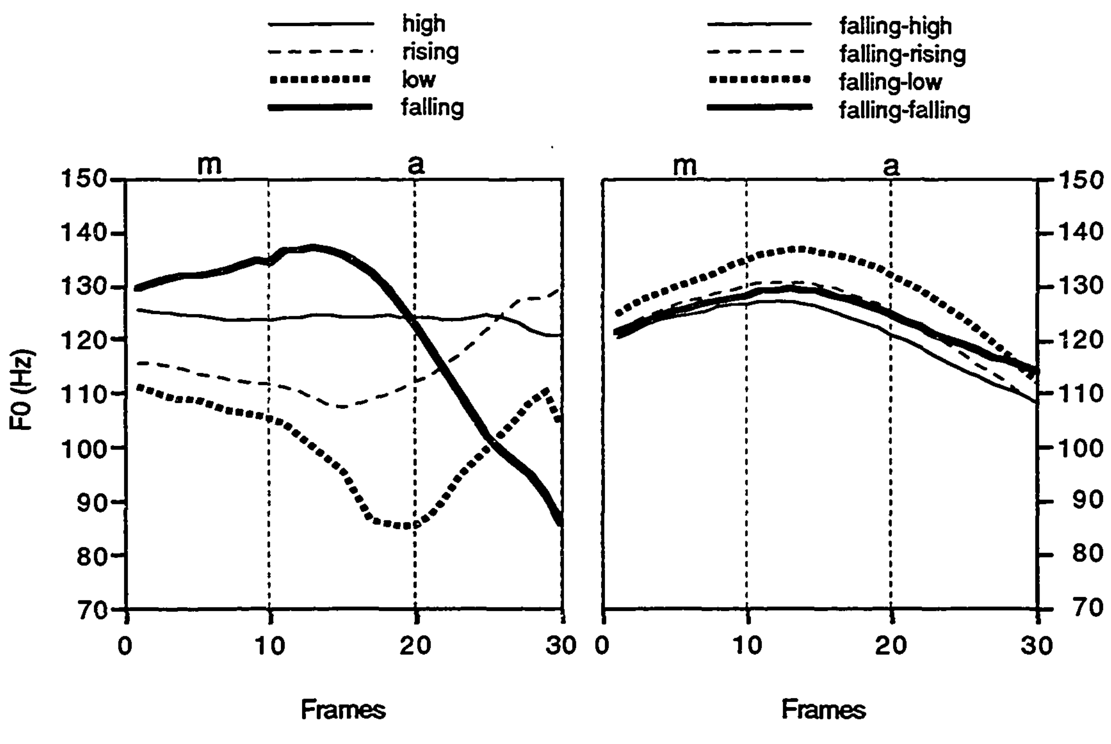


Figure 2.48. Comparison of the falling tone produced in isolation (left) and when followed by different tones (right).

d) *Counteraction of declination.* Alternatively, the anticipatory dissimilation serves to counteract declination, "the tendency of pitch to drift downwards over the course of an intonation group," (Pierrehumbert, 1979, p. 363), which has been found in many languages including Mandarin (Maeda, 1976; Gårding, 1987; 't Hart & Cohen, 1973). Because the normal declination is already going from high to low, to produce the tonal differences, there is a need to make the tonal patterns distinct from the declination contour. When the tonal pattern is HH or LH across the tonal boundary, it is already different from the intonation. When the tonal pattern is HL or HM across the tonal boundary, however, there is the potential of confusing the tonal pattern with the declination pattern. To reduce this potential confusion, the difference between the H and L target is exaggerated. However, due to the physical limit of the lower threshold, this exaggeration is accomplished by fully implementing the H target rather than by lowering the L target. An indirect support for this hypothesis is the finding by Pierrehumbert (1979) that when two stressed syllables are heard as equal in pitch, the second is actually lower, indicating that speakers normalize for declination in judging the relative height of peaks in the intonation contour. Thus, in the case of tone perception, this normalization process has the potential of reducing or even washing out the actual relative pitches in case the pitch sequence across the tone boundary is HL or HM.

A problem with this account, however, is that according to the hypothesis, the likelihood of tonal differences being confused with declination should be greater in the sequence HM than in HL. Yet, in the F0 analysis, it was found that the F0 values were higher before a low tone than before a rising tone, and the low tone has a lower F0 value than the rising tone when produced in isolation. A possible explanation might be found in another finding by Pierrehumbert (1979, p. 363) that "a greater correction for declination was made for wide pitch range stimuli than for narrow pitch range stimuli." Since the pitch range in HL is greater than that in HM, the speaker expects greater correction by the listener, hence, he would make greater effort in counteracting this expected correction.

Pending further investigation, the last two accounts seem equally plausible for the time being, because both can accommodate most of the data. Everything considered, account (c) is slightly more favored over account (d), because the latter has to presume speaker's sensitivity to F0 declination, which is yet to be proven.

