Production and perception of coarticulated tones

Yi Xu

Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511 and Department of Linguistics, U-145, The University of Connecticut, Storrs, Connecticut 06269

(Received 26 August 1992; revised 3 September 1993; accepted 3 November 1993)

In the present study, the distinctive tones of Mandarin were found to coarticulate with adjacent tones in running speech. However, the amount of deviation of a tone from its canonical form due to coarticulation varied depending on the nature of the tonal context. In a context where adjacent tonal values agree (a "compatible" context), the deviation was relatively small. In a context where adjacent tonal values disagree (a "conflicting" context), the deviation was much greater, sometimes even to the extent of changing the direction of a dynamic tone. To examine the perception of coarticulated tones, naturally produced words and phrases were used as stimuli for tone identification. However, their semantic information was removed through waveform editing while all the tonal information contained in the signal remained intact. Identification of tones in the compatible context was highly accurate with or without the original tonal context. Tonal identification for the conflicting context remained accurate only when the tones were presented with the original tonal context. Without the original context, i.e., in isolation, correct tone identification dropped below chance. When the original tonal context was altered, listeners compensated for the altered context as if it had been there originally. It thus seems that, in tone perception, listeners compensate for variations due to coarticulation. Nevertheless, even with the presence of the original context, perceptual identification was better for the compatible context than for the conflicting context, indicating that variation due to coarticulation is not always completely compensated for by listeners.

PACS numbers: 43.71.Hw, 43.70.Fq, 43.71.An

INTRODUCTION

Tones in tone languages, such as Thai, Vietnamese, or Mandarin, have been shown to coarticulate with one another in connected speech (Abramson, 1979; Han and Kim, 1974; Wu, 1984, 1988; Shih, 1987; Shen, 1990) just like segments. Wu (1984, 1988) finds that in trisyllabic words in Mandarin, a tone varies its *F*0 contours in different tonal contexts. Shih (1987), in examining disyllabic tone sequences in Mandarin, finds that under the influence of the preceding tone, the initial target of a tone becomes a transition between the preceding tone and the following tone. Shen (1990) reports that in trisyllabic sequences in Mandarin, tones are affected by both carryover and anticipatory coarticulations, and that the bidirectional effects are symmetric. She also finds that in Mandarin, not only are tonal onsets and offsets affected by coarticulation, but entire tonal contours are shifted up or down by the surrounding tones. There has not been general agreement, however, on the extent to which the contour of a tone can deviate from its canonical form due to coarticulation. While it has been reported that both the onset and offset of the tone and the overall tonal height are affected by the tonal contexts for Vietnamese and Mandarin (Han and Kim, 1974; Shen, 1990), it is also suggested that tonal coarticulation does not affect *F*0 direction (Shen, 1990). Wu (1984), however, finds that under certain circumstances (e.g., when preceded by a high tone and followed by a low tone), a rising tone becomes what he calls a "downward gliding."

The effects of tonal coarticulation on perception are less well documented. Lin and Wang (1984) found that the perceived tone of the first syllable in a disyllabic sequence can depend on the *F*0 of the second syllable. Han and Kim (1974) also found that two different tones having overlapping tonal contours and *F*0 height in different tonal environments could still be correctly identified when presented in the original tonal contexts.

The present study provides more evidence for production and perception of tonal coarticulation by taking a closer look at the contextual tonal variations in Mandarin Chinese. In Mandarin, words are composed of morphemes that are, in most cases, monosyllabic. A monosyllabic morpheme in Mandarin 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 vocalic portion of a syllable. There are four stressed lexical tones and one neutral tone in Mandarin. The four stressed tones have the pitch contours high, rising, falling–rising, and falling when the syllables carrying them is produced in isolation.[1] In connected speech, some of the lexical tones may change their tonal category in tonal context. For example, the low tone in Mandarin changes into a rising tone when followed by another low tone. When that happens, the derived rising tone is perceptually indistinguishable from the lexical rising tone (Wang and Li, 1967). This kind of categorical tone shift due to tonal context is usually referred to as "tone sandhi."

A systematic investigation of contextual variations requires comparison between phonetic variations caused by different phonetic environments. The present study distin-

 0001-4966/94/95(4)/2240/14/$6.00

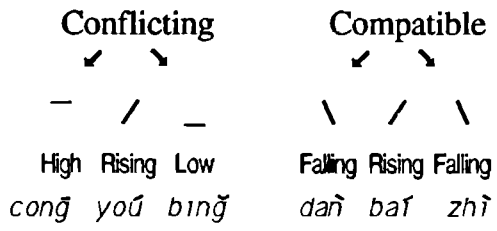

FIG. 1. Two kinds of phonetic environments—compatible and conflicting contexts.

guishes between two kinds of phonetic context, namely, "compatible" and "conflicting." A "compatible" context is an environment in which adjacent phonetic units have 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 Mandarin word *congyoubing*[2] "green onion pancake" (see Fig. 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 Mandarin word *danbaizhi*, "protein," has compatible contexts at both syllable boundaries, because in both places the pitch values are similar.

In the word *congyoubing*, the ending pitch[3] in the tone of the first syllable is high, while the starting pitch in the tone of the second syllable is low;[4] likewise, the starting pitch in the tone of the third syllable is low, while the ending pitch in the tone of the second syllable is high. To produce the rising tone in the second syllable in connected speech, the articulators responsible for *F*0 control would need to change very quickly from one state to another at the syllable boundaries if the underlying tonal values were to be fully realized within the second syllable. Not only is this rather difficult due to inertia of the articulatory muscles, but it is also unnecessary because it is possible to have the neighboring phonetic units coproduced, i.e., produced with temporal overlap (Fowler, 1977). The consequence of temporal overlap would be a compromise between adjacent phonetic units in the actually realized surface values, especially when they differ from one another substantially. Even in careful speech, 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 stressed the syllables in trisyllabic sequences evenly. In more fluent speech, the overlap may be more extensive. In that case, the combined influence of the first tone and the last tone on the second tone could be in the direction of flattening its contour. In a compatible context, however, 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. Therefore, for the two Mandarin words shown in Fig. 1, it is predicted that the one with conflicting contexts, i.e., *congyoubing*, is likely to have a somewhat flattened tonal contour for the second syllable; and the one with compatible contexts, i.e., *danbaizhi*, is likely to preserve its rising tonal contour for the second syllable.

