CONTEXT EFFECTS IN THE PERCEPTION OF LEXICAL TONE*

Robert Allen Fox
The Ohio State University
and
Ying-Yong Qi
University of Arizona

ABSTRACT:

This study examined the effect of a short precursor context on the perception of three Tone 1-Tone 2 continua. Eighteen listeners (9 Mandarin subjects and 9 English subjects) heard speech tokens which were either presented in an isolated-token condition, or in a paired-token condition and then rated them in terms of how similar their F0 contours were to either Tone 1 (high-level) or Tone 2 (mid-rising). In the paired-token condition, each test token was preceded either by a precursor with a high-level Tone 1 contour (*ba*1) or a mid-rising Tone 2 contour (*ba*2). For both Chinese and English listeners, there was only a borderline effect of context which was assimilatory in nature, which was consistent with the contention that the contextual effect can involve auditory perceptual mechanisms.

1. INTRODUCTION.

It has often been noted in the phonetic literature that the identification of a particular sound can be significantly affected by the surrounding phonetic context. For example, as a large number of experimental studies have shown, the nature of the phonetic context can significantly affect the identification of vowel quality (e.g., Ladefoged and Broadbent, 1957; Fry, Abramson, Eimas, and Liberman, 1962; Thompson and Hollien, 1970; Ainsworth, 1975; Repp, Healy and Crowder, 1979; Assmann, Nearey, and Hogan, 1982; Crowder and Repp, 1984; Fox, 1985). Contextual effects have also been obtained with consonant stimuli (often utilizing the selective adaptation paradigm) in terms of voice onset time (Eimas and Corbit, 1973; Eimas, Cooper, and Corbit, 1973; Eimas and Miller, 1978), place of articulation (Miller and Eimas, 1977), the nasal-oral distinction (Miller and Eimas, 1977) and manner of articulation (Eimas, Tartter, Miller and Keuthen, 1978). The effect of context has been shown to extend to the temporal domain in consonant perception as well. For example, Miller (1981) has shown that the perceived "speech rate" of the surrounding phonetic context can affect how listeners identify formant transitions as representing stops (rapid transition) or glides (slower transition).

The vast majority of studies directed at contextual effects in speech perception have been directed at segmental phenomena. Would the same context effect occur when listeners are processing suprasegmental aspects of speech such as tone (which is the utilization of pitch differences in the production of linguistic distinctions)? Two separate threads of research suggest that this may be the case.

The phenomenon of tone sandhi occurs in many, if not all, of the tone languages of the world (Gandour, 1978). Tone sandhi refers to the situation in which the tonal pattern of a syllable is modified as a consequence of its juxtaposition with the tonal pattern of an adjacent syllable. For example, in Mandarin Chinese Tone 3 (low-dipping) becomes Tone 2 (mid-rising) when it occurs before Tone 3 (Wang and Li, 1967). This phonological rule clearly affects a speaker's production of the tonal pitch contrast, but it may also have a significant influence on listener's perception of the tones. In particular, a listener's tonal judgments may be biased by such phonological constraints present in their language. Even without such a context such a bias may be evident. Chuang, Hiki, Sone and Nimura (1972) required native Chinese (Mandarin) listeners to identify the tones of a set of Chinese monosyllables produced by four native speakers. The listeners generally identified the tones correctly, but the most common errors involved a confusion between Tone 2 and Tone 3. Of these errors nearly twice as many involved the misidentification of Tone 3 as Tone 2 as opposed to the reverse. Gandour (1978) remarked that the direction of perceptual confusion found in the Chuang et al. (1972) data might reflect the phonological relationship between Tone 2 and Tone 3 noted above (although Gandour expressed caution in this interpretation since the confusions obtained were not equally distributed across the four speakers).

These context effects do not seem to be limited to the phonological/ phonetic level or even to speech stimuli, however. Context effects have been found to affect the perception of both pure tones and single resonance complex tones as well. Using an anchoring task, Shigeno and Fujisaki (1979) found that the identification of stimulus tokens from both a pure tone continuum and a complex tone continuum could be significantly affected by the presence of an immediately preceding "anchor" token which represented a stimulus outside the frequency range of the stimulus continuum. These effects could be either assimilatory or contrastive in nature. When an assimilatory effect occurred, the category boundary of the stimulus continuum shifted in the direction opposite to that of the anchor. When a contrast effect occurs the boundary shifts in the direction of the anchor stimulus. Shigeno (1986), obtained similar results using an AXB discrimination task in which he found that the perceived pitch of the X token shifted as a function of temporal proximity to either token A or B. Context effects may thus be a product of general auditory processes as well as more specialized phonetic processes.

Except for research directed at tone sandhi, relatively few studies have explicitly sought to investigate the effect which context might have on the perception of fundamental frequency (F0) in terms of lexical tone distinctions. We thus know little about the influence of context upon auditory and/or phonetic level processing utilized in the identification of tonal distinctions, although two studies (Lin and Wang, 1985; and Leather, 1983) have provided some data supporting the existence of such influence. Lin and Wang (1985) demonstrated that the initial frequency of a Tone 4 (high-falling) syllable could significantly change Chinese (Mandarin) listeners' identification of an immediately preceding Tone 1 (high level) syllable.

