Thai Tonal Contrast under Changes in Speech Rate and Stress

Rattima Nitisaroj

Department of Linguistics
Georgetown University, Washington D.C.
rn29@georgetown.edu

Abstract

This study investigates how the five lexical tones in Thai are realized on primary-, secondary-, and unstressed syllables produced at fast, normal and slow rate. The results revealed that 1) speech rate does not have any significant effect on F0 height, excursion size and F0 peak and valley locations of Thai tones, 2) tones on primary-stressed syllables have a larger excursion size than those on secondary- and unstressed syllables, and 3) the five-way tonal contrast in the language is maintained regardless of changes in speech rate and stress.

1. Introduction

Thai has a five-way tonal contrast: mid, high, low, falling, and rising. It is agreed that on stressed syllables, the five-way contrast is maintained [1, 5, 10, 14]. Whether or not this is the case for unstressed syllables is controversial. Different studies show different results regarding tone neutralization in unstressed syllables. Studies such as [7] and [13] observe tone neutralization while others [4, 15, 16] do not. The discrepancy may be due to different speech rates used in each study. Cross-linguistic research investigating rate effects on F0 has concentrated on F0 level and F0 peak location. Higher F0 values are detected in faster speech [3, 11, 19]. However, peak locations do not move as a result of rate change [8, 12, 20].

At present, there has been only one earlier study on Thai that compares tone realizations in stressed and unstressed syllables at different speech rates [6]. However, the experiment did not control very well for the position of the target syllables, which renders the results problematic. Stressed syllables were put in a phrase-final position while unstressed syllables stayed in a non-final position. The smaller speech rate effects on stressed syllables may turn out to be related to the phrase-final position instead of stress degree.

This study aims to compare phonetic realizations of the five tones under various stress conditions at different speech rates. Different from [6], the target syllable position will be controlled for. Also, the number of stress and rate levels will each be increased from two to three. Tonal height, slope, and alignment pattern will be examined with the following hypotheses:

- At fast rate, the tones will be realized with a higher F0 value and a smaller excursion size. At slow rate, F0 values of the tones will be lower and the contours will have a larger excursion size.
- Despite phonetic variation between rates, the five-way tonal contrast will be maintained on primary- and secondary-stressed syllables at every rate. On unstressed syllables, the contrast will be reduced to a three-way contrast.

The next section provides details of the experiment to test the hypotheses.

2. Experiment

2.1. Materials

In writing, a sequence of words may be ambiguous between a syntactic phrase and a compound. To illustrate, ม้านั่ง /máː nâŋ/, which consists of a noun meaning 'horse' followed by a verb meaning 'sit', can represent either a syntactic phrase meaning 'the horse sat' or a compound referring to 'a bench'. The 'phrase' meaning can be achieved by putting primary stress on both syllables. All else being equal, when the first syllable receives a secondary stress, the sequence will be recognized as a compound. In the experiment, such phrase-compound minimal pairs differing in stress pattern (primary-primary vs. secondary-primary) were chosen as tokens to include in the stimuli set in order to obtain primary-stressed and secondary-stressed syllables for further analysis. As for unstressed syllables, they were represented in the stimuli set by the first syllable of disyllabic words such as /nǎŋ sɯ̌ː/ 'book' (unstressed-primary stress pattern). Perfect minimal triplets cannot be made, especially for non-compound disyllabic words, which are rare. The target syllables, i.e. the first syllables of the phrases, compounds, and disyllabic words, are CVː, CVN, and CVːN syllables carrying each of the five tones. To minimize carry-over and declination effects, all 45 tokens (15 phrases + 15 compounds + 15 disyllabic words) were embedded in the initial position of appropriate sentence frames. These sentences were then preceded by one or two sentences of disambiguating context.

2.2. Subjects

Nine female speakers of Thai participated in the study. All subjects were born and raised in Bangkok, Thailand until finishing at least the Bachelor's degree. Three of the subjects went abroad for a couple of years to earn their Master's. At the time of recording, all of the subjects worked in Bangkok. The average age of the subjects is 29.1 years old (SD = 2.67).