Note that the kind of coarticulation discussed here is somewhat different from the classic cases of coarticulation. 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 definitely defined. Explained in terms of feature spreading, the specified feature of the influencing segment is said to spread to the influenced segment (Daniloff and Hammarberg, 1973). In terms of coproduction or gestural overlap, different gestures may overlap in time (Fowler, 1977; Browman and Goldstein, 1986). As a result, acoustic consequences produced by a gesture for the influenced segment may co-occur with acoustic consequences produced by gestures for the influencing segment. In the conflicting context defined here, however, the adjacent tones have opposing values for the same feature or gesture. In case there is any compromise, the overlap would have to be at a more abstract level, that is, the surface value of the influenced tone would be approximately of a weighted average of its own pitch value and the pitch value of the influencing tone, and the value of the weighting factor would depend on the amount of coarticulation as well as the direction of coarticulation.

If different tonal variations are indeed found in compatible and conflicting contexts, it would be desirable to compare the different perceptual effects of those variations. In the present study, perceptual experiments are designed to examine whether native speakers of Mandarin are able to identify the intended tones when the tones are presented with the original tonal context, without the tonal context, or with altered tonal context. It is expected that the results of these perceptual experiments not only would reveal how tones produced in connected speech are perceived, but also would shed light on the nature of tonal variation due to different tonal contexts.

I. *F*0 MEASUREMENTS OF TONES IN CONTEXT

A. 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 Fig. 2, that satisfy the description of the two categories. All of them were used in the reading list.

B. 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.

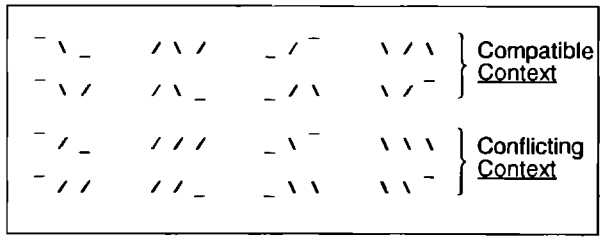

FIG. 2. Tone patterns of trisyllabic words and phrases used for *F*0 measurements.

C. Recording

For each of the 16 patterns listed in Fig. 2, two words or phrases were found. They were repeated five times and printed in Chinese in random order. 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 gaòsunǐ_shì zeňmo huíshì*.[5] "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 pulses played through a loudspeaker by a separate cassette player. The speakers were required to produce each sentence between two pulses occurring at an interval of 3 s. The recording was made with a condenser microphone and an Otari tape recorder at a tape speed of 7.5 in./s.

D. *F*0 analysis

All the utterances were digitized by the Haskins PCM system (Whalen *et al.*, 1990) at a 20-kHz sampling rate. The extraction of the *F*0 values was done manually to achieve high accuracy. A label was placed on each vocal pulse in the waveform 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 fricatives, nasals, the lateral, or unaspirated stops, *F*0 measurement started from the point of release, i.e., where a nasal or lateral segment ends and the vocalic segment starts. For syllables with initial aspirated stops or affricates, *F*0 measurement started from the onset of voice in the syllable. For all the target syllables, the *F*0 measurement ended right before the closure or constriction for the initial consonant of the following syllable.

For each token, three raw measurements were obtained for the second syllable, namely, duration, mean *F*0, and slope of the *F*0 contour. Duration measures the time in milliseconds between the first marked vocal pulse and the last. Mean *F*0 is the average of the *F*0 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 *F*0 value of the turning point in a tone contour, i.e., the point

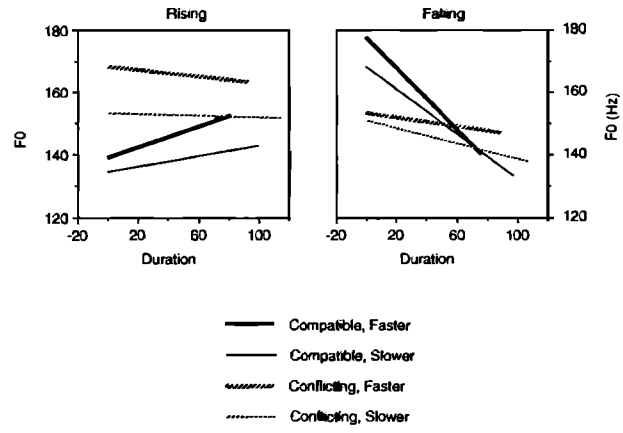

FIG. 3. Mean *F*0 slopes, mean *F*0 height and mean duration for the rising tone and falling tone in compatible and conflicting contexts. For each line, the midpoint represents mean *F*0; the slope of the line represents the mean slope; and the end of the line represents mean duration.

where a contour changes from rising to falling or vice versa, as important measurements (Gårding and Zhang, 1986; Gårding, 1987; Shen, 1990; Shen and 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 is. For this reason, turning points were not measured in this study, and the use of slope as the major measurement for tonal contours ignored any possible curvature within a contour. Also because this study is concerned mainly with comparing tonal variations under different conditions rather than examining the exact contour shapes, slope measurement should be sufficient.