A Note about the Implementation of Tone Sandhi Rule

As discussed at the beginning of this thesis, although the perceptual experiments by Wang and Li (1967) have demonstrated that for listeners, the rising tone derived from a low tone when followed by another low tone is perceptually equivalent to the real rising tone produced in the same tonal context, the acoustical equivalence of the two has been questioned by some studies (Shen, 1990; Zee, 1980a). It is noticed in those studies that the derived rising tone is not exactly the same as the real rising tone in the same context. The data in the present study seem to provide more solid evidence for the difference between the real and the derived rising tones, as seen in the following statistical comparison on the acoustic data.

Table 2.8 shows the mean values of different F0 measurements on the original rising tone and derived rising tone (Figures 2.9 (a) and 2.12 shown earlier can be referred to for actual differences in F0 contours). Each mean is an average over 80 tokens (5 repetitions x 4 carriers x 4 speakers). In Table 2.8, the mean value is greater for the original rising tone than for the derived rising tone on Maxf0, and Meanf0 in both the nasal and the vocalic segments. On Diff0, the derived rising tone has the greater means than the original rising tone in both segments.

Table 2.8. Mean values of different F0 measurements on original and derived rising tones.

	Maxf0		Meanf0		Minf0		Diff0	
Segment	Original	Derived	Original	Derived	Original	Derived	Original	Derived
Nasal	118.88	117.20	115.16	111.94	111.89	107.32	6.99	9.88
Vocalic	123.00	114.61	111.91	101.44	106.14	94.46	16.31	21.53

A group of ANOVAs were conducted to compare the difference between the derived and the original rising tones, and the results are shown in Table 2.9. The independent variables are (1) Pre-word Pitch (Hi / Lo) and (2) Original Tone of the first /ma/ (rising / low). The dependent variables are the F0 measurements, Maxf0, Meanf0, Minf0, and Diff0. The inclusion of Pre-word Pitch as an independent variable is because it was seen before that Pre-word Pitch has a strong effect on the tone of the first syllable. However, because there is no interaction between Pre-word Pitch and any of the F0 measurements, only the effect of Tone is listed in Table 2.9.

Table 2.9. ANOVA Results for differences between the original and the derived rising tone.

	Maxf0		Meanf0		Minf0		Diff0	
Segment	F	P	F	P	F	P	F	P
Nasal	.42	.5291	2.10	.1695	6.20	.0260	3.63	.0774
Vocalic	11.43	*.0045*	36.81	*.0001*	25.30	*.0002*	1.93	.1864

Total df = 14

As shown in Table 2.9, the difference between the original and the derived rising tone is highly significant on Maxf0, Meanf0 and Minf0 in the vocalic segment, but the difference is significant only on MinfO in the nasal segment. The difference on Diff0 is not significant in either of the two segments.

The results shown here may have interesting implications for our understanding of speech production and perception. A recent study by Zsiga (1993) demonstrates very convincingly that in English the /s/ variant before a /j/ is articulatorially different from a lexical /ʃ/, although the former has been previously believed to be equivalent to the latter after palatal assimilation. She finds that, in contrast to the constant palatal articulation for a lexical /ʃ/, a /s/ before /j/ is gradient during the course of its articulation, changing gradually from very close to [s] to very close to [ʃ]. It is not known yet whether the /s/ before /j/ is perceptually equivalent to the lexical /ʃ/ in English because her study does not include perceptual experiments. It could be the case that, similar to the so-called derived high tone in Mandarin as shown in Chapter 1 of this thesis, a /s/ before /j/ in English is not perceived as a lexical /ʃ/. Whichever is eventually found to be the case, Zsiga's (1993) finding nevertheless raises questions about our previous understanding of phonological rules. What is intriguing about the case of tone sandhi is that there has already been a study (Wang & Li, 1967) showing that the derived rising tone and the original rising tone are perceptually equivalent. With the present finding that the two are not exactly the same acoustically, the apparent indication is that even in case of a true phonological change, the representation of the identity of the underlying form is maintained to some extent even at surface phonetic output. It has been suggested that no post-lexical phonological rules are categorical (Liberman & Pierrehumbert, 1984).The current finding about the phonetic implementation of the low-tone sandhi rule provides further support for that proposal.

Summary of Chapter 2

In this chapter, both assimilation and dissimilation are found in contextual tonal variation in Mandarin. It was found that carryover contextual effects, i.e., influence of the preceding tone on the following tone, are assimilatory in nature, whereas anticipatory effects are mostly dissimilatory. In both cases, it is either the starting or the ending pitch value rather than the entire tonal curve of a tone that influences the neighboring tones. While a low ending value in a preceding tone tends to lower the F0 value of the following tone, a low starting value in a following tone tends to raise the F0 value of the preceding tone. Carryover effects were found to influence not only the immediately adjacent tone, but also the second and probably the third tone in a row. Anticipatory effects, however, did not show any consistent cross-syllabic influence.

