

## Perceptual Evidence of Tonal Coarticulation

Shu-hui Peng

peng@ling.ohio-state.edu

**Abstract:** The current study examines the extent to which tonal coarticulation is perceptible. Ten tokens each of two Taiwanese syllables (high tone and mid-falling tone) were placed in three contexts in which the following tone was high, mid, or mid-falling. The task of the listeners was to identify the whole phrase in a three-way forced choice after listening to the first syllable. Acoustic analysis showed that the high tone was contextually more stable than the mid-falling tone. Anticipatory assimilation of F0 was found between the mid-falling tone and its following tone. The results of the identification test indicated that contextual variability was perceptually detectable by listeners and contributed to the recognition of the following tone. Consistent with the patterns of assimilation, the following tone was more predictable when the phrase started with a mid-falling tone than when it started with a high tone. As is the case with segmental coarticulation, tonal coarticulation changes phonetic features of neighboring tones which contribute to the recognition of the tone. However, the variability of the tonal features are constrained by the phonological system of the language to maintain phonological contrast.

### Introduction

By representing assimilation as feature-spreading, non-linear phonologies recognize that these processes are rooted in patterns of segmental coarticulation attested in every language that has been studied instrumentally (e.g. Boyce, 1990). Production studies of languages such as Vietnamese (Han & Kim, 1974), Mandarin (Shih, 1988; Shen, 1990), Yoruba (Laniran, 1992), and Taiwanese (Lin, 1988; Peng, 1994) suggest that tonal coarticulation, like segmental coarticulation, is universal. Previous studies showed that tonal assimilation can be anticipatory or perseveratory in terms of direction of feature spreading. Anticipatory coarticulation refers to the phenomenon that tonal features of one tone assimilate to features of the following tone. Perseveratory assimilation is the carry-over effect of one tone on its following tone. However, previous studies also indicated that the effects of tonal coarticulation are sometimes dissimilatory. Tonal features of tones are changed to be less similar to each other in tone sequences.

The present study extends a production study of tonal coarticulation in Taiwanese (Peng, 1994). Taiwanese has seven distinctive tones: high, mid, high-falling, mid-falling, mid-rising, and two entering tones. All of the Taiwanese tones, except the entering tones, were examined in different tonal contexts in the previous production study. The productions of native Taiwanese speakers showed that anticipatory tonal assimilation occurred between contour tones (high-falling tone, mid-falling tone and mid-rising tone) and the following tone. For example, the F0 offset of the high-falling tone was higher in the context of high tone than in the context of mid-falling tone. Furthermore, in some cases, the pitch range of the

target tone shifted according to the pitch height of the following tone. The coarticulation patterns between level tones (high tone and mid tone) and the following tone were dissimilatory in the production of some subjects. The pitch contour of the high tone and the mid tone ended at a higher F0 value when followed by a low-onset tone, e.g. mid-falling tone, than when followed by a high-onset tone, e.g. high tone.

The current study examines the extent to which tonal coarticulation is perceptible. That is, can a tone be recognized from its influence on the F0 of a preceding syllable excised from context in the way that a following /i/ vs. /u/ can be recognized in the spectrum of an /s/ or /ʃ/ excised from a CV syllable (Yeni-Komshian & Soli, 1981)? Taiwanese underlying high-falling and mid tones in the context of high, mid and mid-falling tones were investigated in production and perception. Because of tone sandhi, these two target syllables in context are actually high tone and mid-falling tone. Acoustic analysis was done to show pitch contours of the target syllables. The target syllables were excised from the phrases to be used as stimuli in the perception test. The task of the listeners in the perception test was to identify the whole phrase in a three-way forced choice after listening to the target syllable.

## Method

### *Materials*

The six two-syllable phrases used in the experiment are shown in the Appendix. The first syllable was the target syllable which was either the high-falling (51) or mid (33) tone underlyingly, changed to the high tone (55) and mid-falling tone (21) respectively after tone sandhi. The target syllable was followed by high tone, mid tone or mid-falling tone. All the phrases were read ten times. There were sixty tokens in total. The first syllables were excised from the phrases to be used as stimuli in the identification test.