Lin and Wang presented listeners with the syllable *bai*1 followed by one of four versions of the syllable /de^4/. The F0 contour for the Tone 1 syllable, which was 210 ms in duration, began at 115 Hz and ended at 115 Hz. In each of the Tone 4 syllables, which were 140 ms in duration, F0 dropped 40 Hz from start to end, but the starting frequency of the fundamental frequency differed across the four versions of the syllable. Initial F0 was either 110, 120, 130 or 140 Hz. Lin and Wang found that as the starting frequency of the second syllable increased, the more likely were listeners to identify the first syllable as being Tone 2 (mid-rising). A similar result was obtained when the initial syllable was *da*1 instead of *bai*1. These data suggest that a listener's identification of a tone is not absolute, but depends to some degree on its relationship to the wider F0 context. In particular, it was difficult for listeners to identify a tone as high-level when its F0 is clearly lower than the frequency of a following tone which supposedly begins at a high tonal level. The actual cause of the response shift cannot be adequately determined since the methodology used by Lin and Wang did not allow one to ascertain whether the effect was an automatic function of auditory/phonetic processing or merely a consequence of response bias whenever a clearly discernible F0 mismatch between the two syllables was present.

In a second study directed at this general issue Leather (1983) sought to examine the effect of speaker context upon the perception of lexical tone in Mandarin speakers using an experimental technique similar to that used by Ladefoged and Broadbent (1957). He created two sets of synthetic Mandarin tone stimuli based on the vowel [y] each of which varied from Tone 1 to Tone 2. The particular F0 frequencies used in the synthesis were based upon two different speakers who demonstrated two different F0 ranges (one low, one high).

Each step of the two separate tone continua were embedded in one of four natural speech phrases produced by each of the two speakers. The mean F0 for the low-F0 speaker was 141 Hz across these four phrases while the mean F0 for the high-F0 speaker was 170 Hz. One test tape was created for each separate speaker. Each tape contained 10 different occurrences in three random orders of each resulting sentence type (phrase + target syllable)1. In a subsequent listening task listeners were presented with two different blocks of sentences, each block containing only the precursor phrases from one individual speaker. Listeners were required to assign the target syllable to either the Tone 1 or Tone 2 category.

The pattern of the tone categorizations in the Leather (1983) study (described only in terms of a chi-squared contingency table) demonstrated that three of the five listeners categorized target syllables with identical F0 characteristics differently as a function of the precursor phrase. Leather concluded that his experiment provided support for the hypothesis that the inferred F0 range of the speaker (as determined from the precursor phrase) was a determining factor in the categorization of lexical tone. A similar argument could be made for the Lin and Wang (1985) data.

Although Leather's (1983) study does provide some indication that some type of "speaker/pitch normalization" (one possible type of contextual effect) is operating in the perception and identification of lexical tones, there are several unanswered questions. First, Leather (1983) stated only that the pattern of tone categorizations were different from one context to another. He did not describe the nature of the difference, that is, whether the effect was one of contrast (e.g., underestimation of F0 of the target syllable when preceded by a high F0 phrase) or assimilation (e.g., over-estimation of F0 when preceded by a high F0 phrase). In fact, although Leather computed the chi-squared difference values separately for each listener for each of four stimulus pairs, he did not include information to indicate (1) if the contextual effect was in the same direction (assimilatory or contrastive) for each listener, (2) if the contextual effect was in the same direction for a given listener across all four stimulus pairs, or (3) whether an overall context effect would be obtained if the data were collapsed across listeners.

It is important to determine the nature (direction) of the effect because different stimuli may produce different types of effects, which may reflect the influence of context upon different levels of perceptual processing. Shigeno and Fujisaki (1979), using an anchoring task, found assimilation effects for pitch judgments of pure tones at short ISIs but contrast effects for phonemic judgments of vowels at similar ISIs (the majority of context effects involving phonemic distinctions have found contrast effects).

Second (and related to the first point), it is unclear whether the changes in listener's responses in Leather's study were auditory or phonetic in nature. That is, did the context affect the auditory mechanisms responsible for the processing of pitch (including non-speech stimuli such as pure tones) or did they affect higher-level phonetic processes, or both? Leather (1983) discussed the data only in terms of "speaker-adaptation" and "inferred" speaker F0 range but did not describe any specific normalization algorithm that might have been utilized by the listeners and did not attempt to separate auditory from phonetic effects.

On the other hand, Shigeno and Fujisaki (1979, 1980; see also Shigeno, 1986)--following the perceptual model developed by Fujisaki and Kawashima (1971)--hypothesized that assimilatory shifts were produced by affecting information contained in a precategorical acoustic memory store. Contrastive shifts, they speculated, were produced by influencing a categorical memory (which is often labeled phonetic memory). The shifts found in Leather's study could have been produced by either effect, particularly since the tonal distinction also represented a linguistic (categorical) distinction for all listeners.

There is, in addition, at least one other possible source of the obtained identification shifts in Leather's data, that of response bias. This explanation would claim that the contextual effects were a result of changes in subjects' labeling strategies rather than a modification of internal perceptual images in the perceptual process. Such effects might have been obtained given that listeners heard the stimuli in two separate blocks with only a single speaker's voice representing the precursor phrase in each block.

