LINGUISTICS OF THE TIBETO-BURMAN AREA
VOLUME 15:1 — SPRING 1992

Tonal coarticulation in Thai disyllabic utterances: a preliminary study*

Jack Gandour and Siripong Potisuk
Purdue University
Sumalee Dechongkit
Mahidol University
Suvit Ponglorpisit
Prasart Neurological Hospital

One of the major problems in describing speech is that all utterances involve *coarticulation*, which refers to the phenomenon whereby a given speech sound is altered in its phonetic manifestation depending upon influences from adjacent sounds. Of particular interest is *anticipatory* or *look ahead* coarticulation, which involves one speech sound being influenced by subsequent sounds, and *perseverative* or *carryover* coarticulation, which involves one speech sound being influenced by preceding sounds. In consonants and vowels, coarticulatory effects are bidirectional, which means that both anticipatory and perseverative effects are found to occur in languages of the world. Though present, the effects may be asymmetric in particular languages. For example, by comparing the effect of adjacent vowels on the articulation of consonants in English and French, Ladefoged (1972:280) suggested that anticipatory effects are greater than perseverative in English whereas perseverative effects are equal in influence to anticipatory effects in French. In tones, coarticulatory effects occur in much the same way as they do between consonants and vowels. It has been suggested that there seems to be a tendency for coarticulatory effects to be perseverative rather than anticipatory between tones (Ladefoged 1982:230-231).

Research on coarticulation has focused mainly on consonants and vowels. Only a few acoustic-phonetic studies have been carried out on coarticulation involving tones. Those studies have focused on tone languages in the Far East: Vietnamese, Mandarin, and Thai. The six lexical tones of

* Acknowledgments: This material is based upon work supported in part by the National Institute on Deafness and other Communication Disorders under Grant No. DC00515-05 and the United States Information Agency in conjunction with the Thailand-United States Educational Foundation under a Fulbright Senior Researcher Award (1st author). The speech data were collected in Bangkok with the approval of the National Research Council of Thailand during the period from December 1988 to July 1989. In Bangkok, the first author was affiliated with the Division of Communication Disorders, Faculty of Medicine, Mahidol University, and Prasart Neurological Hospital. Reprint requests should be sent to: Jack Gandour, PhD, Department of Audiology and Speech Sciences, Heavilon Hall, Purdue University, West Lafayette, IN 47907-1353.

Vietnamese show considerable variation in both slope and overall height of fundamental frequency (F_0) (Han & Kim 1974). Perseverative effects are greater than anticipatory. The four lexical tones of Mandarin show changes in F_0 height but not slope (Shen 1990). Anticipatory and perseverative effects are comparable in magnitude of effect. The five lexical tones of Thai show some changes in F_0 height and slope as a function of the preceding or following tone (Palmer 1969). Though no overall assessment of the magnitude of anticipatory and perseverative effects was made, Palmer pointed out that the high tone shows considerable variation in F_0 height and slope when it occurs in final position but not initial position. Of the data presented in Abramson (1979), the mid tone in final position shows considerable variation in F_0 height and slope in the first 30% of its duration. The initial portion of the falling tone in final position, on the other hand, shows little variation in F_0 height and slope except when following a mid tone.

Because of methodological limitations in the aforementioned studies on tonal coarticulation, it is difficult to determine to what extent coarticulatory effects are symmetric or asymmetric across tone languages, and to what extent both F_0 height and slope vary as a function of preceding or following tones. Only Shen provides statistical treatment of coarticulatory effects. But her statistical analysis of tonal coarticulation in Mandarin is restricted to variation in F_0 height of three measurement points only: tonal onset, turning point of tonal contour, and tonal offset. Differences in F_0 slope are interpreted by visually inspecting graphs of the tonal contours. In the other studies on Vietnamese and Thai, both F_0 height and slope differences are interpreted by visual inspection of graphs exclusively. In order to develop a model of tonal coarticulation, it is essential that our measurement procedure allows for a quantitative assessment of F_0 height and slope throughout the duration of the tonal contours. Without such quantification, we are constrained by the limits of our eyeballs. Only inferential statistics provides a technique for us to distinguish between significant differences in tonal contours and random error. Importantly, such quantitative measures will also make it possible to determine the extent of coarticulatory influences into adjacent tones. In the case of perseverative coarticulation, how far into the following syllable does the influence of the preceding tone extend? In the case of anticipatory coarticulation, how far into the preceding syllable does the influence of the following tone extend?