While the assimilatory tonal variation is similar to previous findings about tonal coarticulation, anticipatory tonal dissimilation has not been discussed systematically before. Several accounts for the anticipatory tonal dissimilation are discussed, including phonological deletion, wide-swing, limit of low pitch register, and counteraction of declination.

The finding of acoustical difference between the original rising tone and the rising tone derived from the low tone before another low tone due to the tone sandhi rule raises serious questions about our current understanding of phonological alteration and phonetic implementation in speech production.

Chapter 3 — Toward a More Unified View of Tonal Variation in Mandarin

Explaining Contextual Tonal Variation

In Chapters 1 and 2, various aspects of the production and perception of contextual tonal variations in Mandarin are examined. Results of the experiments and analyses either confirm previous findings or reveal phenomena that have not been observed or have not been systematically observed before. In the course of analyzing the data, pieces of observations made in previous studies as well as in the present one concerning contextual variation of Mandarin tones seem to be gradually converging to form a picture with much more consistency than previously found possible. The following is a list of the various aspects of the emerging unified view about Mandarin tonal variation.

1) Tones, just like segments, undergo variations due to different phonological and phonetic processes in speech production.

2) In general, a tone is influenced by both carryover effect and anticipatory effect. The two kinds of effect are different in nature, and thus have asymmetrical influences on a tone.

3) The carryover tonal influence is assimilatory, and it is the ending pitch rather than any other part of a tone that is exerting the carryover influence. A high ending pitch of a tone raises the onset F0 as well as the overall F0 height of the immediately following tone as well as the next following tone, and possibly even the third tone in the row.

4) The anticipatory tonal influence is dissimilatory, and it is the starting pitch of a following tone that is exerting the anticipatory influence. A low or mid starting pitch raises the maximum F0 as well as the overall F0 height of the preceding tone except the low tone.

Unlike carryover effect, however, no consistent cross-syllabic anticipatory effect was found in the present study.

5) The magnitude of assimilatory tonal variation is positively related to speaking rate but negatively related to the level of stress of the tone being affected.

6) Listeners are able to compensate for assimilatory tonal variations rather successfully when the tonal context is present. Listeners' performance in identifying the perturbed tones is not directly proportional to the magnitude of the perturbation.

7) The perceptual effect of anticipatory tonal dissimilation is unknown. Further studies are needed to address this question.

With the above list, it is now possible to provide more systematic explanations for many phenomena found in earlier studies as well as in the present one.

a) The phonetic mechanism for Chao's rising-tone variation rule is the asymmetry of anticipatory and carryover effects. The carryover effect is assimilatory, while the anticipatory effect is dissimilatory. The magnitude of the two effects is negatively related to the stress level of the tone being influenced. Since the second syllable in a trisyllabic sequence is usually prosodically weak, it is subject to strong influence from the preceding tone as well as the following tone. Looking again at Figure 1.3, for the rising tone, we see that its onset, as well as its overall F0, is raised by the high ending pitch of a preceding tone; at the same time, its maximum, as well as its overall F0, is raised by the low starting pitch of a following tone. As a result of the combined influences, both the onset F0 and the overall F0 of the rising tone are greatly elevated, resulting in a contour with high overall F0 and slightly negative slope. When a rising tone is preceded by a tone with low ending pitch and followed by a tone with high starting pitch, its onset and maximum F0 height are lowered by the preceding tone, and the overall F0 height is further lowered by the following tone, resulting in a rising contour with low overall F0 height. Similarly, for the falling tone, its F0 onset, as well as its overall F0 height, is lowered by the low ending pitch of the preceding tone. At the same time, both its maximum and its overall F0 height are lowered by the high starting pitch of the following tone. Consequently, the falling tone becomes a slightly falling contour with rather low overall F0 height. When the falling tone is preceded by a tone with high ending pitch and followed by a tone with low starting pitch, its onset is raised by the preceding tone, and its overall F0 height is boosted by the following tone, resulting in a high falling contour with rather steep negative slope.

In case a rising tone is preceded and followed by high pitches in the surrounding tones, the high pitch that precedes it elevates its onset F0, and the high pitch that follows reduces its maximum F0 in the final rise. As a result, the surface form approaches a flat contour without much of a rising slope.

b) When a falling tone is followed by another falling tone, it is said to have only a half fall (Chao, 1948). Shen (1990) argues that the half fall of a falling tone before another falling tone results from tonal coarticulation rather than from tone sandhi. From Figures 2.22 (b) and 2.30 (b), however, it seems that the falling tone falls to roughly the same height by the end of the syllable carrying it regardless of what the following tone is. Shih (1986, p. 12) also reports that the L target of a non-final falling tone "only reaches the middle pitch range regardless of the following tonal context." As listed in Table 2.7 of the present thesis, her measurements show that the ending F0 of the falling tone is more or less the same when followed by different tones (actually the highest before the low tone, the lowest before the high tone, and the same before the rising and the falling tones).