E. Results and discussion

For each word or phrase spoken by each speaker at one of the speaking rates (with or without carrier), the mean slope, mean *F*0, and mean duration were obtained by averaging over the individual values of those measurements for all five tokens of the word. Figure 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 midpoint represents mean *F*0; the slope of the line represents mean slope; and the end of the line indicates mean duration. Several things are immediately apparent from this figure. (1) For the rising tone, the slope is positive in compatible context but negative in conflicting context. [In fact, 11 out of the 16 words in this condition (rising tone in conflicting context) had negative mean slopes]. For the falling tones, the slope is steeper in compatible context than in conflicting context. And, although the overall mean slope remained negative even in the conflicting context, three out of the 16 words in the conflicting context had positive mean slopes. (2) 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;

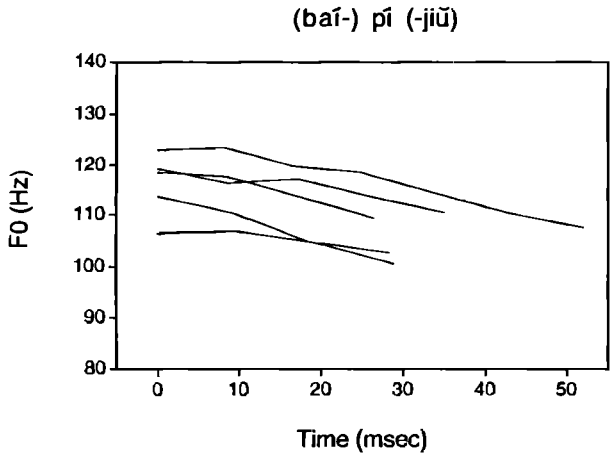

FIG. 4. *F*0 contours of five tokens of pí in the word "baípíjiǔ" (light beer) produced by speaker LC.

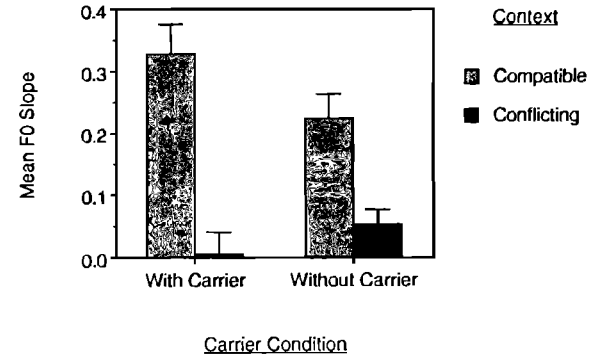

FIG. 5. Interaction between context and carrier. The horizontal bars indicate standard error.

however, in conflicting context, the slopes are steeper for the rising tone at a faster speaking rate than at a slower speaking rate, whereas for the falling tone, the slopes are steeper at a slower speaking rate than at a faster speaking rate. (3) While the rising tone has higher mean *F*0 in conflicting context than in compatible context, the falling tone has lower mean *F*0 in conflicting context than in compatible context. (4) The duration is longer at a faster speaking rate than at a slower speaking rate.

Figure 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 /pi/, not of the entire syllable.

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 "directional change," 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 directional change, 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 also conducted with mean *F*0 or mean duration as dependent variables.

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 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 *F*0 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 stepper 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 (85.58 ms) than those produced in isolation (104.95 ms), $F(1,56)=6.25$, $p=0.015$, indicating a difference in speaking rate between the two conditions.

The main effect of carrier on adjusted slope is not significant. However, there is a significant interaction between carrier and context, $F(1,56)=10.26$, $p<0.0$. Figure 5 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 with the carrier sentence (i.e., when they are short), which shows that the smaller value for the conflicting context is not simply a matter of articulatory undershoot.

There is a highly significant interaction between the effects of tone and context on the mean *F*0 of the target tones. The rising tone had a higher mean *F*0 in the conflicting context that in the compatible context, whereas the falling tone had higher mean *F*0 in the compatible context

than in the conflicting context. This does not make much sense until we take into consideration what the context is 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 *F*0 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 (cf. Fig. 3). Conversely, for the falling tone in the conflicting context, the preceding tone ends low, so the low mean *F*0 seems more like a result of pulling down its beginning value rather than raising its ending value in producing the flattened contour. 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 *F*0 in the rising tone indicates greater influence from the preceding falling tone that has a low ending *F*0 value; and the higher mean *F*0 in the falling tone indicates greater influence from the preceding rising tone that has a high ending *F*0 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, which seems to disagree with Shen's report (1990) on general symmetry in tonal coarticulation in Mandarin. However, because only the tonal influence of the first and the last syllables upon the second syllable was examined here, this finding is not yet conclusive.[6]

In summary, the results of the *F*0 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 *F*0 indicates possible greater carryover than anticipatory tonal coarticulation in Mandarin.

II. PERCEPTION EXPERIMENTS

The *F*0 analysis showed that in connected speech, a tone varied its pitch contour extensively in different tonal context, 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 *F*0 contours at their face value and perceive tones that are different from the intended ones in case the *F*0 contours deviated much from the canonical forms? If listeners are able to recover the intended tones, how do they do it? Three perceptual experiments are designed to answer these questions. Experiment 1 looks at identification of tones presented in the original contexts. Experiment 2 examines perception of tones produced in context but presented in isolation. Experiment 3 tests identification of tones presented in tonal contexts that are different from the original ones.

A. Perception experiment 1

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 nonphonetic linguistic knowledge in accomplishing the task. To achieve this goal, phonetic manipulations were performed on the trisyllabic words. For each word, certain modifications of the acoustic signal were made so that the phonetic structure of one of the syllables was changed. The phonetic manipulations were of two kinds: excision or substitution, and they were applied to either the first or the last syllable in a word. No phonetic manipulation was applied to the second syllable in a word. In the process of excision, the initial part of the frication noise in a syllable was cut out to the extent 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].

Table I is an illustration of the phonetic manipulations 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 [tɕ^h] 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. The syllable that has undergone phonetic manipulation is represented by a different character, so are the other two syllables. 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 I. Illustration of phonetic manipulation and the resulting nonsense strings used as response choices for subjects in experiment 1.

Characters	transcription	character glosses	word glosses
气象站	[tɕʰì ɕiaŋ̀ ʈʂàn]	"air situation station"	original word: "weather station"
替香战	[tʰì ɕiaŋ̀ ʈʂàn]	"replace incense fight"	nonsense strings used as response choices
替详战	[tʰì ɕiaŋ̀ ʈʂàn]	"replace detail fight"	
替想战	[tʰì ɕiaŋ̀ ʈʂàn]	"replace thought fight"	
替向战	[tʰì ɕiaŋ̀ ʈʂàn]	"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 they 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 recorded onto magnetic tape 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 I. 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 manipulation, the character used represented the new syllable. For the last (or first) syllable that had not been manipulated, 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 for a complete list of the target words and their corresponding nonsense strings used as response choices.)