Another methodological limitation of previous studies on tonal coarticulation is the lack of any quantitative measure of intrasubject or intersubject variability in tone production. It is well known that variability occurs both between different speakers on the same occasion and within the same speaker on different occasions. Only a few speakers participated in previous studies of tonal coarticulation. In the absence of a quantitative measure of variability, it it difficult to determine the degree to which the coarticulatory effects are consistent between different speakers and within the same speaker. Using a quantitative measure of variability of production of Thai tones in isolated monosyllabic words, Gandour, Potisuk, Ponglorpisit, & Dechongkit (1991) found that the degree of interspeaker variability is greater than intraspeaker across all five tones, that young and old speakers exhibited the same pattern of variability, and that variability in tone production differed depending on the lexical tone. Such an approach can easily be extended to the study of variability of production of tones in connected speech.

The aim of this preliminary study is to introduce a new methodology for investigating tonal coarticulation that is quantitative in nature and that is appropriate for inferential statistical analysis. Quantitative measures of both F_0 height and slope are to be developed for assessment of coarticulatory effects. Data are drawn from a set of disyllabic tonal sequences that occur in Thai. The study is preliminary in the sense that only a few of the 25 possible combinations of the five Thai tones are investigated, and that those particular tone sequences permit an assessment of perseverative tonal coarticulation only. There is nothing preliminary, however, about the need for such a quatitative approach if we are to make verifiable statements about tonal coarticulation within a particular language and across languages.

As such, it represents the first in a series of studies already underway in our laboratory on tonal coarticulation in Thai. It is also part of a larger series of studies designed to explore tonal coarticulation in Thai brain-damaged patients with unilateral left hemisphere and right hemisphere lesions. By including both young and old normal speakers, we are able to assess whether there are any age-related effects on variability in production of tone in connected speech. In our future studies of tonal coarticulation in Thai brain-damaged patients, it will be important to distinguish whether differences in tone production are to be attributed to the normal aging process rather than neurological lesions.

I. Method

1. Subjects

Nineteen adult speakers of Thai participated in the study: 10 'young' adults and 9 'old' adults. All 10 speakers in the young group were male and had completed 13 years of formal education; the average age of the young group was 26.7 (*SD* = 2.6). Four of the speakers in the old group were male, five were female; all had completed 4 years of formal education; the average age of the old group was 56.9 (*SD* = 2.3). All speakers had spent their entire lives in the Bangkok metropolitan area. Two different age groups as well as female speakers in the old group were included for purposes of comparison in a larger project on tonal coarticulation in Thai brain-damaged adults.

2. Speech Materials

Three pairs of disyllabic phrases were selected for preliminary investigation of tonal coarticulation in Thai (see Table 1). Each utterance consisted of a noun followed by a verb. Within pairs, phrases were identical segmentally except for a vowel length difference in pair (1). A plosive was selected to begin the second syllable of phrases to facilitate segmentation in acoustic analysis. Phrases were minimally or near-minimally distinguished by the tone on the first syllable. The tone on the second syllable was held constant. These particular tonal sequences made it possible to evaluate perseverative or carryover coarticulatory effects.

Table 1

Speech Stimuli

Pair (1)	falling-low /klɔŋkaw/ 'old camera'	mid-low /klɔɔŋkaw/ 'old drum'
Pair (2)	falling-mid /khaawdii/ 'good rice'	low-mid /khaawdii/ 'good news'
Pair (3)	high-rising /maakhaaw/ 'white horse'	rising-rising /maakhaaw/ 'white dog'

3. Recording Procedure

All phrases were printed in large Thai script on 3" x 5" cards. A total of 60 cards (3 pairs x 2 phrases x 10 tokens) was presented in random order to subjects. They were instructed to read the phrases at a comfortable speaking rate. To avoid start and end effects, extra cards were placed at the top and bottom of the deck. To avoid list reading effects, a sufficient pause was provided between items to ensure that subjects maintained a uniform speaking rate. Subjects' utterances were recorded in a reasonably quiet room in a single session using a Sony ECM-66B microphone and a Marantz PMD-420 taperecorder.

4. Listening Procedure

A closed-set identification procedure was used in the listening tests. Under this procedure, a trial consisted of the presentation of a single phrase. Each subject's productions of the six phrases were presented in random order. The response interval for each trial was 3.5 seconds.

Five native Thai adults participated in the listening tests. Listeners were instructed to identify each stimulus by circling one of the six phrases on their response sheets corresponding to what they heard or by circling 'other'. A block of five practice trials consisting of the same set of stimuli was presented before proceeding with the actual experiment. Test tapes were played on a Marantz PMD420 taperecorder, and the signal was presented through TDH-39 headphones at a comfortable listening level.

5. Measurement Procedure

All utterances were low-pass filtered at 4 kHz and digitized at a 10 kHz sampling rate for speech analysis using ILS (Interactive Laboratory System, Signal Technology, Inc.) programs implemented on a Zenith 386 PC-AT computer. F_0 contours were extracted using a cepstral method of analysis. To minimize effect of consonantal perturbations at the beginning of F_0 contours, changes in F_0 that were 10 Hz or greater across a 10 ms time frame were excluded. Cepstral analysis sometimes failed to extract an F_0 contour from an obviously periodic audio waveform or from an audio waveform with lengthy stretches of aperiodicity. Because of such measurement problems, 1% and 4% of utterances produced by young and old speakers, respectively, were eliminated from the corpus.