Listeners could have developed an task-dependent strategy in such a situation--a strategy that would not represent the processes used in "normal" speech perception. This possibility is entertained by Leather who stated that "relatively long attention to a single speaker while concentrating on a single phonetic contrast could possibly affect subjects' phonetic criteria in some way uncharacteristic of normal listening" (1983, p. 379). It is also interesting to note that the two subjects who heard the sets of stimuli in the opposite order from the other three subjects demonstrated identification shifts that were "not so significant as for the other subjects" (1983, p. 379).

A final question that needs to be addressed involves the extent to which the two different acoustic cues to the tonal distinction (F0 rise-rate and overall F0 height) may be separated in terms of the context effect. In particular, the majority of the significant contrast effects (all but one) obtained by Leather (1983) were confined to a single pair of test tokens (stimuli 4 and 6). Each of these tokens had a F0 rise rate of approximately 7 semitones/second which Leather had previously determined was the location of the Tone 1/Tone 2 toneme boundary. Leather thus obtained context effects (using two contexts which differed primarily in terms of overall F0 height) only when the rise-rate cue had been suppressed. Listeners, presumably, could be making their tonal categorizations of these ambiguous stimuli strictly on the basis of their initial or final frequency compared to the overall F0 of the precursor context. These data support the contention that the perception of pitch can be affected by context, but suggests nothing about whether similar effects can be obtained in the perception of F0 rise rate.

The purpose of the present study is to examine the effect which a preceding context has upon the tonal categorization of a target stimulus using the paired-token paradigm (Crowder and Repp, 1984). Complementary to Leather (1983), this study will concentrate upon the F0 rise-rate cue and will suppress the overall F0 cue. The experimental stimuli will include precursors which represent either good Tone 1 or Tone 2 exemplars, but which do not differ from the test tokens in terms of overall F0. This should allow us to separate the F0 rise-rate effect from the effect(s) involving overall inferences about the speaker's mean F0 or range.

The listeners in the experiment will include both native speakers of Mandarin Chinese (for whom the tonal distinction will be phonemic) as well as monolingual American English listeners--in this way we hope to separate possible auditory processing from phonemic processing. The stimuli will not be blocked in terms of precursor phrase, rather, the nature of the precursor will be randomly varied across the stimulus set. This should eliminate any task-dependent processing strategies that might arise in the processing of a single repetitive precursor phrase over a long sequence of test tokens.

2. METHOD.

2.1. Listeners.

A total of 18 listeners, naive to the purpose of the study, participated in the experiment. Nine of these listeners were native speakers of Mandarin Chinese from Beijing with no known speech or hearing impairment. Seven of the Mandarin Chinese speakers were graduate students at The Ohio State University and two were student spouses. The remaining nine listeners were undergraduate or graduate students at The Ohio State University. All were monolingual and spoke a Midwestern dialect of American English (central Ohio) with no known speech or hearing impairment.

2.2. Stimuli.

Mandarin Chinese has been recognized as having four distinctive tones for the citation syllable (Chuang, Hiki, Sone and Nimura, 1971; Howie, 1976; see Gandour, 1978). Of these four tones, the high-level tone (Tone 1) and the mid-rising tone (Tone 2) were selected as the endpoints of three Tone 1-Tone 2 stimulus continua--these two tones are the same ones used in the Leather (1983) experiment.

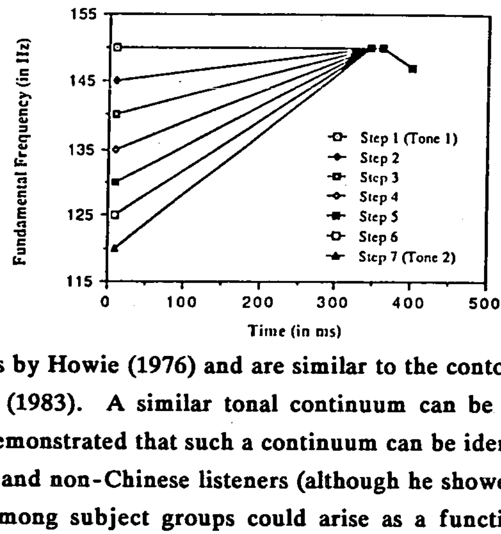
The three tonal continua were based on the syllables *ba*, *da* or *bi*. These continua were created using a cascade/parallel software synthesizer (Klatt, 1980) implemented on a PDP 11/23 computer. Each stimulus was 400 msec in duration. Each stimulus had a steady-state vowel approximately 350 msec in duration. The steady-state frequencies of F1, F2, and F3 for [a] was 700, 1220, and 2600 Hz, respectively. The steady-state frequencies of F1, F2, and F3 for [i] were 300, 2020, and 2960 Hz, respectively. The frequencies of formant 4 and 5 were kept at 3300 and 3850 Hz, respectively, throughout the token.

The *ba* stimulus had a 5 msec stop release burst created using a broad-band noise source with the frequencies of F1, F2, and F3 set at 360, 1126 and 2267 Hz, respectively. The CV formant transitions for the *ba* stimulus were 30 ms in duration. During the transition, the frequencies of F1, F2, and F3 changed from 440, 1139, and 2360 Hz, respectively, to the formant values for the steady state vowel [a]. All transitions were based on linear interpolation.