The manipulation 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.

2. Subjects

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

3. 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.

4. Results

Figure 6 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.

A four factor ANOVA was conducted with context, tone, carrier, and native speaker as independent variables and identification score[7] as dependent variable.

Figure 7 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 8 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 inter-

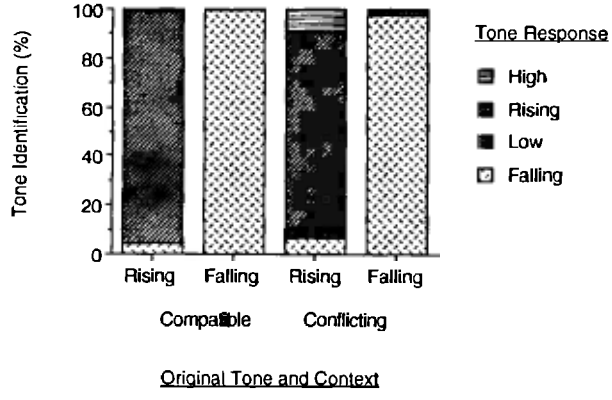

FIG. 6. Tone responses for the rising and falling tones produced in the compatible and conflicting tonal contexts in experiment 1.

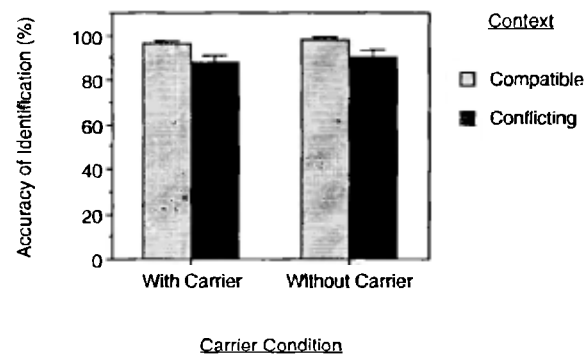

FIG. 7. Accuracy of tone identification under influence of tonal context and carrier sentence in experiment 1. The horizontal bars indicate standard error.

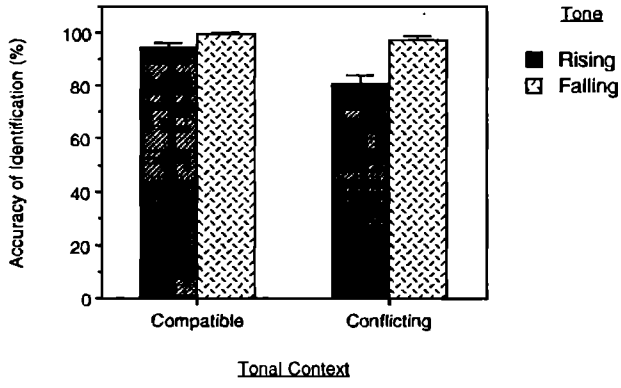

FIG. 8. Tone identification accuracy under the influence of tonal Context and Tone in experiment 1. The horizontal bars indicate standard error.

action between the effects of context and tone, $F(1,9) = 13.39$, $p < 0.01$. 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 indistinct 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, $F(1,9) = 6.33$, $p = 0.033$ (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.

5. Discussion

The results we have seen 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 *F*0 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 *F*0 contours were closer to the underlying ones due to the benefit of a compatible tonal context, the recognition was better than when *F*0 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 may not be fully compensated for by listeners.

B. Perception experiment 2

1. Stimuli

Although experiment 1 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 compensate 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.

2. Subjects

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

3. Procedure

The stimuli were recorded onto a magnetic tape in random order, each repeating 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.

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

4. Results

Subjects' response patterns are shown in Fig. 9. 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 identifi-

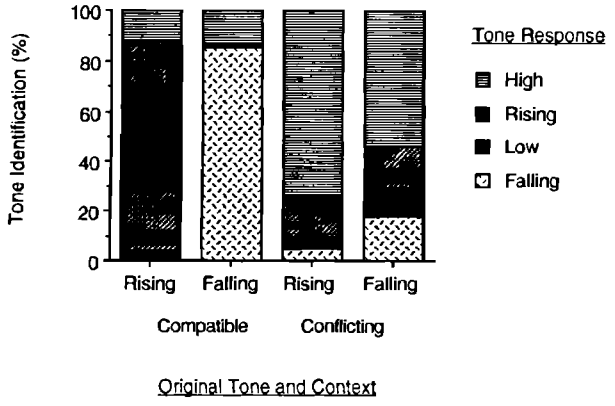

FIG. 9. Tone response patterns for syllables whose original context was replaced by white noise in experiment 2.

cation 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.

5. Discussion

The results of experiment 2 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 the target tone is coproduced with, 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 *F*0 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 *F*0 contours were heard mostly as the high tone, except when the absolute *F*0 is very low.

C. Perception experiment 3

Experiments 1 and 2 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 two experiments do not show definitely, however, what exactly in the context that is responsible for the accurate tone identification in experiment 1 but low accuracy of tone identification in experiment 2. To further determine that it is the tonal information in the context 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.

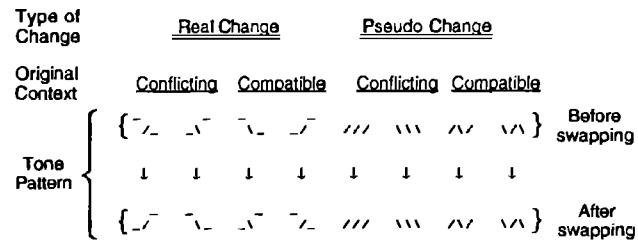

FIG. 10. Swapping design for experiment 3.

1. Stimuli

The first and last syllables in the trisyllabic words and phrases used in experiment 2 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 manipulation on the initial consonant of the first or the last syllable.