6. Data Analysis

Only those utterances which were identified correctly by at least four of the five native Thai listeners were retained for analysis. By this criterion, 99% of utterances of both young and old speakers were retained for data analysis.

Because all tokens of each tone were not of equal duration, the F_0 contours were equalized for duration on a percentage scale. In light of well-known differences in absolute F_0 between male and female speakers, the F_0 contours were also normalized on a *z* score scale (cf. Rose 1987). By such normalization, F_0 contours may be compared across speakers.

Average F_0 contours do not provide any indication of intraspeaker token-to-token variation in tonal production. Unless we quantify intra-speaker variation, it is difficult to assess inter- and intraspeaker variability in the production of Thai tones. To compare the point-to-point variability of F_0 contours for all tokens of each tone, an ensemble-averaging approach was taken (cf. Atkinson 1976:441-442). All tokens of a tone made up an ensemble of F_0 contours for that tone by that speaker. The ensemble mean and standard deviation were determined at each 10% interval. The ensemble mean can be considered as the target F_0 contour for a particular tone, the ensemble standard deviation as a measure of the random perturbations that result in deviations from the target contour.

To assess changes in F_0 height and slope of the final syllable as a function of the tonal category of the initial syllable, four measurement intervals and measurement points, respectively, were selected in the final syllable. In the case of F_0 height, there were four different time intervals varying in duration. The first time interval provided a measure of average F_0 height over 10% of the final syllable starting from the beginning, the second a measure of average F_0 height over 30%, the third a measure of average F_0 height over 50%, and the fourth a measure of average F_0 height over 100% or the total duration. In the case of F_0 slope, there were four different points of measurement. The first point was at 10% of the total duration of the final syllable, the second at 30%, the third at 50%, and the fourth at 100%. The normalized F_0 contour of each syllable was fitted with a polynomial curve of fifth degree using the least squares method ($f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$). The fifth degree polynomial provided the best overall fit to the data. Higher degree polynomials resulted in little or no improvement in overall fit. Given an F_0 contour, the slope of $f(x)$ at any point x was computed by evaluating the derivative of $f(x)$.

II. Results

Figures 1-4, respectively, display average F_0 contours of each of the three tonal sequences for a young male speaker, an old female speaker, a group of young speakers, and a group of old speakers.

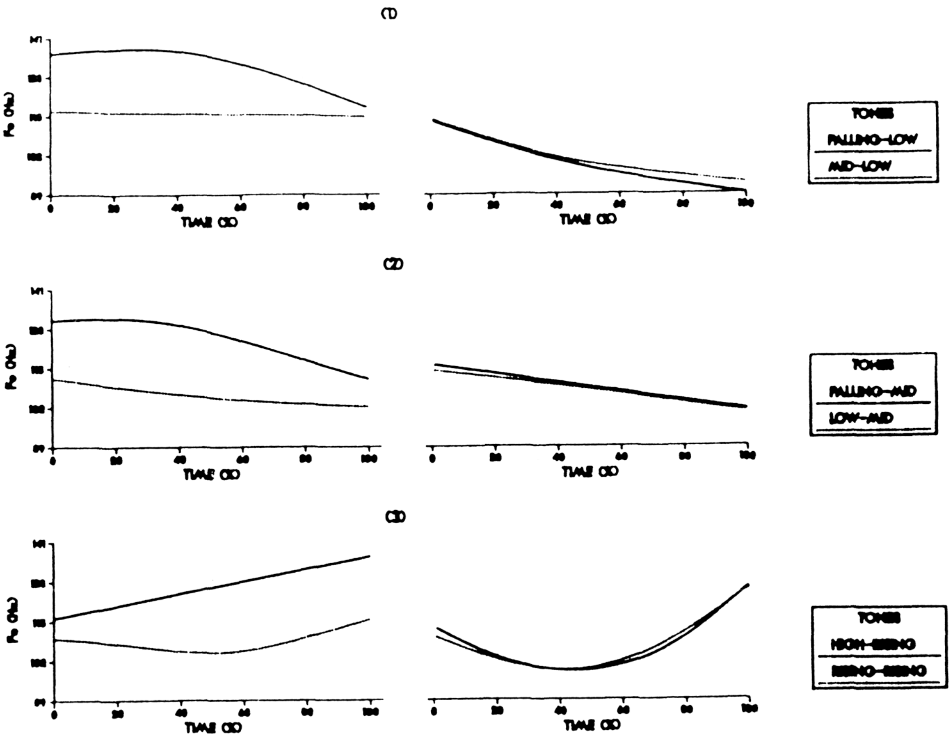

Figure 1.