The *bi* stimulus also had a similar 5 ms stop release burst with the frequencies of F1, F2, and F3 set at 263, 1235, and 2296 Hz, respectively. The CV formant transition was 30 msec in duration. During the transition, the frequencies of F1, F2, and F3 started at 294, 1504, and 2369 Hz, respectively, and changed to the steady-state formant values for the [i] vowel in a linear manner. The *da* stimulus had a 10 ms stop release burst created using a broad-band noise source with the frequencies of F1, F2 and F3 set at 322, 1520, and 2600 Hz, respectively. The CV formant transition was 50 ms in duration. During the transition, the frequencies of F1, F2, and F3 went from 382, 1520, and 2600 Hz, respectively, to the steady-state formant frequencies of the steady-state vowel [a]. Again, all transitions were based on linear interpolation of the formant frequencies.

Since Tone 1 and Tone 2 have approximately the same F0 level at their endpoints and the frequency of F0 rises in a linear fashion for the mid-rising tone, it was possible to produce Tone 1-Tone 2 continua by varying the starting F0 in discrete steps and linearly interpolating the F0 over the duration of the syllable. There were seven steps in each tone continuum. The frequency of the Tone 1 endpoints began at 150 Hz where it remained steady for 360 ms; it then fell in a linear manner to 147 Hz at the end of the token. The F0 of the Tone 2 endpoints began at 120 Hz and rose linearly to 150 Hz after 360 ms. The F0 of all tokens fell from 150 Hz at the 360 ms point to 147 Hz at the end of the token to increase the perception of naturalness. The starting F0 values for the other 5 continua steps were 5 Hz apart and are shown schematically in Figure 1. These contours were based on

Figure 1. Schematic representation of F0 changes along each 7-step continuum.



measurements by Howie (1976) and are similar to the contours used in Fox and Unkefer (1983). A similar tonal continuum can be found in Wang (1976) who demonstrated that such a continuum can be identified easily by both Chinese and non-Chinese listeners (although he showed that response differences among subject groups could arise as a function of language background and experience with psychophysical experiments).

Given the three 7-step continua, there were a total of 21 different stimulus tokens. From these stimuli, two experimental tapes were created. The first tape was an isolated-token identification tape which contained 20 examples of each of the stimulus tokens in randomized order for a total of 420 tokens with an interstimulus interval of 3 s. These tokens were output at 10 kHz (with an accuracy of 12 bits), low-pass filtered at 4.5 kHz and recorded using a high-quality cassette recorder (Harmon/Kardon 191). The interstimulus interval was 3000 ms.

Prior to the test tokens, a set of 21 example tokens and 10 practice tokens were included on the identification tape. The example tokens included each step of each tone continuum. The steps from each tone continuum were played in order from the Tone 1 endpoint to the Tone 2 endpoint. The instruction sheet described each of the tone continua and listeners read the appropriate portion of the instruction sheet prior to hearing the example tokens. Following the example tokens, there was a set of 10 practice tokens. The isolated-token condition represents the "neutral" context, that is, the immediately preceding token (context) for any target token was a random distribution of tones and segments. This is in contrast to the paired-token condition2 in which each target token was preceded by only one of two different tokens.

The second tape created the paired-token stimuli. In this tape, pairs of stimulus tokens were presented with a 500 ms interstimulus interval. The second token in each pair was the target token representing one of the 21 tone stimuli. The first token in a pair was either the Tone 1 endpoint from the *ba* continuum (ba1) or the Tone 2 endpoint from the *ba* continuum (*ba*2). As will be noted in the procedure section, listeners made no response to the first token. Rather, the first token served as the context for the second token. Each of the 21 different target tokens were paired with either *ba*1 or *ba*2 20 times for a total of 840 stimulus pairs. These pairs were randomly ordered with an interval of 3 s between stimulus pairs. There was short rest period of 2 minutes midway through the experiment. The paired--token tape also included both a set of example pairs (15) followed by a set of practice pairs (10).

All listening tests were conducted in a sound-conditioned booth (IAC). The tapes were played on a high-quality cassette recorder (Harmon/Kardon 191) through high-quality headphones (Sennheiser HD 420) at a comfortable listening level.

2.3. Procedure.

Listeners completed the isolated-token condition on one day and the paired-token condition on a second day. The identification test took approximately 25 m to complete, and the paired-contrast test took approximately 1 h to complete. Listener read a separate instruction sheet for each different condition. Each instruction sheet described the set of examples which occurred first on the tape.

The instructions for the isolated-token test described the tonal variations present in each of the three stimulus continua, and listeners heard all 7 steps from each of the three continua, in order. There were separate instruction sheets for each language group. The instruction sheet for the English listeners described Tone 1 as a high, unchanging (i.e., level) pitch, and Tone 2 as a mid-rising pitch. The Chinese instructions referred to the tones using the appropriate linguistic terminology. Prior to hearing the example set, the listeners were told that they would hear three groups of tokens consisting of *ba*, *da*, and *bi* and that within each of these groups of tokens there were seven different pitch variations, ranging from a clear Tone 1 to a clear Tone 2. The other five tokens were described as somewhat "ambiguous," representing intermediate pitch variations between Tone 1 and Tone 2.