### *Subjects*

All phrases were read by one female native Taiwanese speaker. Nineteen native Taiwanese speakers participated in the perception test (11 female, 8 male). None of them reported having speech or hearing problems. All the subjects also speak Mandarin and English.

### *Recording & Acoustic Measurements*

Tokens were recorded using a TEAC V-427C stereo cassette deck in a sound-proof booth at the Ohio State University Linguistics Laboratory. The subject was asked to keep a steady speech rate and equal loudness in reading the phrases.

Duration and fundamental frequencies of the target syllables were measured. Duration was measured from wide-band spectrograms on a Kay DSP Sona-Graph 5500. The duration of each target syllable was measured from the release of the initial stop consonant to the offset of the diphthong. The fundamental frequency of each phrase was analyzed with the *Waves*<sup>TM</sup> signal editor (Entropic Research Laboratory, INC., 1993). The F0 of each target syllable was measured at three different points along the F0 contour: onset, midpoint and offset.

## Procedure

The identification test consisted of two sessions: high tone and mid-falling tone. Before each session, subjects were asked to read the three two-syllable phrases starting with the target syllable. There was a practice session before each test session. The practice session contained nine trials selected from all following tone conditions including high tone, mid tone and mid-falling tone. All the first syllables were presented to subjects four times in random order over headphones. There were 120 trials in each test session. Subjects heard the target tone at the beginning of each trial and could listen to it as many times as necessary by clicking on a button shown on the computer screen in front of them. There were also three other buttons labeled with the three phrases, Chinese characters. The task of the subject was to identify the phrase in a three-way forced choice after listening to the first syllable.

## Results

### Production data

Acoustic analysis showed similar patterns of assimilation to those found in previous studies (Lin, 1988; Peng, 1994). Pitch heights of tones were affected by tonal

environments. Measurements of duration and F0 indicated that the high tone was contextually more steady than the mid-falling tone in the productions of this speaker. Figure 1 shows mean durations and F0 values of the high tone in different tonal environments. Three different F0 measures are shown: onset, mid-point and offset of syllable. The effects of the following tone on the duration and F0 of the high tone were not significant. Although the F0 contour of the high tone ends at a slightly higher F0 value in the environment of mid-falling tone than in the other two tonal environments, the difference was not statistically significant.

Mean duration and F0 values of the mid-falling tone are shown in Figure 2. The effect of the following tone on duration of the mid-falling tone was marginally significant, [ $F(2, 27) = 4.929, p < 0.05$ ]. The duration of the mid-falling tone was slightly longer when followed by a mid-falling tone than when followed by a mid

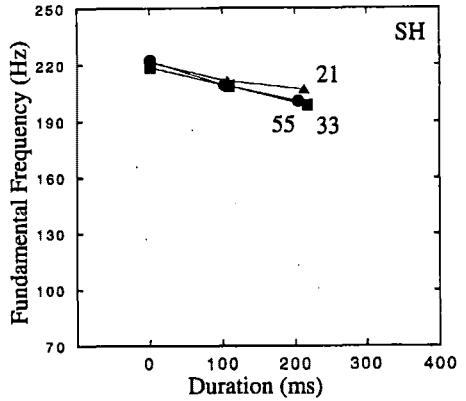


Fig. 1. Mean durations and F0s of the high target tone followed by different tones.

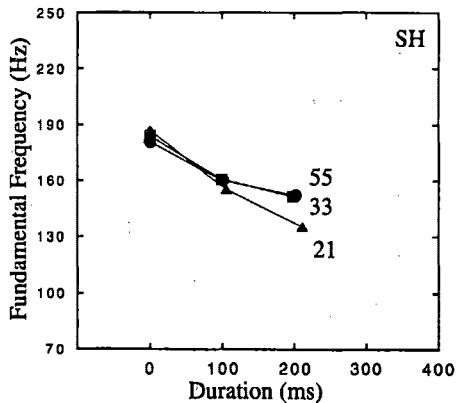


Fig. 2. Mean durations and F0s of the mid-falling tone followed by different tones.

tone. Tonal context had a significant effect on the F0 of the mid-falling tone at the offset of the F0 contour, [ $F(2, 27) = 22.444, p < 0.001$ ]. The F0 value of the mid-falling tone was lower when followed by the mid-falling tone than when followed by the other two tones.