2.3. Recording Procedure

The target sentences preceded by a few sentences providing a context requiring the intended reading were typed in Thai script on index cards. The subjects were asked to read the cards at three speech rates. The "normal" rate means a relaxed and comfortable tempo. The "slow" rate refers to the slowest tempo that the subjects could produce without inserting a pause between words within the same sentence. The "fast" rate is defined as the fastest tempo possible without making errors or losing naturalness. Before the recording began, the subjects were allowed to familiarize themselves with the sentences. During the recording session, the subjects were first asked to read all the sentences at the normal rate. Then, they read at the slow and the fast rate. Seven filler sentences were also added between the slow and fast rates. The subjects were asked to read these filler sentences at the normal rate. These intervening filler sentences were put in to remind the subjects what the "normal" rate was and helped them to produce the "fast" rate as actually faster than the "normal" one. The materials were recorded directly onto a PC at a 22.05 kHz sampling rate using the Praat program [2]. It took each subject approximately one hour to finish the task

2.4. Measurements

F0 was extracted via an autocorrelation algorithm in the Praat program, using a 30 ms Hanning window with a step size of 5 ms. To be able to compare F0 configurations across syllable and tone types, F0 data were time-normalized by taking measurements at 11 locations throughout the rhyme of each target syllable. Measurements started from the beginning of the rhyme (0%) and moved every 10% of the rhyme duration until the end.

2.5. Data Normalization

Pitch range differences among the subjects were normalized by transforming Hertz values to a z-score scale [9, 18] using the following formula:

$$F0_{norm} = (F0_i - \bar{F}0)/s \tag{1}$$

where s is one standard deviation about the mean F0 ($\bar{F}0$)

Normalization parameters, i.e. mean and standard deviation, were calculated from raw F0 values of all tokens for a subject. The normalized F0 values thus express individual F0 values as the number of standard deviations above or below a subject's mean.

3. Results

The rate manipulation was successful as confirmed by the mean rhyme durations for the different rates: 126, 163, and 205 msec for fast, normal, and slow rate respectively. Then all the five tones on primary-stressed, secondary-stressed, and unstressed syllables uttered at fast, normal, and slow rate are examined and compared with respect to the four properties of F0 height, excursion size, highest F0 location, and lowest F0 location. A series of three-way repeated measure ANOVAs (rate x stress x tone) were conducted to evaluate the extent to which these four properties of all five tones varied as functions of speaking rate and stress. Results are reported below.

3.1. F0 Height

F0 height defines the value averaged over the 11 measurement points of a syllable. A three-way repeated measure ANOVA on F0 heights indicated a main effect of tone, $F(4, 32) = 414.38$, $p < .0001$, and a significant two-way interaction between stress and tone, $F(8, 64) = 12.39$, $p < .0001$. Neither rate nor stress produced significant effects. Also, all interactions between rate and other factors were not significant. Post-hoc Student Newman-Keuls (SNK) multiple paired comparisons ($\alpha = .05$) revealed significant differences between all tones. By F0 heights, Rising < Low <Mid < High < Falling. Under each stress condition, differences between all tones except Low and Rising on unstressed syllables were significant.

3.2. Excursion Size

Excursion size was calculated by taking the difference between the highest and the lowest F0 for each syllable. In a three-way (rate x stress x tone) repeated measure ANOVA, stress and tone exhibited significant effects on tonal excursion sizes, $F(2, 16) = 11.04$, $p < .05$ for stress, and $F(4, 32) = 49.19$, $p < .0001$ for tone. Besides, stress was found to interact with tone to affect excursion sizes, $F(8, 64) = 2.69$, $p < .05$. Post-hoc SNK tests ($\alpha = .05$) revealed that primary-stressed syllables had a significantly larger excursion size than secondary-stressed and unstressed syllables. Low had the largest excursion size, followed by Rising and High. Falling had a smaller excursion size than High. The smallest excursion size belonged to Mid. Post hoc SNK tests ($\alpha = .05$) indicated significant differences between all tones except between Rising and High. Under each stress condition, the five-way contrast was not maintained. On primary-stressed syllables, Rising and High could not be distinguished on the basis of excursion size. Within the secondary-stressed category, Low constituted the only tone whose excursion size was significantly different from other tones. For the rest, no significant difference was found between Mid and Falling, Falling and Rising, and Rising and High. For the case of unstressed syllables, difference between Falling and High failed to reach a statistically significant level.