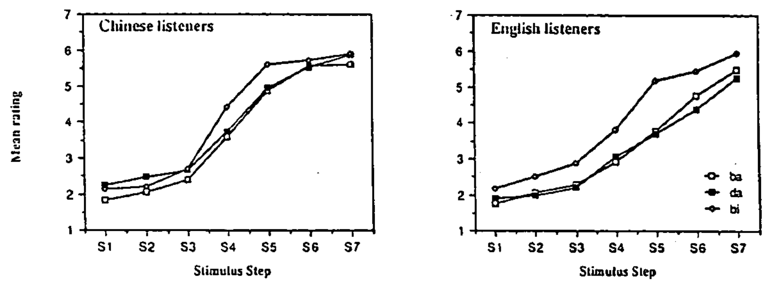
Listeners were required to identify a given the stimulus token by rating how similar it was to either Tone 1 or Tone 2 on a 7-point scale. On this 7-point scale, "1" indicated a pitch contour identical to Tone 1 while "7" was identical to Tone 2. Listeners completed 15 practice items before the actual test items began.

In the paired-token condition each listener was instructed (in either English or Chinese) that they would be hearing pairs of stimulus tokens. The first token of each pair would either be *ba* with a high-level pitch or *ba* with a mid-rising pitch. In all cases they were told to ignore the first token of each pair. They were required to identify the second token of each pair using the 7-point scale described above. Each listener was informed that the tokens that would appear in the second position were the same set of tokens heard in the isolated-token condition. Again, sets of stimulus pair examples appeared first on the tape followed by a set of practice items.

3. RESULTS AND DISCUSSION

3.1 Isolated-token data.

Figure 2. Mean rating responses for each of the two language groups for the isolated-token condition.



Shown in Figure 2 are the results from the isolated-token condition for each of the three stimulus continua, in terms of mean rating response. The left panel shows the data from the Chinese listeners; the right panel displays the data from the English listeners. Several things are apparent from a visual inspection of the data. First, the rating functions of the Chinese listeners show a somewhat sharper slope near the the toneme boundary than do the English listeners, similar to those found for "categorical" identification functions (Liberman et al. 1967). This is to be expected since the Tone 1-Tone 2 distinction is a linguistic contrast in Mandarin but not English. Second, for both listener groups the rating function for the *bi* continuum is consistently different from the *ba* and *da* continua. The *bi* tokens (especially the more ambiguous tokens, steps 3-5) are more likely to be rated as more Tone 2 like than their *ba* and *da* counterparts. This is probably due to the difference in intrinsic pitch for the [i] as opposed to [a]. In particular, the vowel [i] tends to have a higher fundamental frequency than does the vowel [a] (Fairbanks, 1953; Lehiste, 1970). Listeners might then expect this F0 difference and perceive a particular F0 as relatively lower in pitch when it occurs with [i] than when it occurs with [a]. This might cause listeners to perceive a *bi* token at a given F0 as more similar to Tone 2 (which has a lower staring frequency than Tone 1) than a *ba* or *da* token at the same F0. Chuang and Wang (1978) found a similar pitch bias based on vowel quality differences using a paired-vowel pitch discrimination task.

These mean ratings were analyzed with a three-way analysis of variance with the factors language (Chinese and English listeners), stimulus continuum (*ba*, *da* and *bi*), and continuum step (steps 1-7). Language was a between-subject factor while stimulus continuum and continuum step were within-subject factors. There was, of course, a significant main effect of continuum step ($F(6,336)=201.7$, $p < .0001$) demonstrating that position along the Tone 1/Tone 2 continua had a reliable effect upon the tonal ratings.

The main effect of language was also significant ($F(1,336)=23.4$, $p<.0001$). This effect was obtained because the mean rating of the English listeners (3.50) was lower than the Chinese listeners (3.91). There was also a significant language X continuum step interaction ($F(6,336)=4.02$, $p<.001$). This interaction effect demonstrates that the slope of the rating functions were different for the two language groups.

A significant main effect of stimulus continuum was also obtained ($F(2,336)=16.03$, $p<.001$), reflecting that fact that the mean rating of the *bi* continuum (4.04) was significantly higher than for the *ba* (3.51) or *da* (3.57) continua. This effect interacted with language ($F(2,336)=4.58$, $p<.001$) demonstrating that the difference between the *ba*, *da*, and *bi* continua were not as great for the Chinese listeners (mean ratings, 3.72, 3.92, and 4.08, respectively) as for the English listeners (mean ratings, 3.30, 3.22, and 3.99, respectively).

The isolated-token results show that all three continua produce reasonably comparable ratings functions for both the Chinese and English listeners. This condition has also served as training session for the paired-token condition.

3.2 Paired-token data.

The results from the paired-token condition are shown in Figure 3 for the Chinese listeners and Figure 4 for the English listeners. It is evident from these figures that any contextual effect obtained (whether it be assimilatory or contrastive in nature) is relatively small.

Figure 3. Mean rating responses for the Chinese listeners for the pairedtoken condition.