### Identification data

The results of identification tests for the high tone and the mid-falling tone are shown in Figures 3 and 4. I values derived from the lower bound of the Green's area measure (Green, 1964) represent the measurement of signal detectability. The sensitivity of subjects to the signals they heard ranged from zero to one. A value of one refers to perfect sensitivity or performance. For example, if subjects identify a phrase with complete accuracy and without any false alarm response, then the I value for the phrase is one.

Figure 3 shows the I values with standard error for the high tone in different environments of the following tone. The effect of response type was significant, [ $F(2, 54) = 22.801, p < 0.001$ ]. Tukey post-hoc tests showed that the I value for the high tone in the context of the mid tone was higher than in the context of the high tone which was in turn higher than in the context of the mid-falling tone. However, t-tests indicated that the performance of subjects was better than chance ( $I = 0.33$ ) only in identifying the phrase with high tone followed by mid tone, [ $F(1, 18) = 2.397, p < 0.05$ ]. Figure 4 shows the I values of the mid-falling tone in different tonal environments. The effect of response type was significant, [ $F(2, 54) = 8.736, p < 0.01$ ]. The performance of subjects was better in the environment of the mid-falling tone than in the other two environments. t-tests indicated that performance for all response types was better than chance, [high tone  $F(1, 18) = 5.175, p < 0.001$ ; mid tone  $F(1, 18) = 3.669, p < 0.01$ ; mid-falling tone  $F(1, 18) = 2.528, p < 0.05$ ].

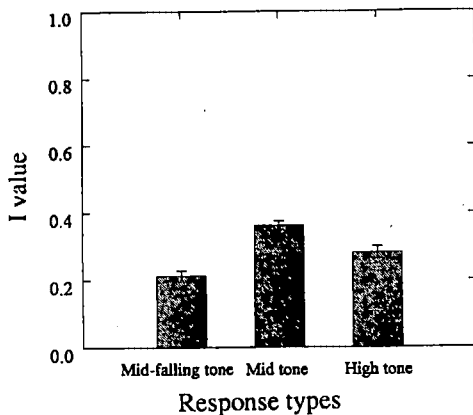


Fig. 3. Mean I values of different response types for the high tone.

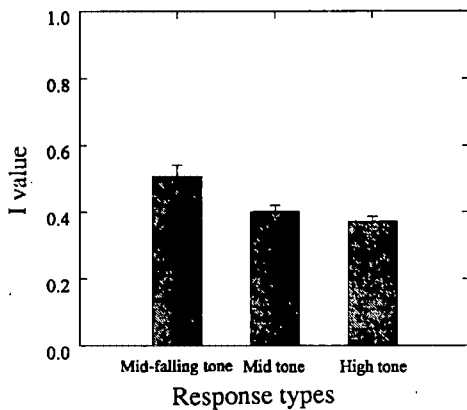


Fig. 4. Mean I values of different response types for the mid-falling tone.

## Discussion

Just as segmental coarticulation does, tonal coarticulation changes phonetic features of neighboring tones which contribute to the recognition of the tone. However, the variability of the tonal features are constrained by the phonological system of the language, to maintain phonological contrast. The mid-falling tone showed more acoustic variation according to tonal context than the high tone. Duration and F0 values of the mid-falling tone changed according to the tonal quality of the following syllable. The mid-falling tone was longer in duration and lower in F0 at the offset of the F0 contour when preceding mid-falling tone than when preceding the high tone and the mid tone. The effect was stronger on F0 than on duration. Anticipatory assimilation of F0 was found between the mid-falling tone and its following tone. The F0 height of the mid-falling tone at the offset of the F0 contour assimilated to the onset F0 height of its following tone. The high tone, on the contrary, did not show much variation in both duration and F0 values in different tonal contexts.