Average F_0 contours in each pair of disyllabic utterances produced by a young speaker. Each syllable is normalized for time on a percentage scale. Pair (1) = /klɔŋkaw/ falling-low 'old camera', /klɔɔŋkaw/ mid-low 'old drum'; Pair (2) = /khaawdii/ falling-mid 'good rice', /khaawdii/ low-mid 'good news'; Pair (3) = /maakhaaw/ high-rising 'white horse', /maakhaaw/ rising-rising 'white dog'.

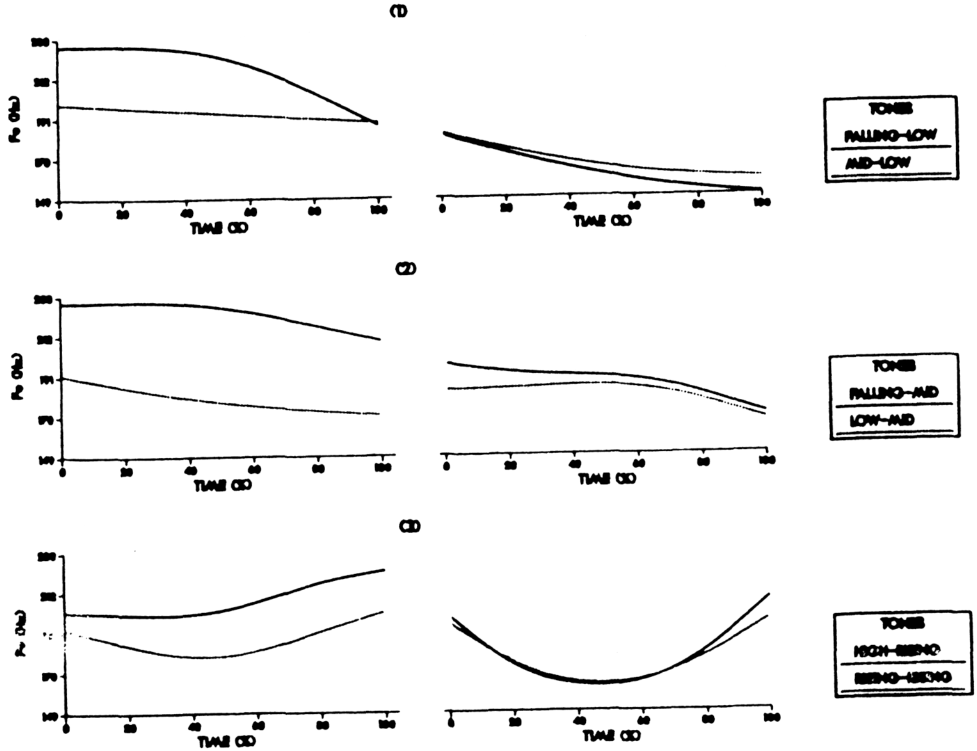

Figure 2

Average F_0 contours in each pair of disyllabic utterances produced by an old speaker. Each syllable is normalized for time on a percentage scale. See also caption to Figure 1.