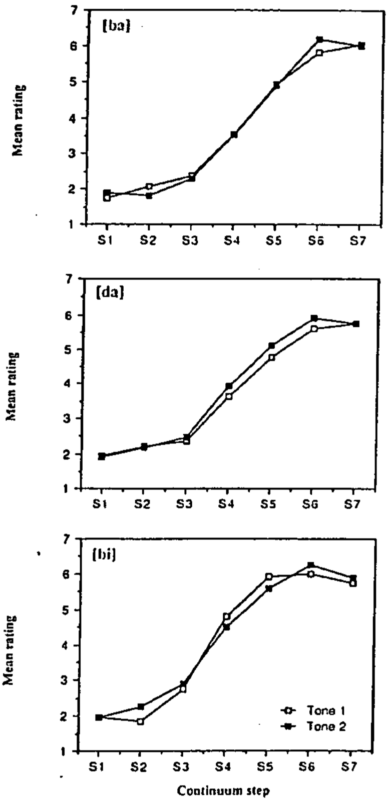
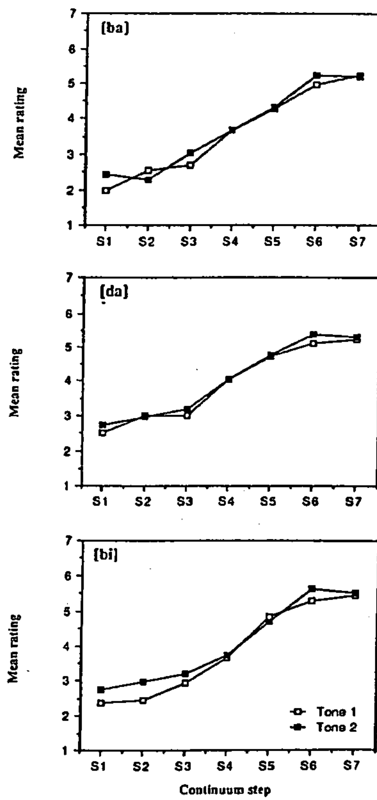


Figure 4. Mean rating responses for the English listeners for the paired-token condition.



These data were analyzed using a four-way analysis of variance with the factors language (Chinese and English listeners), precursor (*ba*1 or *ba*2), stimulus continuum (*ba*, *da*, and *bi*), and continuum step (steps 1-7). Language was a between-subject factors while the others were withinsubject factors.

Unlike the isolated-token condition, there was no significant main effect of language (F(1,672)=0.65, p>.42). The overall mean rating of the Chinese listeners (3.93) was very similar to that of the English listeners (3.88). The main effect of precursor was only of borderline significance (F(1,672)=3.48, p<.06). The mean rating following *ba*1 precursor was 3.85, while the mean rating following *ba*2 precursor was 3.96. The same trend was obtained in both the Chinese (mean rating following *ba*1 = 3.89; following *ba*2 = 3.96) and the English (mean rating following *ba*1 = 3.81; following *ba*2 = 3.95) data. These data suggest a slight contextual effect of assimilation, similar to the effects found by Shigeno and Fujisaki (1979) in the identification of the pitch of pure tones and the timbre of single resonance tones at a 500 ms ISI.

Both the main effects of stimulus continuum (F(2,672)=11.28, p<.0001) and continuum step (F(6,672)=395.14, p<.0001) were significant. As in the isolated-token condition, listeners tended to rate the *bi* tokens higher on the Tone 1/Tone 2 scale (mean rating 4.07) than either the *ba* tokens (mean rating 3.74) or the *da* tokens (mean rating 3.91).

There were only two significant interaction effects: the language X stimulus continuum interaction (F(2,672)=4.31, p<.02) and the language X continuum step interaction (F(6,672)=15.40, p<.0001). The language X stimulus continuum interaction was obtained because the two language groups demonstrated a slightly different pattern in their ratings of the *ba*, *da*, and *bi* continua. In particular, for the English listeners the mean rating for the *da* continuum (3.82) was closer to that for the *ba* continuum (3.79) than for the *bi* continuum (4.17). In the English data, the mean rating for the *da* continuum (4.00) was closer to that for the *bi* continuum (3.96) than for the *ba* continuum (3.68). As in the isolated token condition, the significant language X continuum step interaction was obtained because the slopes of the rating functions for the English listeners were less steep than were those for the Chinese listeners.

Table 1. Values of chi-squared values computed for each listener for continuum steps 3, 4 and 5. Values were computed on the basis of 2 X 2 contingency tables of tone category (Tone 1 or Tone 2) and precursor (*ba*1 or *ba*2). For those values that are significant, we have indicated whether the obtained effect was assimilatory (Assim.) or contrastive (Contr.).

	Continuum Step		
Listener No	Step 3	Step 4	Step 5
	Chinese Listeners		
1	0.00	0.47	0.30
2	4.18* (Assim.)	3.33	0.20
3	0.30	0.00	0.09
4	0.22	1.83	0.08
5	0.11	0.00	0.56
6	0.92	0.14	5.96* (Assim.)
7	0.00	1.39	0.00
8	6.86** (Assim.)	5.96* (Assim.)	6.26* (Assim.)
9	14.8*** (Assim.)	2.28	0.13
	English Listeners		
10	0.00	1.64	0.14
11	4.54* (Assim.)	0.44	0.61
12	1.23	1.73	0.51
13	0.19	0.84	0.38
14	0.00	4.81* (Contr.)	1.64
15	38.5*** (Assim.)	21.9*** (Assim.)	23.2*** (Assim.)
16	1.57	1.64	1.50
17	0.00	0.70	0.00
18	0.27	0.19	0.03

* $p<.05$ ** $p<.01$ *** $p<.001$

At first glance, the data obtained in Leather (1983) seem to provide much stronger support for context effects on tonal categorization than do the present data. However, as pointed out in the introduction, only the chi-squared values were published in the study which somewhat obscured the nature of the contextual effects obtained. In order to better compare the data from his study with those presented here, we have reduced our data to binary categorizations and analyzed them using chi-squared statistics in a manner similar to that used by Leather.