If there is not much difference in the ending point of a falling tone, where then did Chao's impression of the half fall come from? The findings of anticipatory dissimilation in this study show that a falling tone before another falling tone has lower maximum as well as overall F0 than before a rising or low tone. So, it is highly likely that the impression of a half fall is the result of hearing the reduced falling contour due to anticipatory dissimilation.

c) Shih (1986, p. 11) reports that "following a H target, a tone 3 consistently starts higher than a tone 2, which is unexpected since both tones have the same initial target." In Figures 2.21 and 2.29 of the present thesis, the same phenomenon can be clearly seen: when the first syllable has a high or rising tone, a following low tone would have a higher starting F0 than a following rising tone. Furthermore, however, both the rising and the low tone in the second syllable have a higher starting F0 than the high and the falling tone. When the two phenomena are considered together, it is clear that it is anticipatory dissimilation that is responsible for the difference in the onset F0 of the second syllable. When the second syllable has a tone with low starting pitch, the ending F0 of the preceding high or rising tone is elevated to such a degree that the raised F0 contour extends into the initial nasal segment of the second syllable. In other words, the anticipatory tonal dissimilation is reflected back onto the onset of the influencing tone. Because the low tone has an even lower starting pitch than the rising tone, the amount of elevation of the ending F0 of the preceding tone is larger when followed by a low tone than by a rising tone, and thus there is greater reflection of the anticipatory dissimilation for the low tone than for the rising tone.

d) Shih (1986, p. 11) observes,

> Tonal alignment in disyllabic words is basically the same as in monosyllabic words. The most important difference is the absence of the first tonal target at the very beginning of the second syllable. That is exactly where two different tonal targets may clash. By removing the first target, some transitional time is allowed during the consonantal region. When the second syllable begins with a glide, the tonal transition occurs mostly during the glide. In cases where the second syllable begins with a vowel, transitional time is pushed into the vowel.

The F0 analyses in Chapter 2 of the present study agrees with Shih's observation in that the fastest transition between the end of the first tone and the beginning of the second tone occurs during the initial voiced segment of the second tone. The range of ending F0 values of a tone is much narrower than the range of its onset F0 values. This is why the second tone seems to Shih (1986) to lack an initial tonal target.

Presumably, this phenomenon is due to different mechanisms for the carryover and anticipatory effects. On the one hand, since the carryover effect is a form of coarticulation, and since tonal coarticulation is apparently uni-directional, i.e., only from left to right, whenever a compromise between two neighboring tones is called for, the later tone simply gives in to the earlier one due to lack of assimilating power. As a result, the later tone simply starts from where the previous tone ends, and then proceeds from there to approach its own pitch targets. On the other hand, the anticipatory effect seems not to directly affect how two neighboring tones join one another, but rather it is more like a process of adjusting the relative F0 height and range of the previous tone. Hence, it does not seem to affect the ending F0 of the previous tone in particular. As a result, the ending range of a tone looks much narrower than its starting range.

It is also observable in the F0 curve tracings in the present study (see Figures 2.29 and 2.32) that when the pitch heights disagree across a syllable boundary, the first turning point occurs sometimes not at the onset of the initial nasal segment of the second syllable, but somewhere after the onset of the initial nasal segment, sometimes even as late as half way through the initial nasal segment. Presumably, this is due to inertia of the articulatory movement. By the end of the syllable, although the articulation for the tone associated with that syllable may have stopped, the movement toward the last goal of a tone sometimes cannot stop instantaneously. Thus we see an extension of that movement into the initial segment of the next syllable.

Remaining Questions

Although the findings in the present study make the picture about contextual tonal variations in Mandarin much clearer than it has been shown previously, a lot of relevant details of the picture are still lacking.

First, it is still not very clear how exactly anticipatory and carryover effects interact with stress. Although it is fairly certain that the amount of variation of a tone is negatively related to the stress level of the syllable carrying it, it is still not clear how carryover and anticipatory effects interact with stress in determining the surface shape of a tone.

Second, the effect of speaker's intention to speak clearly and the effect of deliberate emphasis are both believed to contribute to the final F0 contour of a tone. It is also relevant to our understanding of the entire speech communication process. Unfortunately, few studies have been conducted regarding these effects, especially the first one (but see Picheny, Durlach, & Braida, 1986).

Third, the domain of the anticipatory and carryover effects needs to be further explored. It is found in the present study that carryover effects can reach beyond the adjacent syllable, but its influence beyond the next following syllable can only be inferred from the fact that the difference it causes remains till the end of the next following tone. More solid evidence is definitely needed for the long distance tonal coarticulation. As for the anticipatory effect, the inconsistency in the effect across a prosodic boundary needs to be solved, or satisfactory explanations need to be found.