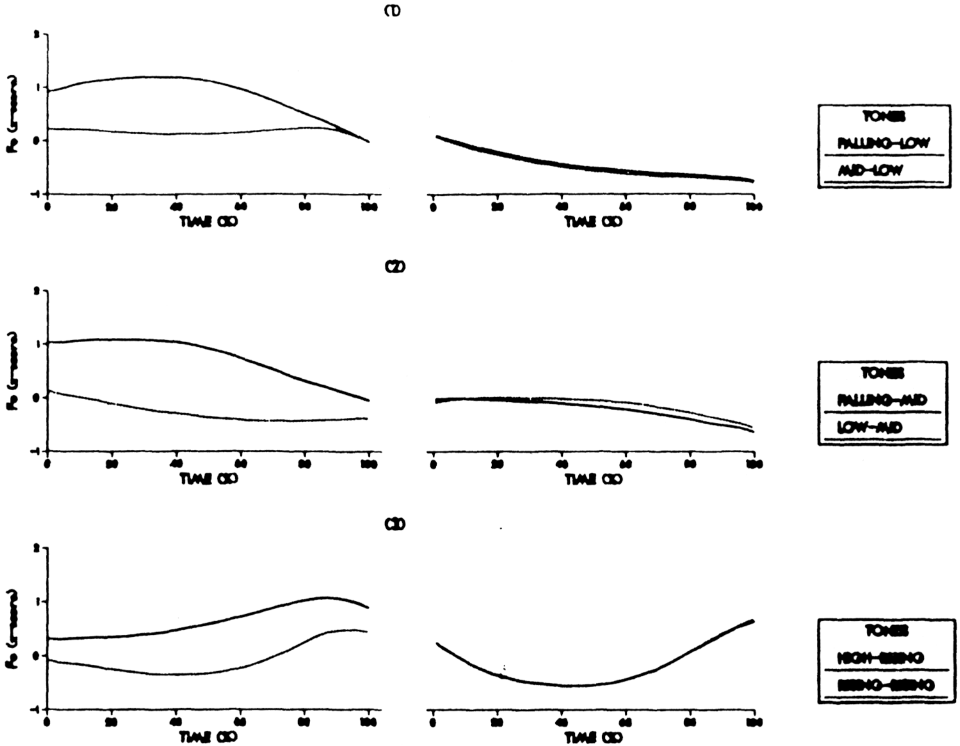

Figure 3

Average F_0 contours in each pair of disyllabic utterances produced by a group of young speakers (n = 10). Each syllable is normalized for time on a percentage scale. See also caption to Figure 1.

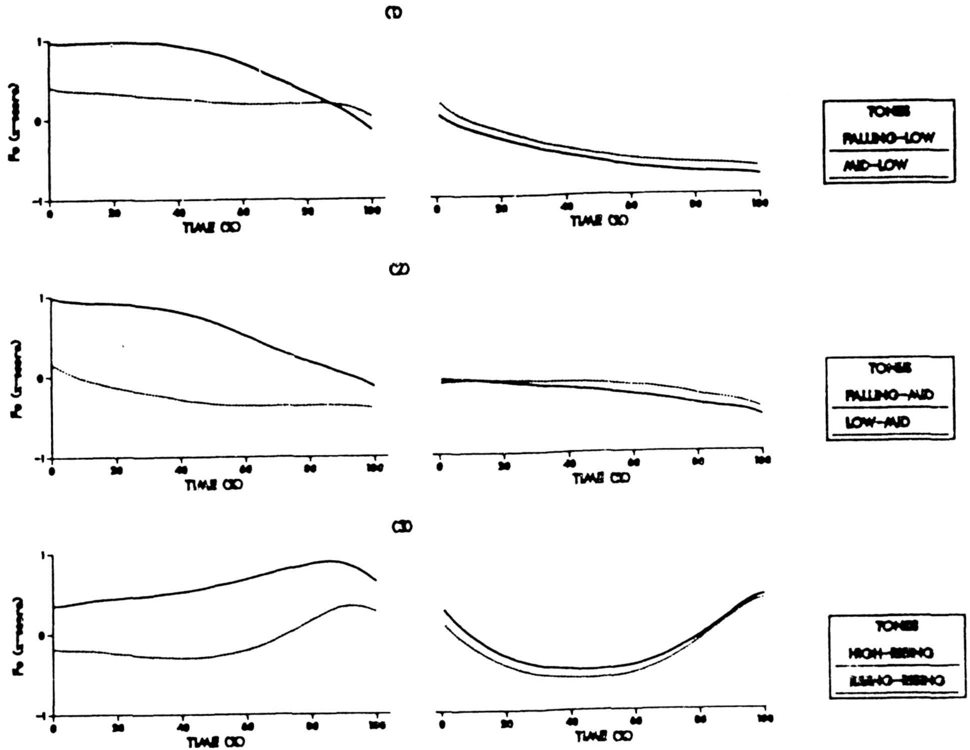

Figure 4

Average F_0 contours in each pair of disyllabic utterances produced by a group of old speakers (n = 9). Each syllable is normalized for time on a percentage scale. See also caption to Figure 1.

To evaluate the magnitude of differences in F_0 height and slope of the final syllable as a function of the tonal category of the initial syllable, a series of two factor analyses of variance (group x final syllable) with final syllable as repeated measures was performed on F_0 values and slope coefficients, respectively, at the 10%, 30%, 50%, and 100% timing positions in the final syllable. In the case of F_0 height, results indicated a significant main effect of final syllable at each of the four time intervals [10%- $F(5,1078) = 27.80$, $p < .0001$; 30%-$F(5,1078) = 7.77$, $p < .0001$; 50%-$F(5,1078) = 56.58$, $p < .0001$; 100%-$F(5,1078) = 56.25$, $p < .0001$]. Neither the group main effect nor the two-way interaction of group and final syllable reached significance. A set of three planned comparisons of F_0 height in the final syllables (falling-low vs. mid-low, falling-mid vs. low-mid, high-rising vs. rising-rising) at each of the four measurement intervals revealed no significant differences (see Table 2). In the case of F_0 slope, results indicated a significant main effect of final syllable at each of the four time points [10%- $F(5,1068) = 234.86$, $p < .0001$; 30%-$F(5,1068) = 121.52$, $p < .0001$; 50%-$F(5,1068) = 223.43$, $p < .0001$; 100%-$F(5,1068) = 61.14$, $p < .0001$]. Neither the group main effect nor the two-way interaction of group and final syllable reached significance. A set of three planned comparisons of F_0 slope in the final syllables (falling-low vs. mid-low, falling-mid vs. low-mid, high-rising vs. rising-rising) at each of the four measurement points revealed a significant difference only for falling-mid vs. low-mid at the 30% (see Table 3). Inspection of the F_0 contours show that the mid tone is rising at the 30% point of the duration of the final syllable when it occurs after the low tone, but falling when it occurs after the falling tone (see Figures 3-4).