As a result of this simple swapping, two kinds of changes were introduced into the tonal context for the target tone, as shown in Fig. 10. The first kind of change, referred to as pseudochange, occurs in trisyllabic words whose first and last tones are the same, for instance, words with the tonal pattern of "/ / /." The second 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 " ¯ / _." 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 pseudochange 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 2, 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 onto magnetic tape in random order with ten 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

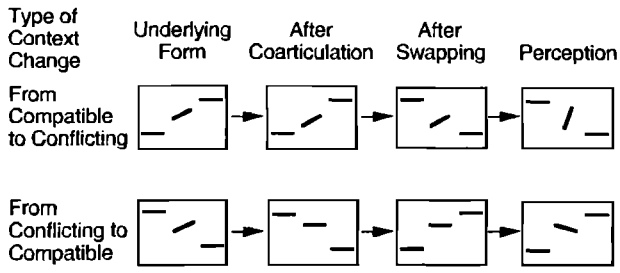

FIG. 11. Predicted effect of swapping upon perception of coarticulated tones for syllables produced in the conflicting tonal context.

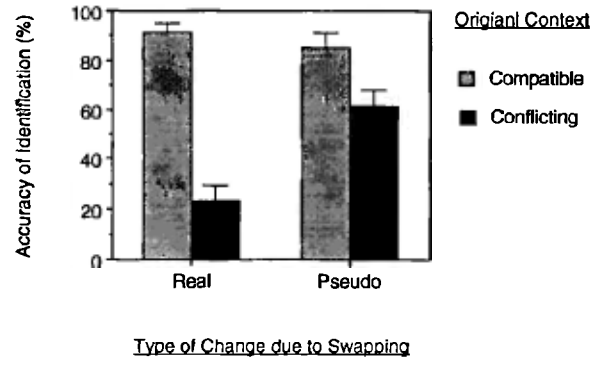

FIG. 12. Accuracy of tone identification under the influence of original tonal context and type of change due to swapping in experiment 3. The horizontal bars indicate standard error.

same *CV* structure as the original syllable. The first character and the last were printed only once for each trial. They both represent the original syllables except that their positions are exchanged. They might or might not be the same characters as the original ones, depending on how difficult it would be for the subjects to guess the original word. The essential thing was to guarantee that those characters would neither remind the subjects of the original word, nor form a new word when combined with any one of the four middle characters.

2. 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 pseudochange 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 11 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 *F*0 values of the two tones differ significantly, but also completely alters the original tonal context which is responsible for the *F*0 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.

3. Subjects

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

4. 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.

5. Results

Figure 12 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 pseudochange.

Figures 13 and 14 break up the results shown in Fig. 12 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 pseudochange as a result of swapping operation.

In Fig. 13, it can be seen that under the pseudochange condition, a rising tone was mostly heard as rising tone whether the original context was compatible or conflicting,

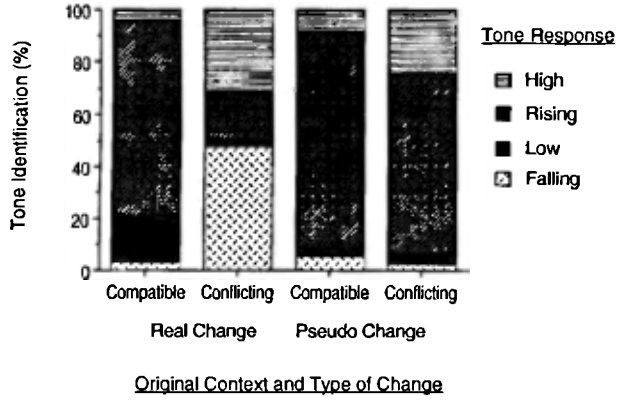

FIG. 13. Tone responses for the rising tone after swapping in experiment 3.

although as in experiment 1, the compatible context was a little better than the conflicting context. For the real change, as expected, there was dramatic difference between compatible and conflicting contexts. Most tokens originally in compatible context 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 (49.7%) 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 subjects 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 Fig. 11.

Figure 14 shows tone response patterns for the falling tone. For the pseudochange 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'

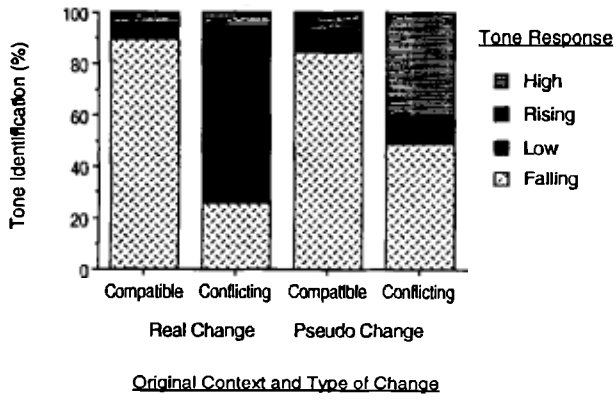

FIG. 14. Tone responses for the falling tone after swapping in experiment 3.

compensation for the reversed declination of overall *F*0 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 pseudochange by the swapping operation) reverses the *F*0 declination, which has been found in many languages including Mandarin ('t Hart, 1973; Maeda, 1976; Pierrehumbert, 1979; Gårding, 1987), 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 \ \ \ would 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 *F*0 values in the stimuli. This double raising of perceived overall pitch plus the flattened *F*0 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 *F*0 analysis described earlier. This carryover coarticulation would lower the starting as well as the overall *F*0 value of the third syllable due to the low ending value of the falling tone in the second syllable. This would then raise the pitch of the second syllable relative to that of the first syllable, and thus increasing the number of high tone responses for the second syllable after the swapping operation.

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 indistinct 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 subject would compensate for the influence of the altered context introduced by the swapping operation as if it had been there originally.

6. Discussion

The results of experiment 3 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 pitch values of their adjacent tones. The clearest demonstration of this is that in the cases where the *F*0 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 experiment 1, it must have been primarily the presence of the tonal information in the context, that helped the subjects to achieve their high accuracy of identification of the target tones.