Finally, the exact mechanism for the anticipatory dissimilation needs to be further investigated. Although it is argued in Chapter 2 in favor of the limit-of-low-pitch-register account or the counteraction-of-declination account, more supporting evidence is definitely needed. Or, if there isn't any, alternative accounts should be considered.

Appendix 1

Original Words and Response Choices for Experiment 1

shānhúdǎo	zhuōmícáng	zǐdìbīng	zǒnglùxiàn	nǎoshénjīng	shǒuliúdàn	cūzhìpǐn	fāngxiàngpán	← Original Words
珊瑚岛	捉迷藏	子弟兵	总路线	脑神经	手榴弹	粗制品	方向盘	
chānhudǎo	shuōmicáng	sǐdibīng	zǒngludiàn	nǎoshendīng	shǒuliucàn	tūzhipǐn	bāngxiangpán	
搀忽导	说眯藏	死低冰	总撸电	恼伸丁	守溜灿	突知品	帮香盘	Response Choices
搀胡导	说弥藏	死迪冰	总炉电	恼神丁	守刘灿	突直品	帮祥盘	
搀虎导	说米藏	死底冰	总鲁电	恼审丁	守柳灿	突止品	帮想盘	
搀护导	说蜜藏	死第冰	总陆电	恼甚丁	守六灿	突治品	帮象盘	
sānjíguǎn	dīniánjí	zǒngbàgōng	tǔdìmiào	běijíxīng	xiǎobáicài	jīxièshǒu	jīqìrén	
三极管	低年级	总罢工	土地庙	北极星	小白菜	机械手	机器人	
zānjiguǎn	dīnianxí	dǒngbagōng	zǔdimiào	féijixīng	xiǎobaitài	qīxieshǒu	dīqirén	
簪机馆	低拈习	董八功	祖低妙	匪机兴	晓掰太	七歇守	低期仁	
簪集馆	低黏习	董拔功	祖迪妙	匪集兴	晓白太	七斜守	低齐仁	
簪几馆	低碾习	董把功	祖底妙	匪几兴	晓百太	七写守	低起仁	
簪计馆	低念习	董霸功	祖第庙	匪计兴	晓败太	七谢守	低气仁	
xuélíngqián	luóxuánjiǎng	bànshìchù	jìniànbēi	dànbáizhì	hàoqíxīn	zájìtuán	máoxìguǎn	
学龄前	螺旋桨	办事处	纪念碑	蛋白质	好奇心	杂技团	毛细管	
juélingqián	luóxuanxiǎng	fànshichù	xìnianbēi	sànbaizhì	gàoqixīn	dájituán	máoxihuǎn	
决拎乾	罗宣想	饭失触	系拈悲	散掰制	告期新	答机团	茅西缓	
决灵乾	罗悬想	饭实触	系年悲	散白制	告奇新	答级团	茅习缓	
决领乾	罗选想	饭使触	系碾悲	散摆制	告起新	答几团	茅洗缓	
决令乾	罗镟想	饭是触	系念悲	散败制	告器新	答计团	茅戏缓	
shíniánqián	báipíjiǔ	qìxiàngzhàn	shìjièguān	zìliúdì	xiàngpígāo	rénzàogé	lánlùhǔ	
十年前	白啤酒	气象站	世界观	自留地	橡皮膏	人造革	拦路虎	
chīnianqián	báipixiǔ	tìxiangzhàn	zhìjieguān	sìliudì	jiàngpigāo	shénzaogé	lánlugǔ	
持拈乾	白批朽	替香战	治接关	四溜第	降披高	神遭格	兰撸古	
持黏乾	白疲朽	替祥战	治节关	四刘第	降疲高	神凿格	兰炉古	
持碾乾	白痞朽	替想战	治姐关	四柳第	降痞高	神早格	兰鲁古	
持念乾	白僻朽	替向战	治借关	四六第	降僻高	神躁格	兰陆古	

Appendix 2

Original Words and Response Choices for Experiment 2

shānhúdǎo	zhuōmícáng	zǐdìbīng	zǒnglùxiàn	fēijīchǎng	zhōngyīxué	fǎxīsī	xiǎnwēijìng	← Original Words
珊瑚岛	捉迷藏	子弟兵	总路线	飞机场	中医学	法西斯	显微镜	
chānhuzǎo	shuōmitáng	sǐdipīng	dǒngludiàn	bēijitǎng	chōngyijué	bǎxizī	jiǎnweidìng	
搀忽早	说眯堂	死低乒	董撸电	悲基躺	冲衣决	把希资	简微定	Response Choices
搀胡早	说迷堂	死敌乒	董卢电	悲及躺	冲移决	把习资	简围定	
搀虎早	说米堂	死底乒	董鲁电	悲几躺	冲以决	把洗资	简委定	
搀护早	说秘堂	死第乒	董陆电	悲计躺	冲义决	把细资	简卫定	