Table 2

A Priori Multiple Planned Comparisons: F_0 Height

	Time Interval within Final Syllable			
Tonal Sequence	10%	30%	50%	100%
FL vs. ML	2.53	3.24	3.50	4.31
FM vs. LM	4.31	3.69	1.63	0.43
HR vs. RR	0.00	0.00	0.00	0.83

Notes: Values are expressed as F with 1,5 degrees of freedom. FL = falling-low, ML = mid-low; FM = falling-mid, LM = low-mid; HR = high-rising, RR = rising-rising.

Table 3

A Priori Multiple Planned Comparisons: F_0 Slope

Tonal Sequence	Time Point within Final Syllable			
	10%	30%	50%	100%
FL vs. ML	1.79	0.23	0.01	1.31
FM vs. LM	3.41	11.15 *	3.04	1.32
HR vs. RR	1.05	0.38	0.05	0.86

Notes: Values are expressed as F with 1,5 degrees of freedom. FL = falling-low, ML = mid-low; FM = falling-mid, LM = low-mid; HR = high-rising, RR = rising-rising. * = $p < .05$.

Figures 5 and 6 display the ensemble-averaged F_0 contours and their standard deviations for young and old speakers, respectively. To evaluate the magnitude of differences in variability as a function of group and tone sequence, an analysis of variance was performed on log-transformed ensemble standard deviations in z-score units. Results from a two factor analysis of variance (group x sequence) with sequence as repeated measures revealed a significant main effect of sequence, $F(5,85) = 18.348$, $p < .0001$, indicating that there were differences in variability of production of the six tonal sequences. Neither the main effect of group, $F(1,17) = 1.324$, $p < .2657$, nor the two-way interaction of group and sequence, $F(5,85) = 2.050$, $p < .0797$, reached significance, meaning that young and old speakers exhibited essentially the same parallel patterns of variability across the six tonal sequences.

Pooling across groups, the mean ensemble standard deviations in z-score units were 0.29, 0.23, 0.34, 0.29, 0.39, and 0.36 for falling-low, mid-low, falling-mid, low-mid, high-rising, and rising-rising, respectively. The tone sequences were ranked in order from greater to lesser variability: high-rising, rising-rising, falling-mid, low-mid, falling-low, mid-low. Posthoc Student-Newman-Keuls multiple paired comparisons indicated that variability in F_0 of the high-rising sequence was significantly (α=.01) greater than the others, the mid-low sequence lesser than the others. The remaining four sequences were intermediate with rising-rising and falling-mid greater in variability than low-mid and falling-low. No other differences in variability among tonal sequences reached significance.

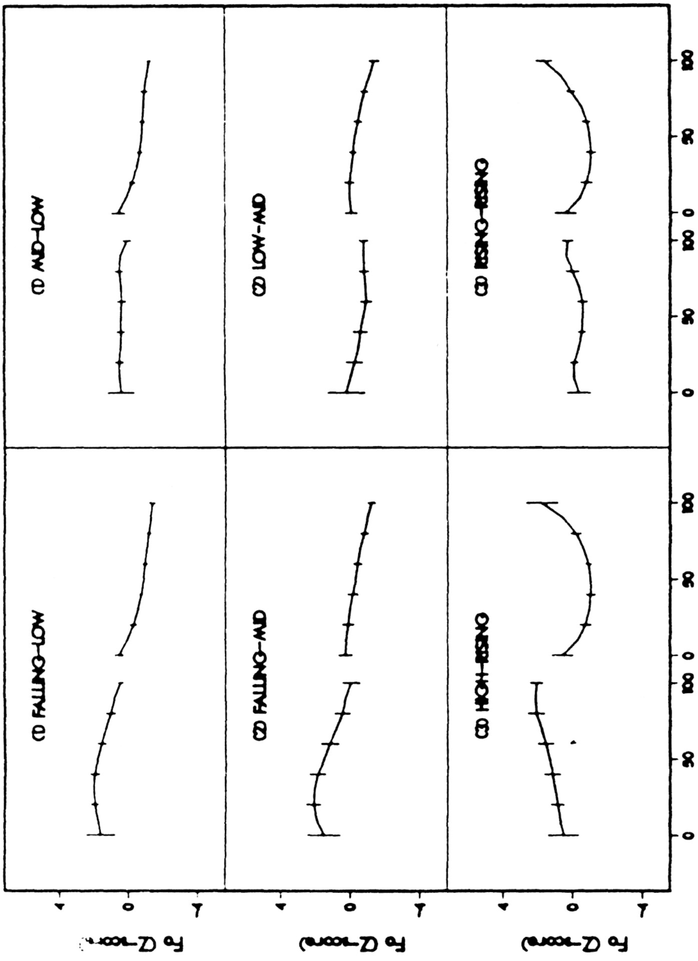

Figure 5
Variability of F_0 contours in each pair of disyllabic utterances produced by a young speaker. Each syllable is normalized for time on a percentage scale. See also caption to Figure 1.