In converting the rating responses to binary responses ratings of 1-3 were were considered a Tone 1 response and ratings of 4-7 were considered a Tone 2 response (this conversion scheme is somewhat arbitrary but is based on the fact that the mean overall rating of the both groups of listeners was between 3.00 and 4.00). For each of the three middle steps of the continua (steps 3, 4, and 5) and each listener, a chi-squared value was computed from a 2 X 2 contingency table of tone category (Tone 1 or Tone 2) and precursor (*ba*1 and *ba*2). Since there was no significant precursor X stimulus continuum interaction, the response from each step were collapsed across the three stimulus continua.

The obtained chi-squared values are shown in Table 1. For those values which are statistically significant we have indicated whether the observed effect was assimilatory (i.e., more Tone 1 responses were obtained when the test token followed the *ba*1 precursor than the *ba*2 precursor) or contrastive (i.e., fewer Tone 1 responses were obtained when the test token followed the *ba*1 precursor than the *ba*2 precursor). It is evident that some degree of listener variability is evident, as was found in Leather (1983). For both language groups there is at least one listener who demonstrates a significant context effect at all three steps and several listeners who demonstrate no significant effect at any step. The majority of the significant effects obtained were assimilatory (82%) rather than contrastive (18%) which is consistent with our previous analysis of the data. The percentage of significant to nonsignificant responses in our study is less than, but roughly comparable to that obtained by Leather. In particularly, Leather found that 7 of the 20 listener by stimulus combinations (35%) showed significant differences as a function of context. In the present study, 6 of the 27 combinations (22%) showed significant differences. Given this comparison, it is possible that the contextual effects observed by Leather (1983) would not have attained statistical significance if the data were analyzed in a more global fashion without eliminating the direction of the contextual effect from the analysis.

4. GENERAL DISCUSSION.

The experiment reported here provides only limited support for the hypothesis that context (in the form of F0 rise rate) can significantly affect lexical tone identification responses for either Chinese or English listeners. The small context effect that was obtained with assimilatory in nature and was similar for both groups of listeners. This would argue that the contextual effect was a product of auditory rather than phonetic processing (Shigeno and Fujisaki, 1979). The fact that there was no precursor X stimulus continuum effect (i.e., the same assimilatory effect occurred for three stimulus continua) supports the contention that the effect was strictly suprasegmental in nature and was unaffected by segmental differences between the precursor and the target syllable.

Although F0 rise rate may produce limited contextual effects, the Leather (1983) study suggested that the overall F0 height of the context may have significant effect upon tonal categorization. However, as we have pointed out, the method of analysis used by Leather ignored the direction of the contextual effects obtained and do not necessarily demonstrate a significant and consistent effect of precursor F0 rise-rate on tonal categorizations. Thus the nature of contextual effects (even in the form of overall F0 height) upon the identification of tonal contours is very unclear at this point.

Clearly more controlled experiments examining the influence of context (especially in terms of pitch variations) on the perception of suprasegmental phenomena, in general, and lexical tone, in particular need to be done. Especially needed are experiments including both F0 rise-rate and overall F0 height differences in both the context and the target tokens. The present study used a single temporal interval (500 ms) between the precursor token and the test token (Leather did not provide specific details about the exact ISI used in his study) and did not systematically examine sequential order effects between individual stimulus tokens. However, both these factors may play a significant role in the identification of tone. In particular, the data from Shigeno and Fujisaki (1979, 1980) and Shigeno (1986) demonstrate that the direction of the context effect may change as a function of the time between the presentation of the context and the onset of the test token. In addition, Chuang and Wang (1978) demonstrated that listeners may show a pitch bias related to the sequential order of stimulus presentation. They found that when listeners were required to compare the pitch of pairs of vowels, there was a 12% likelihood that listeners would judge the pitch of the second vowel as higher than the first. Clearly, such temporal factors should be systematically controlled in future experiments.

NOTES

*The authors would like to thank Prof. William Wang for his suggestions on an earlier version of this study and for directing our attention to several important reference articles. Copies of this articles may be obtained from the first author, at Division of Speech and Hearing Science, The Ohio State University, Columbus, OH 43210-1372

[1]The description of the stimulus set by Leather (1983) is somewhat ambiguous. In particular, he states that "four natural speech carrier sentences...(meaning 'What word is this: X?')" were recorded by each of the two speakers. He does not specify in what ways these four carrier sentences differed. It is also unclear whether each target syllable was embedded in each of the four carrier sentences or in only one particular version. He does not state the total number of tokens identified during the listening test.

[2]In the relevant literature, the "paired-token" condition is often referred to as "paired-contrast." However, since we do not know, a priori, whether any observed contextual effect will be assimilatory or contrastive, we will use the more neutral term.