sānjíguǎn	dīniánjí	zǒngbàgōng	tǔdìmiào	jiāotōngjǐng	wēifēnxué	hǎoyīshēng	běifānghuà
三极管	低年级	总罢工	土地庙	交通警	微分学	好医生	北方话
zānjikuǎn	jīniandí	dǒngbahōng	zǔdimiào	xiāotongdǐng	fēifenjué	gǎoyizhēng	fěifangkuà
簪基款	机拈敌	董八轰	祖低妙	肖通顶	飞分决	稿衣争	匪方跨
簪及款	机年敌	董拔轰	祖敌妙	肖同顶	飞坟决	稿移争	匪房跨
簪几款	机碾敌	董把轰	祖底庙	肖统顶	飞粉决	稿以争	匪访跨
簪计款	机念敌	董爸轰	祖第妙	肖痛顶	飞奋决	稿义征	匪放跨

xuélíngqián	luóxuánjiǎng	bànshìchù	jìniànbēi	shíhuīshí	yuánzhūbǐ	bànfēngjiàn	jiàokēshū
学龄前	螺旋桨	办事处	纪念碑	石灰石	圆珠笔	半封建	教科书
juélingxián	luóxuanxiǎng	fànshizhù	dìnianfēi	chíhuīzhí	xuánzhupǐ	fànfengxiàn	xiàokezhū
决拎闲	罗宣想	饭失柱	第拈飞	持挥直	悬朱痞	饭风现	效棵朱
决灵闲	罗悬想	饭十柱	第年飞	持回直	悬竹痞	饭冯现	效壳朱
决领闲	罗选想	饭使柱	第碾飞	持毁直	悬主痞	饭讽现	效可朱
决令闲	罗镟想	饭是柱	第念飞	持汇直	悬助痞	饭奉现	效客朱

shíniánqián	báipíjiǔ	qìxiàngzhàn	shìjièguān	kuánghuānjié	báihuānxǐ	dàduōshù	jiùshēngyī
十年前	白啤酒	气象站	世界观	狂欢节	白欢喜	大多数	救生衣
zhínianxián	páipixiǔ	tìxiangshàn	zhìjiehuān	huánghuāndié	páihuanqǐ	sàduozhù	xiùshengdī
直拈闲	排批朽	替香善	治阶欢	黄欢叠	排欢起	萨多柱	秀升低
直年闲	排疲朽	替详善	治节欢	黄环叠	排环起	萨夺柱	秀绳低
直碾闲	排痞朽	替想善	治解欢	黄缓叠	排缓起	萨朵柱	秀省低
直念闲	排僻朽	替向善	治界欢	黄换叠	排换起	萨舵柱	秀胜低

References

Abramson, A. S. (1962). *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments.* Bloomington: Indiana University Research Center in Anthropology, Folklore, and Linguistics, Pub. 20.

Abramson, A. S. (1976). Thai tones as a reference system. In T. W. Gething, J. G. Harris, & P. Kullavanijaya (Eds.), *Thai linguistics in honor of Fang-Kuei Li* (pp. 1-12). Chulalongkorn University Press: Bangkok.

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, **21**, 319-325.

Abramson, A. S. (1979a). The coarticulation of tones: An acoustic study of Thai. In T. L. Thongkum, P. Kullavanijaya, V. Panupong, & K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugénie J. A. Henderson* (pp. 1-9). Bangkok: Chulalongkorn University Press.

Abramson, A. S. (1979b). The noncategorical perception of tone categories in Thai. In B. Lindblom & S. Öhman (Eds.), *Frontiers of speech communication research* (pp. 127-134). Academic Press: London.

Bai, D. (1934). Guanzhong shengdiao shiyan lu [Experiments with tones of Guanzhong dialects]. In *Shiyusuo Jikan [A Collection by Shiyusuo]* (pp. 355-361).

Benguerel, A.-P., & Cowan, H. A. (1974). Coarticulation of upper lip protrusion in French. *Phonetica*, **30**, 41-55.

Browman, C. P., & Goldstein, L. M. (1985). Dynamic modeling of phonetic structure. In V. A. Fromkin (Ed.), *Phonetic Linguistics* (pp. 35-53). New York: Academic Press.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook*, **3**, 219-252.

Catford, J. C., Jusczyk, P. W., Klatt, D. H., Liberman, A. M., Remez, R. E., & Stevens, K. N. (1991). Panel discussion: The motor theory and alternative accounts. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and The Motor Theory of Speech Perception* (pp. 192-194.). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

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

Chao, Y. R. (1948). *Mandarin Primer.* Cambridge: Harvard University Press.

Chao, Y. R. (1956). Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. In M. Halle (Eds.), *For Roman Jakobson* (pp. 52-59). Mouton: The Hague.

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

Chistovich, L. A., & Kozhevnikov, V. A. (1969). Some aspects of the physiological study of speech. In L. D. Proctor (Ed.), *Biocybernectics of the Central Nervous System*, Boston: Little, Brown.