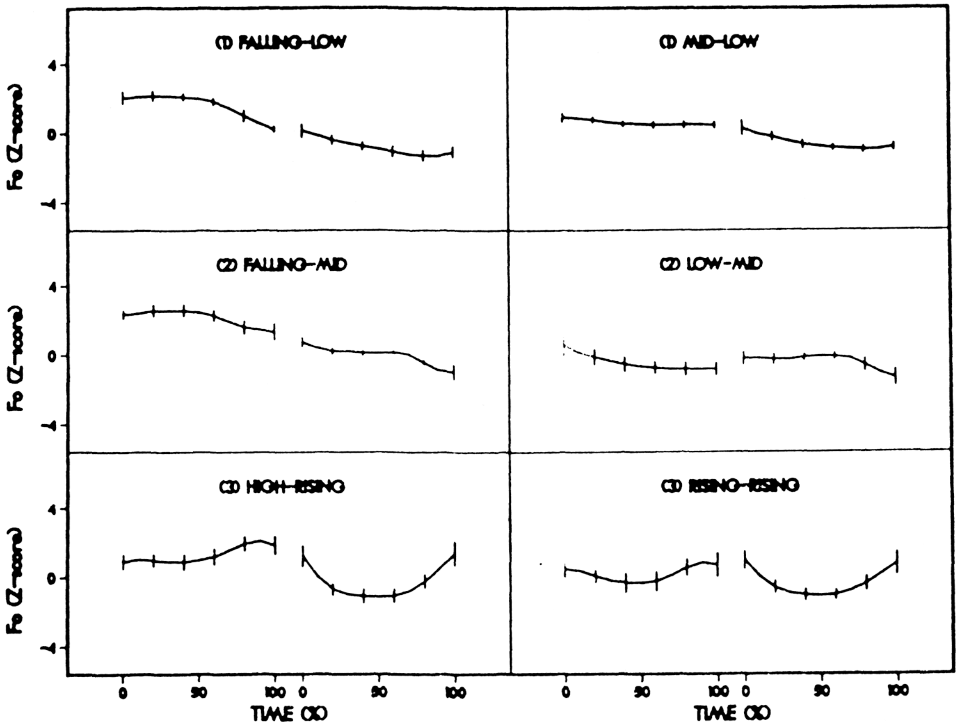

Figure 6

Variability of F_0 contours in each pair of disyllabic utterances produced by an old speaker. Each syllable is normalized for time on a percentage scale. See also caption to Figure 1.

III. Discussion

The findings from this preliminary study of tonal coarticulation in Thai seem to indicate that perseverative or carryover effects on F_0 height and slope are minimal. No differences in F_0 height of the final syllable can be attributed to differences in tonal category of the preceding syllable. Indeed, only the F_0 slope of the mid tone varies depending on the tonal category of the preceding syllable, rising in the beginning portion when following a low tone, falling when following a falling tone. Of the three pairs, pair 2 is the only one in which the two opposing tones in the preceding syllable differ in terms of the relationship of F_0 height of the preceding syllable to that of the final syllable. In pair 1, the terminal portions of both the falling and mid tones in the preceding syllable are higher than the beginning portion of the low tone in the final syllable; in pair 2, the falling tone is higher, the low tone is lower than the mid tone in the final syllable; in pair 3, both the high and rising tones are higher than the rising tone in the final syllable. Thus, the extent to which tonal coarticulation is triggered depends upon the magnitude of the difference in F_0 height and slope of the preceding tone. Similar perseverative effects on F_0 height as well as slope are anticipated in a full-scale investigation of tonal coarticulation in Thai.

The current findings are in partial agreement with those in Palmer (1969) and Abramson (1979). In Palmer's study, the mid tone in final position similarly shows more variation as function of the preceding tones than either the low or rising tones. But in Abramson's study, the mid tone appears to differ substantially in F_0 height but not slope when following the falling and low tones. Differences in findings between this study and those of Palmer and Abramson, however, are most likely to be attributed to differences in methodology. First, the disyllabic phrases were produced in isolation instead of being embedded in a sentence frame. Changes in timing patterns in longer utterances could lead to an increase in tonal coarticulation. Second, the initial consonant of the final syllable was a plosive in all three pairs, and vowel quality was held constant within each pair. In the studies by Palmer and Abramson, sonorants and fricatives as well as plosives occurred as the initial consonant of the final syllable, and vowel quality was not held constant. Tonal coarticulation may be inhibited by the presence of plosives. Because of aerodynamic effects on the larynx as a result of complete obstruction in the supraglottal cavity, perhaps it is only with intervening fricatives and sonorants that tonal coarticulation manifests itself strongly. Third, unlike the previous studies by Palmer and Abramson, only a subset of all possible disyllabic tonal sequences were examined, and all three pairs were designed to assess the effects of perseverative tonal coarticulation. In this preliminary study, the effects of anticipatory tonal coarticulation were not investigated. Further research may reveal that anticipatory tonal coarticulation in Thai is more robust than perseverative.