The pitch contour of the high tone was contextually stable to maintain perceptual contrast with other tones. A perception study (Lin & Repp, 1989) of Taiwanese tones using synthesized pitch contours showed that high tone was easily confused with mid-rising tone and high-falling tone if its pitch movement (tone shape) was changed. When a syllable was synthesized with the pitch height of the high tone, but with the pitch movement of the mid-rising tone or the high-falling tone or with the pitch movements intermediate between the high tone and the mid-rising tone or the high-falling tone, the syllable tended to be perceived by listeners as mid-rising tone or high-falling tone. Therefore, pitch movement is an important perceptual cue for the high tone and shows little variation in different tonal contexts. The perception study (Lin & Repp, 1989) also showed that pitch movement was also an important cue for perceiving the mid-falling tone. However, because the mid-falling tone begins at a low F0 value and ends at the bottom of the pitch range of Taiwanese tonal system, it is not likely to be perceived as other tones unless the tone shape is drastically changed to be flat as the mid tone or rising as the mid-rising tone, or the F0 onset is raised as high as the high tone or high-falling tone. The change of F0 offset due to tonal assimilation does not affect the tonal quality of the mid-falling tone, but provides some perceptual cues for recognizing the following tone.

Contextual variability was perceptually detectable by listeners and contributed to the recognition of the following tone. The results of the identification test indicated that listeners could predict the tone of the second syllable in a phrase from the variation of the tonal contour of the first syllable. As found in perception of segmental coarticulation (Yeni-Komshian & Soli, 1981), listeners could anticipate the tonal features of one tone from the preceding tone possibly to increase efficiency of speech perception. Consistent with the patterns of assimilation, the following tone was more predictable when the phrase started with a mid-falling tone than when it started with a high tone.

## Future Study

It was found in the previous study (Peng, 1994) that tonal dissimilation occurred between the high tone or the mid tone and the following tone. One may wonder whether F0 cues derived from tonal dissimilation will contribute to the perception of the high tone or the mid tone as found in tonal assimilation. A perception study

using synthesized pitch contours will indicate the extent to which tonal dissimilation is perceptually detectable. The synthesized pitch contour of the target tone will have a lower F0 offset when followed by a high-onset tone than when followed by a low-onset tone. This perception study will show whether listeners can predict one tone from the F0 cues in the preceding tone generated by tonal dissimilation.

## References

- Boyce, S. (1990) Coarticulatory organization for lip rounding in Turkish and English. *Journal of Acoustical Society of America*, 88 (6), 2584-2595.
- Entropic Research Laboratory, INC. (1993) *Waves+ 5.0*. AT&T Bell Laboratories. Washington, D. C..
- Green, D. M. (1964) General prediction relating yes-no and forced choice results. *Journal of Acoustical Society of America*, 36, 1042.
- Han, M. & Kim, K.-O. (1974) Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics*, 2, 223-232.
- Laniran, Y. (1992) *Intonation in Tone Languages: The Phonetic Implementation of Tones in Yoruba*. PhD Dissertation, Cornell University.
- Lin, H.-B. (1988) *Contextual Stability of Taiwanese Tones*. PhD Dissertation, University of Connecticut.
- Lin, H.-B. & Repp, B. (1989) Cues to the perception of Taiwanese tones. *Language and Speech*, 32(1), 25-44.
- Peng, S.-H. (1994) Effects of Prosodic Position and Tonal Context on Taiwanese Tones. *OSU Working Papers in Linguistics*, 44, 166-190.
- Shen, X. (1990) Tonal coarticulation in Mandarin. *Journal of Phonetics*, 18, 281-295.
- Shih, C. (1988) Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory*, 3, 83-109.
- Yeni-Komshian, G. & Soli, S. (1981) Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of Acoustical Society of America*, 70 (4), 966-975.

## Appendix

- /kau<sup>51</sup> taŋ<sup>55</sup>/ --> [kau<sup>55</sup> taŋ<sup>55</sup>] nine years  
/kau<sup>51</sup> tai<sup>33</sup>/ --> [kau<sup>55</sup> tai<sup>33</sup>] nine generations  
/kau<sup>51</sup> tā<sup>21</sup>/ --> [kau<sup>55</sup> tā<sup>21</sup>] nine burdens  
  
/kau<sup>33</sup> kaŋ<sup>55</sup>/ --> [kau<sup>21</sup> kaŋ<sup>55</sup>] time-consuming  
/kau<sup>33</sup> kau<sup>33</sup>/ --> [kau<sup>21</sup> kau<sup>33</sup>] thick  
/kau<sup>33</sup> kua<sup>21</sup>/ --> [kau<sup>21</sup> kua<sup>21</sup>] thick lid