Chuang, C. K., Hiki, S., Sone, T., & Nimura, T. (1971). The acoustical features and perceptual clues of the four tones of standard colloquial Chinese. In *Seventh International Congress on Acoustics*, 25 C 13 (pp. 297-300). Budapest.

Connell, B., & Ladd, D. R. (1990). Aspects of pitch realization in Yoruba. *Phonology*, **7**, 1-29.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, **24**, 579-606.

Cooper, F. S., Liberman, A. M., Harris, K. S., & Grubb, P. M. (1958). Some input-output relations observed in experiments on the perception of speech. In *Proceedings of The Second International Congress of Cybernetics*. Namur, Belgium.

Erickson, D. M. (1976). *A Physiological Analysis of the Tones of Thai*. Doctoral Dissertation, The University of Connecticut.

Fang, Z. (1990). Tone sandhi and tone perception. *Acta Psychologica Sinica*(3), 255-259.

Fant, G. (1960). *Acoustic Theory of Speech Production*. 's-Gravenhage: Mouton.

Fant, G. (1973). *Speech Sounds and Features*. Cambridge, Massachusetts: M.I.T. Press.

Fowler, C. A. (1977). *Timing control in speech production* (Indiana University Linguistics Club, Bloomington, Indiana).

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, **8**, 113-133.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, **36**, 359-368.

Fowler, C. A. (1987). Perceivers as realists, talkers too: commentary on papers by Strange, Diehl et al., and Rakerd and Verbrugge. *Journal of Memory and Language*, **26**, 574-587.

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 123-139). Hillsdale, NJ: LEA.

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

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, **11**, 149-175.

Gandour, J. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics*, **12**, 235-293.

Gandour, J. (1992a). Tonal coarticulation in Thai disyllabic utterances: a preliminary study. *Linguistics of the Tibeto-Burman Area* , **15**, 93-110.

Gandour, J. (1992b). Anticipatory tonal coarticulation in Thai noun compounds. *Linguistics of the Tibeto-Burman Area* , **15**, 111-124.

Gårding, E., & Zhang, J. (1986). Tone 4 and Tone 3 discrimination in Modern Standard Chinese. *Language and Speech*, **29**, 281-293.

Gårding, E. (1987). Speech act and tonal pattern in Standard Chinese. *Phonetica*, **44**, 13-29.

Han, M. S., & Kim, K. (1974). Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics*, **2**, 223-232.

Henke, W. L. (1966). *Dynamic articulatory model of speech production using computer simulation*. Ph. D dissertation, Massachusetts Institute of Technology.

Howie, J. M. (1970). *The Vowel and Tones of Mandarin Chinese: Acoustical Measurements and Experiments*. Doctoral Dissertation, Indiana University.

Howie, J. M. (1972). Some experiments on the perception of Mandarin tones. In A. Rigault & R. Charbonneau (Ed.), *The 7th International Congress of Phonetic Sciences*, (pp. 900-904). Mouton: The Hague.

Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica*, **30**, 129-148.

Keating, P. A. (1985). CV phonology, experimental phonetics, and coarticulation. *UCLA Working Papers in Phonetics*, **62**, 1-13.

Keating, P. A. (1988). Underspecification in phonetics. *Phonology*, **5**, 275-292.

Kent, R., & Minifie, F. (1977). Coarticulation in recent speech production models. *Journal of Phonetics*, **5**, 115-133.

Kratochvil, P. (1984). Phonetic tone sandhi in Beijing dialect stage speech. *Cahiers de Linguistique, Asie Orientale*, **13**, 135-174.

Laniran, Y. (1992). *Intonation in Tone Languages: The phonetic Implementation of Tones in Yorùbá.* Doctoral Dissertation, Cornell University.

Liberman, A. M., Cooper, F. S., Harris, K. S., MacNeilage, P. F., & Studdert-Kennedy, M. G. (1967). Some observations on a model for speech perception. In W. Wathen-Dunn (Eds.), *Models for the perception of speech and visual form.* Cambridge, Massachusetts: M.I.T. Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. G. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1952). The role of stimulus variables in the perception of stop consonants. *American Journal of Psychology*, **65**, 497-516.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oehrle (Eds.), *Language Sound Structure* (pp. 157-233). Cambridge, Massachusetts: M.I.T. Press.

Lin, H.-B. (1988). *Contextual Stability of Taiwanese Tones*. Doctoral Dissertation, The University of Connecticut.

Lin, H.-B., & Repp, B. H. (1989). Cues to the perception of Taiwanese tones. *Language and Speech*, **32**, 25-44.

Lin, M. (1965). Yingao xianshiqi yu Putonghua shengdiao yingao texing [The pitch indicator and the pitch characteristics of tones in Standard Chinese]. *Acta Acoutica Sinica*, **2**, 8-15.