In view of the preliminary nature of this investigation, it is difficult to compare these findings with those on tonal coarticulation in Vietnamese (Han & Kim 1974) and Mandarin (Shen 1990). Nevertheless, one potential area of disagreement is worthy of mention. Shen explicitly makes the claim that "tonal coarticulation affects only F_0 height, not F_0 direction [slope]" (p.293). The findings herein contradict her claim. The slope of the mid tone in pair 2 is affected by the preceding falling and low tones, whereas the height of the mid tone is not. Only further research can resolve whether the differential sensitivity of F_0 height and slope to tonal coarticulation is a language-specific phenomenon or one that occurs across tone languages. It is also possible that Shen's claim reflects the limitations of her measurement procedure. Her assessment of F_0 slope is based on visual inspection of three measurement points: onset, turning point, and offset. The method used herein to quantify F_0 slope and height offers a far more sensitive measure of changes in F_0 slope due to tonal coarticulation.

The degree of variability of F_0 production varies depending on the particular combination of tones in a tonal sequence. In this study, the high-rising sequence exhibited the largest degree of variability, the mid-low sequence the smallest. Gandour et al. (1991), however, found that tones produced in isolated monosyllables were ranked in order from greater to lesser variability: high, mid, low, rising, falling. Their findings, for example, would lead us to expect the mid-low sequence to exhibit a larger degree of variability than rising-rising. But just the opposite occurred. This discrepancy in findings between this study and Gandour et al. (1991) suggests that F_0 variability depends not simply on the tone itself, but rather the particular combination of tones. If a particular sequence influences variability, then we might expect F_0 variability to differ depending on particular permutations of the same tones. That is, the high-rising sequence may turn out to be more variable than rising-high. The next logical study is an investigation of all possible disyllabic tonal combinations. The question of variability of tone production in connected speech clearly remains open for further empirical research.

One of the most pervasive characteristics of aging is reduction in speed due to neural and muscular changes. The fine motor control required to adjust vocal fold vibration in connected speech might be expected to change as a function of normal aging. There was no significant increase in variability of tonal productions by older speakers in this study. Ages of the older speakers ranged from 53 to 60 years of age. Their chronological age may not be a valid reflection of physiological age. They may have aged less physiologically than their chronological age would indicate. Gandour, Ponglorpisit, & Dechongkit (1991) similarly reported that variability in the production of voice onset time of Thai word-initial stop consonants did not differ significantly as a function of age. To the extent that chronological and physiological age are not commensurate, the analysis of the relationship of normal aging to speech production may be confounded.

In this preliminary study of tonal coarticulation, we have established a sound, rigorous methodology for exploring the effects of tonal coarticulation as measured by changes in F_0 height and slope. To extend this line of research, two studies of tonal coarticulation in disyllabic tone sequences are currently underway in our laboratory. One study explores the effects of anticipatory tonal coarticulation in bisyllabic noun compounds produced in isolation, the other the effects of both anticipatory and perseverative tonal coarticulation in all 25 possible tonal sequences produced in a sentence context.

REFERENCES

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

Atkinson, J. 1976. "Inter- and intraspeaker variability in fundamental voice frequency." *Journal of the Acoustical Society of America*, 60:440-445.

Gandour, J., Potisuk, S., Ponglorpisit, S., & Dechongkit, S. 1991. "Inter- and intraspeaker variability in fundamental frequency of Thai tones." *Speech Communication*, 10:355-372.

______. 1991. "Age-related effects on the production of voice onset time in Thai word-initial stops." Manuscript submitted for publication.

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

Ladefoged, P. 1972. "Phonetic prerequisites for a distinctive feature theory." In A. Valdman (Ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, pp. 273-285. The Hague: Mouton.

______. 1982. *A Course in Phonetics*. 2nd ed. New York: Harcourt Brace Jovanovich.

Palmer, A. 1969. "Thai tone variants and the language teacher." *Language Learning*, 19:287-299.

Rose, P. 1987. "Considerations in the normalisation of the fundamental frequency of linguistic tone." *Speech Communication*, 6:343-351.

Shen, X-N. S. 1990. "Tonal coarticulation in Mandarin." *Journal of Phonetics*, 18:281-295.