III. GENERAL DISCUSSION

In the present study, by comparing coarticulation in "compatible" and "conflicting" contexts, it was possible to examine coarticulation under a wider range of variation than in previous studies. It was found that a phonetic unit deviated more from its underlying form when produced in a conflicting context than in a compatible context. In the cases examined in this study, 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. Specifically, when its tonal context was conflicting, a rising tone in Mandarin 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. More interestingly, the two kinds of contexts, i.e., compatible and conflicting ones, had greater effects in producing tonal variations when the syllables carrying the target tones were produced at a faster rate (with a carrier sentence) than at a slower rate (without carrier sentence). This suggests that coarticulatory tonal variation is due to overlapping of underlying tonal contours. Hence, the faster the speaking rate, the greater the temporal overlap, and the more compromise or enhancement there is between the adjacent tonal values (cf. Fig. 5).

It is also found in the *F*0 analysis that the tone of the second syllable in trisyllabic words and phrases in Mandarin is affected more by the preceding tone than the following tone, suggesting possible greater carryover than anticipatory tonal coarticulation in Mandarin. This finding differs from Shen's report (1990) that there is general symmetry in Mandarin tonal coarticulation. This difference probably comes from the difference in the reading lists as well as the instructions to the speakers used in the two studies. In Shen (1990), nonsense strings of syllables were used as reading list, whereas in the present study 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.

The finding that the rising tone sometimes becomes slightly falling in a conflicting context partially confirms the proposal by Wu (1984, 1988) that the rising tone becomes falling in the same context as one called conflicting in this study. This confirmation is only partial, however, because in this study only a tendency rather than a reliable rule was found. The tendency varies depending on the rate of speech, and hence, the amount of overlap between neighboring tones. The faster the speaking rate, the greater the tonal overlap, and the more likely there is a change in tonal direction.

There has also been a report of a change in tonal direction for the falling tone in Mandarin (Lin, 1985). However, the tonal variation Lin reported is different from what was observed in this study both in kind and in magnitude. Lin found that in certain areas of Beijing (eastern suburb and the inner city) a falling tone sometimes becomes a rising tone when followed by another falling tone. However, he also reported that when this happened, a narrow-band spectrogram showed a clear rising contour, and the tone was heard as rising tone. In looking for speakers for the present study, care was taken to avoid using speakers from those areas reported by Lin as having the falling tone sandhi. In the present study, only a very small portion of falling tone tokens had direction change. When it occurred, the slope was very close to zero, which is different from Lin's findings. Besides, unlike the perceptual shift of a falling tone to a rising tone reported by Lin, the change of direction observed in the present study did not cause a shift in perceived tonal categories when a tone was presented in its original context.

The dependence of perception of a tone on its tonal context was demonstrated by all the perceptual experiments in this study. Due to the phonetic manipulation through waveform editing in experiment 1, it was possible to examine purely phonetic perception of fluently produced tones. When presented with their original tonal context, 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 2, 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 3, 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 and 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 and Smith, 1986). In the present 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 provided correct coarticulatory information in experiment 1, the intended tones were mostly correctly identified. When the tonal context was different from the original one in experiment 3, 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.

It could be argued that some of the tonal variations in conflicting context, i.e., those involving the rising tone, are cases of phonological tone sandhi. Chao (1968, pp. 27–28) proposed a tone sandhi rule for the rising tone in trisyllabic words or phrases. The rule states that a rising tone changes into a high tone when preceded by a high or rising tone and followed by any other 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 and Li, 1967). The perceptual experiment 1 in this study showed that most of the rising tones produced with flattened pitch contours in conflicting context were still perceived as the 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 intentionally produce a tone that is different from the underlying one despite the distorted surface value due to conflicting context. Evidence is found in a recent study by Shih (1992). She finds 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 *F*0 contours from the high tone. The results of both Shih (1992) and the present study demonstrate that the contextual rising tone variation is due to coarticulation rather than a phonological process that changes the intended tone category from one to the other.

It could also be argued that the variation in the tonal contour of the second syllable is due to neutralization caused by lack of stress. While the magnitude of the variation might be related to the reduction of stress typical of syllables in the mid position of a trisyllabic sequence in Mandarin, neutralization would not fully account for the perceptual results in experiment 1. In that experiment, it was clear that despite severe distortion of the canonical tonal contours in the second syllable, in most cases, the tonal identities were not lost. Thus neutralization was certainly not what happened in the conflicting context.

Although nearly all the target tones had fairly high identifiability in experiment 1, due to the separation of those tones according to their tonal context, 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 was by 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 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 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 coproducing it with adjacent phonetic units (Fowler, 1981; Liberman and Mattingly, 1985; Whalen, 1990), 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 that particular underlying form is in a conflicting phonetic context, and the listener's seemingly imperfect recovery is actually a full recovery of a form that has been intended to be less lucid for some reason, because listeners actually perceive what talkers do (Fowler, 1987). Either of the above two being the 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 *et al.*, 1991)] cannot be explained fully in terms of bottom-up processes of recovering either the phonetic segments or the articulatory gestures.

Whatever the true cause of less perfect identification for tones produced in conflicting context than for those produced in compatible context, the fact that they differ at all also suggest a potential source for language change. As Ohala (1981, 1992) has argued, listener's failure to recover speaker's intention may be a source of language change. More specifically, "hypocorrection," i.e., failure to compensate for distortion due to phonetic context, may cause the listener to take the speech signal at its face value. In Chao's (1968) description of the rule concerning the rising tone in what is called in this study a conflicting tonal context, the tonal change occurs only at conversational speed, not at a more deliberate speed. Indeed, significant tonal variation was observed acoustically in the present study because the utterances were produced at a fluent speed. In perception, the overall high accuracy of identification ob-