Lin, M. (1988). Putonghua shengdiao de shengxue texing he zhijue zhengzhao. *Zhongguo Yuwen [Chinese Linguistics]*, **204**, 182-193.

Lin, M., & Yan, J. (1991). Tonal coarticulation patterns in quadrasyllabic words and phrases of Mandarin. In *Proceedings of the XIIth International Congress of Phonetic Sciences*.Vol.3 (pp. 242-245).

Lin, T., & Wang, W. S. Y. (1984). Shengdiao ganzhi wenti [Perception of tones]. *Zhongguo Yuyan Xuebao [Bulletin of Chinese Linguistics]*, **2**, 59-69.

Liu, F. (1924). *Sisheng Shiyan Lu [Experiments with tones]*. Shanghai: Qunyi Shushe.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review*, **77**, 182-196.

MacNeilage, P. F. (1985). Planning and production of speech: an overview. *Journal of the American Speech and Hearing Association*, **15**, 15-21.

Maeda, S. (1976). *A Characterization of American English Intonation*. Doctoral Dissertation, MIT, Cambridge.

Magen, H. (1989) *An Acoustic Study of Vowel-to-Vowel Coarticulation in English*. Ph.D. dissertation, Yale University.

Mann, V. A. (1980). Influence of a preceding liquid on stop consonant perception. *Perception & Psychophysics*, **28**, 407-412.

Mann, V. A., & Repp, B. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, **28**, 213-228.

Moll, K. L., & Daniloff, R. G. (1971). Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America*, **50**, 678-684.

Moore, C. B. (1993). Phonetic observations on tone and stress levels in Mandarin classifiers. In *Proceedings of the 5th North American Conference on Chinese Linguistics*, (p. 29). Newark, Delaware: University of Delaware.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America*, **75**, 224-230.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, 29, 434-446.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, **66**, 363-369.

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

Recasens, D. (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America*, **76**, 1624-1635.

Recasens, D. (1985). Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. *Language and Speech*, **28**, 97-114.

Shen, X. S. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics*, **18**, 281-295.

Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin Tone 2 and Tone 3. *Language and Speech*, **34**, 145-156.

Shih, C. (1986). *The phonetics of the Chinese tonal system.* Technical memorandum. AT & T Bell Laboratories.

Shih, C., & Sproat, R. (1992). Variations of the Mandarin rising tone. In *Proceedings of the IRCS Workshop on Prosody in Natural Speech No. 92-37*, (pp. 193-200). Philadelphia: The Institute for Research in Cognitive Science, University of Pennsylvania.

't Hart, J., & Cohen, A. (1973). Intonation by rule: a perceptual quest. *Journal of Phonetics*, **1**, 309-327.

Tseng, C. (1981) *An Acoustic Phonetic Study on Tones in Mandarin Chinese*. Doctoral Dissertation, Brown University.

Wang, W. S.-Y., & Li, K.-P. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research*, **10**, 629-636.

Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, **18**, 3-35.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, **49**, 25-47.

Whalen, D. H., Wiley, E. R., Rubin, P. E., & Cooper, F. S. (1990). The Haskins Laboratories' pulse code modulation (PCM) system. *Behavior Research Methods, Instruments & Computers*, **22**, 550-559.

Winer, B. J. (1962). *Statistical Principles in Experimental Design* (2nd ed.). New York: McGraw-Hill.

Wright, S., & Kerswill, P. (1989). Electropalatography in the analysis of connected speech processes. *Clinical Linguistics & Phonetics*, **3**, 49-57.

Wu, Z. (1984). Putonghua sanzizu biandiao guilü [Rules of tone sandhi in trisyllabic words in Standard Chinese]. *Zhongguo Yuyan Xuebao [Bulletin of Chinese Linguistics]*, **2**, 70-92.

Wu, Z. (1988). Tone-sandhi patterns of quadri-syllabic combinations in Standard Chinese. *Report of Phonetic Research, Institute of Linguistics (CASS), Beijing, China*, **PL-ARPR/1988**, 1-13.

Xu, Y. (1991). Depth of phonological recoding in short-term memory. *Memory & Cognition*, **19**, 263-273.

Yang, Y. (1989). Yuanyin he shengdiao zhijue [Perception of Chinese vowels and tones]. *Acta Psychologica Sinica*, (1), 29-34.

Zahorian, S. A., & Jagharghi, A. J. (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *Journal of the Acoustical Society of America*, **94**, 1966-1982.

Zee, E. (1980a). The effect of aspiration on the F0 of the following vowel in Cantonese. *UCLA Working Papers in Phonetics*, **49**, 90-97.

Zee, E. (1980b). A spectrographic investigation of Mandarin tone sandhi. *UCLA Working Papers in Phonetics*, **49**, 98-116.

Zsiga, E. C. (1993). *Features, Gestures, and the Temporal Aspects of Phonological Organization.* Ph. D dissertation, Yale University.