REFERENCES

AINSWORTH, W. 1975. Intrinsic and extrinsic factors in vowel judgments. In G. Fant and M. Tatham, eds., Auditory Analysis and Perception of Speech. London: Academic Press. Pp. 103-113.

ASSMANN, P., T. NEAREY, and J. HOGAN. 1982. Vowel identification: Orthographic, perceptual and acoustic aspects. Journal of the Acoustical Society of America 71. 975-989.

CHUANG, C., S. HIKI, T. SONE, and T. NIMURA. 1971. The acoustical features and perceptual cues for the four tones of standard Chinese. Proceedings of the Seventh International Congress on Acoustics, III. Budapest: Akademiai Kiado. Pp. 297-300.

CHUANG, C. and WANG, W. S-Y. 1978. Psychophysical pitch biases related to vowel quality, intensity difference, and sequential order. Journal of the Acoustical Society of America 64. 1004-1014.

CROWDER, R.G. and B.H. REPP. 1984. Single formant contrast in vowel identification. Perception and Psychophysics 35. 372-378.

EIMAS P.D. and J. D. CORBIT. 1973. Selective adaptation of linguistic feature detectors. Perception and Psychophysics 13.247-252.

EIMAS, P.D., W.E. COOPER, and J.D. CORBIT. 1973. Some properties of linguistic feature detectors. Cognitive Psychology 4.99-109.

EIMAS, P.D. and J.L. MILLER. 1978. Effects of selective adaptation on the perception of speech and visual patterns. In H. L. Pick and R. D. Walk, eds., Perception and Experience. New York: Plenum.

EIMAS, P.D., V.C. TARTTER, J.L. MILLER, and N.J. KEUTHEN. 1978. Asymmetric dependencies in processing phonetic features. Perception and Psychophysics 23. 12-20.

FOX, R. A. 1985. Auditory contrast and speaker quality variation in vowel perception. Journal of the Acoustical Society of America 77. 1552-1559.

FOX, R.A. and J. UNKEFER. 1985. The effect of lexical status on the perception of tone. Journal of Chinese Linguistics 13.69-90.

FRY, D., A. ABRAMSON, P.D. EIMAS, and A.M. LIBERMAN. 1962. The identification and discrimination of synthetic tones. Language and Speech 5. 171-189.

FUJISAKI, H. and T. KAWASHIMA, T. 1971. A model of the mechanisms for speech perception: Quantitative analysis of category effects in discrimination. Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo 30. 59-68.

GANDOUR, J. 1978. The perception of tone. In V. Fromkin (ed.), Tone: A Linguistic Survey. New York: Academic Press. 41-76.

HOWIE, J.M. 1976. Acoustical Studies of Mandarin Vowels and Tones. Cambridge: Cambridge University Press.

KLATT, D. 1980. Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America 67. 1636-1642.

LADEFOGED, P. and D.E. BROADBENT. 1957. Information conveyed by vowels. Journal of the Acoustical Society of America 29. 98-104.

LEATHER, J. 1983. Speaker normalization in perception of lexical tone. Journal of Phonetics 11. 373-382.

MILLER, J.L. 1981. Some effects of speaking rate on phonetic perception. Phonetica 38. 159-180.

MILLER, J.L. and P.D. EIMAS. 1977. Studies on the perception of place and manner of articulation. Journal of the Acoustical Society of America, 61, 835-845.

REPP, B.H., A.F. HEALY, and R.G. CROWDER. 1979. Categories and context in the perception of isolated, steady-state vowels. Journal of Experimental Psychology: Human Perception and Performance 5. 129-143.

SHIGENO, S. and H. FUJISAKI 1979. Effect of a preceding anchor upon the categorical judgment of speech and nonspeech stimuli. Japanese Psychological Research 21. 165-173.

SHIGENO, S. and H. FUJISAKI 1980. Context effects in phonetic and non-phonetic judgments. Annual Bulletin RILP 14. 217-224.

SHIGENO, S. 1986. The auditory tau and kappa effects for speech and nonspeech stimuli. Perception and Psychophysics 40. 9-19.

LIN, T. and WANG, W. Y-S. 1985. Shengdiao ganzhi wenti. Zhongguo Yuyan Xuebao. 2. 59-59 (Beijing).

THOMPSON, C.L. and H. HOLLIEN. 1970. Some contextual effects on the perception of synthetic vowels. Language and Speech. 13. 1-13.

WANG, W. S-Y. 1976. Language Change. Annals of the New York Academy of Sciences. 280. 61-72.

WANG, W. S-Y. and LI, K-P. 1967. Tone 3 in Pekinese. Journal of Speech and Hearing Research. 10. 629-636.

语境对词调感知的影响

Robert Allen Fox
The Ohio State University
Ying-Yong Qi
University of Arizona

本文观察了语境对三个从阴平到阳平连续变化体的影响作用。十八位受试者（其中讲汉语和讲英语的各九位）参加了测试。测试分成两个部分：第一部分这些连续体以单字形式作测试；第二部分这些连续体以成对的形式作测试，前字是阴平或是阳平。受试者要求对连续变化体中每一个音节和阴平（平调）或阳平（升调）的相似程度作出判断。

测试的结果表明，汉语和英语的受试者都显示出语境的影响作用。这说明语境影响作用和听觉的感知机制有关。