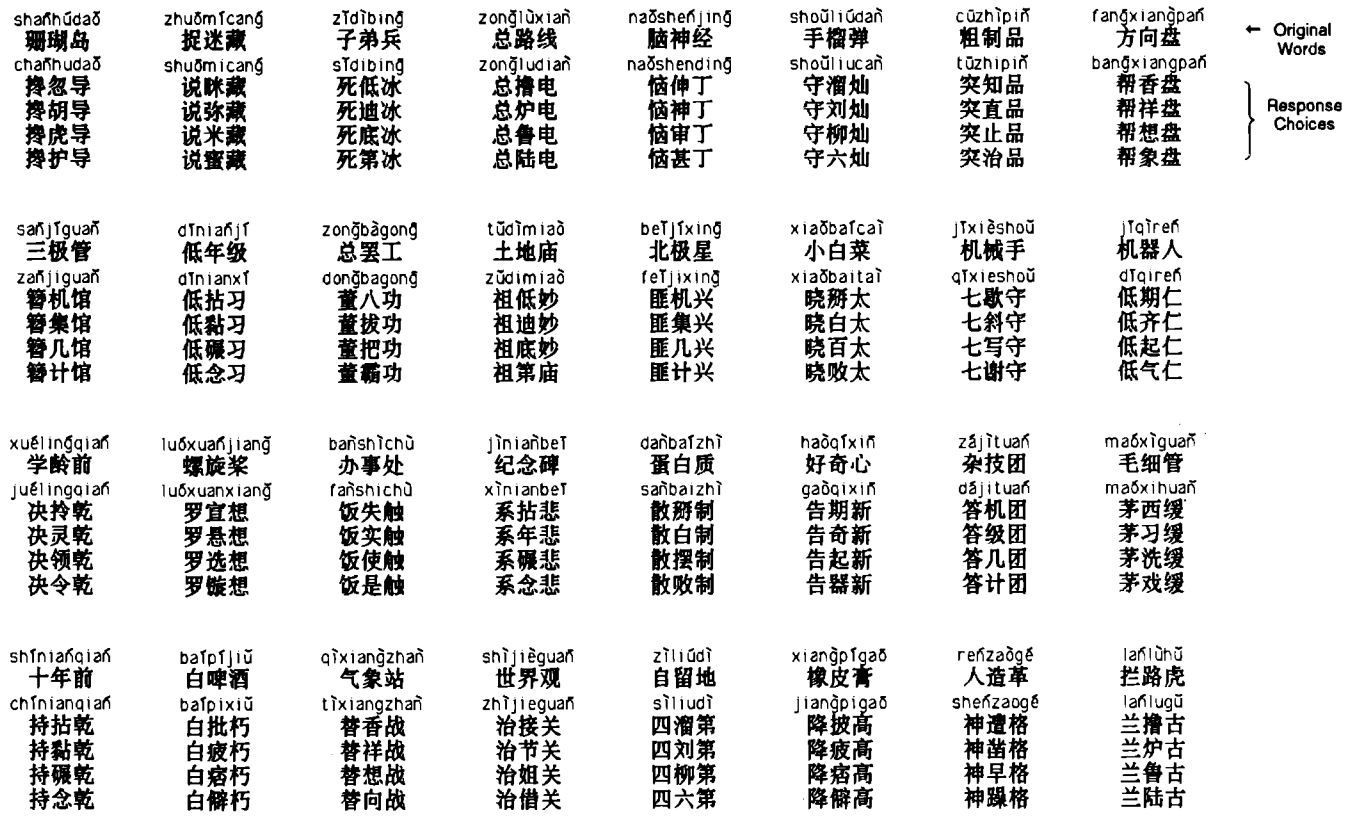

FIG. A1. Original words and response choices for experiment 1.

served in experiment 1 does not favor a phonological interpretation of the tonal variation observed in the acoustic analysis as the term "tone sandhi" used by Chao might have implied. However, the less perfect identification of the tones in conflicting context than those produced in compatible context does suggest a possible direction of this kind of variation developing into a new phonological rule due to hypocorrection.

ACKNOWLEDGMENTS

This work was supported by NIH grant NICHD-HD-01994 to Haskins Laboratories. My thanks go to Ignatius G. Mattingly, Arthur S. Abramson, D. H. Whalen, Catherine T. Best, and Carol A. Fowler, who provided helpful comments on an earlier version of this paper. I would also like to thank Xiaonan S. Shen, John Hogan, and an anonymous reviewer for reviewing my paper and providing precious suggestions. Part of this research was presented at the 122nd Meeting of the Acoustical Society of America, November 1991, in Houston.

APPENDIX

Original words and response choices for experiment 1 are given in Fig. A1.

[1]In the pinyin transcription, a phonemic spelling system for Mandarin, the four stressed tones are represented, rather mnemonically, by the tone marks "¯ " for the high tone, "ˊ " for the rising tone, "ˇ " for the falling-rising tone, and "ˋ " for the falling tone. In connected speech, the falling-rising tone is believed to be realized only as a low tone except when followed by another falling-rising tone or a pause, or at the end of an utterance. In this paper, the falling–rising tone is referred to as the "low tone," and the symbol "_" is used to represent this tone in the graphic representations of the tones. However, in the pinyin transcription, the tone mark "ˇ " is still used to represent this tone.

[2]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 more references.

[3]The term pitch here refers to the underlying pitch value rather than perceived pitch.

[4]The term "low" here is used in a relative sense. While it is generally accepted that the rising tone in Mandarin has a pitch contour of "35" in a five scale tone representation scheme (Chao, 1968), the term "low" here merely means "nonhigh" as opposed to "high."

[5]The word "nǐ" (you), although having a low tone, did not change its tone to a rising tone in cases where the first syllable in the following trisyllabic word also had low tone. This is because, on the one hand, "nǐ" and the following syllable belong to separate prosodic phrases (Shih, 1986; and Speer, 1989), and on the other hand, I believe, it is unstressed.

[6]A recent acoustic study on disyllabic tonal patterns (Xu, in press) found more solid evidence for stronger carryover than anticipatory tonal coarticulation in Mandarin, thus confirming that the asymmetry in tonal coarticulation seen here is attributable to dominance of anticipatory tonal coarticulation.

[7]The identification scores were arcsin transformed for the ANOVA to offset the ceiling effect.

Abramson, A. S. (1979). "The coarticulation of tones: An acoustic study of Thai," in *Studies in Tai and Mon-Khmer Phonetics and Phonology in Honour of Eugénie J. A. Henderson*, edited by T. L. Thongkum, P. Kullavanijaya, V. Panupong, and T. L. Tingsabadh (Chulalongkorn U. P., Bangkok), pp. 1–9.

Browman, C., and Goldstein, L. (1986). "Towards an articulatory phonology," in *Phonology Yearbook 3*, edited by C. Ewan and J. Anderson (Cambridge U. P., Cambridge), pp. 219–252.

Catford, J. C., Jusczyk, P. W., Klatt, D. H., Liberman, A. M., Remez, R. E., and Stevens, K. N. (1991). "Panel discussion: The motor theory and

alternative accounts," in *Modularity and the Motor Theory of Speech Perception*, edited by I. G. Mattingly and M. Studdert-Kennedy (Erlbaum, Hillsdale, NJ), pp. 192–194.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese* (Univ. of California, Berkeley).

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," J. Acoust. Soc. Am. **24**, 579–606.

Daniloff, R., and Hammarberg, R. (1973). "On defining coarticulation," J. Phon. **1**, 239–248.

Fowler, C. A. (1977). *Timing Control in Speech Production* (Indiana University Linguistics Club, Bloomington, IN).

Fowler, C. A. (1981). "Production and perception of coarticulation among stressed and unstressed vowels," J. Speech Hear. Res. **46**, 127–139.

Fowler, C. A. (1984). "Segmentation of coarticulated speech in perception," Percept. Psychophys. **36**, 359–368.

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

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

Gårding, E. (1987). "Speech act and tonal pattern in Standard Chinese: Constancy and variation," Phonetica **44**, 13–29.

Gårding, E., and Zhang, J. (1986). "Tone 4 and Tone 3 discrimination in Modern Standard Chinese," Lang. Speech **29**, 281–293.

Han, M. S., and Kim, K.-O. (1974). "Phonetic variation of Vietnamese tone in disyllabic utterances," J. Phon. **2**, 223–232.

Liberman, A. M., Delattre, P. C., and Cooper, F. S. (1952). "The role of stimulus variables in the perception of stop consonants," Am. J. Psychol. **65**, 497–516.

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

Lin, T. (1985). "Bejinghua qusheng liandu biandiao xin tan." (A new investigation of the falling tone sandhi in Beijing dialect.) Zhongguo Yuyan Xuebao (Bull. Chinese Linguistics) **2**, 99–104.

Lin, T., and Wang, W. S.-Y. (1984). "Shengdiao ganzhi wenti." (The issue of tone perception.) Zhongguo Yuyan Xuebao (Bull. Chinese Linguistics) **2**, 59–69.

Maeda, S. (1976). "A characterization of American English Intonation," doctoral dissertation, MIT, Cambridge, MA.

Mann, V. A. (1980). "Influence of a preceding liquid on stop consonant perception," Percept. Psychophys. **28**, 407–412.

Mann, V. A., and Repp, B. (1980). "Influence of vocalic context on perception of the [ʃ] [s] distinction," Percept. Psychophys. **28**, 213–228.

Ohala, J. J. (1981). "The listener as a source of sound change," in *Papers from the Parasession on Language and Behavior*, edited by C. S. Masek, R. A. Hendrick, and M. F. Miller (Chicago, Linguistic Society, Chicago), pp. 178–203.

Ohala, J. J. (1992). "What's cognitive, what is not, in sound change," in *Diachrony within Synchrony, Language History and Cognition*, edited by G. Kellerman and M. Morrissey (Peter Lang, New York), pp. 309–355.

Pierrehumbert, J. (1979). "The perception of fundamental frequency declination," J. Acoust. Soc. Am. **66**, 363–369.

Shen, X. S. (1990). "Tonal coarticulation in Mandarin," J. Phon. **18**, 281–295.

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

Shih, C.-L. (1986). "The Prosodic Domain of Tone Sandhi in Chinese," doctoral dissertation, University of California, San Diego, CA.

Shih, C.-L. (1987). "The phonetics of Chinese tonal system," Bell Laboratories, Tech. Memor.

Shih, C.-L., and Sproat, R. (1992). "Variations of the Mandarin rising tone," in *Proceedings of the IRCS Workshop on Prosody in Natural Speech* (Univ. of Pennsylvania, Philadelphia, PA), IRCS Rep. No. 92-37, pp. 193–200.

Speer, S. R., Shih, C., and Slowiaczek, M. L. (1989). "Prosodic structure in language understanding: evidence from tone sandhi in Mandarin," Lang. Speech **32**, 337–354.

't Hart, J., and Cohen, A. (1973). "Intonation by rule: a perceptual quest," J. Phon. **1**, 309–327.

Wang, W. S.-Y., and Li, K.-P. (1967). "Tone 3 in Pekinese," J. Speech Hear. Res. **10**, 629–636.

Whalen, D. H. (1990). "Coarticulation is largely planned." J. Phon. **18**, 3–35.

Whalen, D. H., Wiley, E. R., Rubin, P. E., and Cooper, F. S. (1990). "The Haskins laboratories' pulse code modulation (PCM) system," Behav. Res. Methods Instrum. Comput. **2**, 550–559.

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

Wright, S., and Kerswill, P. (1989). "Electropalatography in the analysis of connected speech processes," Clin. Linguistics Phon. **3**, 49–57.

Wu, Z.-J. (1984). "Putonghua sanzizu biandiao guilü." (The rules of trisyllable tone sandhi in Standard Chinese.) Zhongguo Yuyan Xuebao (Bull. Chinese Linguistics) **2**, 70–92.

Wu, Z.-J. (1988). "Tone-sandhi patterns of quadra-syllabic combination in Standard Chinese," Phonetic Laboratory Annual Report of Phonetic Research, **PL-ARPR/1988**, 1–13.

Xu, Y. (in press). "Asymmetry in contextual tonal variation in Mandarin," in *Proceedings of Sixth International Symposium on Cognitive Aspects of the Chinese Language* (Taipei, Taiwan).