3.3. Highest F0 Location

The three-way repeated measure ANOVA conducted on highest F0 locations demonstrated main effects of stress, $F(2, 16) = 6.38$, $p < .05$, and tone, $F(4, 32) = 71.76$, $p < .0001$. Also, rate interacted with tone to affect highest F0 locations, $F(8, 64) = 3.24$, $p < .05$. Highest F0 values were found later in unstressed syllables (43.41% of rhyme duration) than in primary- (37.24%) and secondary-stressed syllables (36.94%). Post hoc SNK tests ($\alpha = .05$) identified no significant difference between stress categories. With respect to the main effect of tone, SNK post hoc tests ($\alpha = .05$) revealed that it came from significant differences between all tones except between High and Falling. The highest value occurred earliest in the case of Low (at 3.99% of rhyme duration). For Mid, highest F0 locations were approximately at 11.89% of the entire rhyme duration. At 24.90% of the rhyme, the highest F0 value of Rising was found. For Falling, the highest value occurred at about three quarters (76.26%) of the rhyme. High tones had their highest F0 values relatively late in the syllable (at 78.93% of the rhyme duration). Within each of the three rates, the difference between Low and Mid did not reach a significant level. Neither did that between Falling and High. At the fast rate, Mid was also not significantly different from Rising.

3.4. Lowest F0 Location

Concerning lowest F0 locations, a three-way ANOVA with rate, stress and tone as independent variables displayed significant effects of stress, $F(2, 16) = 11.40$, $p < .05$, and tone, $F(4, 32) = 110.58$, $p < .0001$. There was also a significant two-way interaction between rate and tone, $F(8, 64) = 3.67$, $p < .05$. Unstressed syllables had their lowest F0 values at the earliest location (52% of rhyme duration). The next location (55.41%) belonged to secondary-stressed syllables. The lowest values of primary-stressed syllables took place latest in the syllables (at 57.51% of rhyme duration). Post hoc SNK tests (α = .05) indicated no significant difference between stress conditions. Regarding the main effect of tone, it came from significant difference between all tones, as indicated by SNK post hoc tests (α = .05). Lowest F0 values could be found earliest in the syllable for Falling (17.90% of rhyme duration). The next earliest location belonged to High (28.72%). For Rising, lowest F0 values could be found around the middle part of the rhyme (59.26%). F0 fell to the lowest point at 77.74% of rhyme duration for Mid. Latest locations for lowest F0 values took place around 77.74% of rhyme duration and it was approximately where low tonal contours reached their lowest values. Within each rate, all tones were significantly different from one another in terms of lowest F0 location. The only pair whose difference did not reach a statistically significant level was Falling and High on tokens spoken with the slow speed.

The following table summarizes significant effects of rate, stress, and tone, as well as their interactions on F0 height, excursion size, highest F0 location, and lowest F0 location.

Table 1: *Significant differences (indicated by *) from 3 x 3 x 5 (rate x stress x tone) repeated measure ANOVAs*

	F0 Height	Excursion Size	Highest F0 Loc	Lowest F0 Loc
Rate				
Stress		*	*	*
Tone	*	*	*	*
Rate x Stress				
Stress x Tone	*	*		
Rate x Tone			*	*
Rate x Stress x Tone				

4. Discussion and Conclusions

This study has investigated four acoustic properties of Thai tones under changes in speech rate and stress. Results in the previous section showed that speech rate did not produce any across-the-board effects on Thai tones. As seen in Fig. 1, contours for each tone are almost identical across rate. With respect to F0 height, the results do not provide support to previous findings in other languages where F0 level raises as speech rate increases [3, 11, 19]. However, consistent with [8, 12, 20], in Thai F0 peak does not move as a result of rate change.

As for stress, the ANOVA results and Fig. 2 show that it significantly affected tonal realizations. In particular, tones on primary-stressed syllables have a larger excursion size than their counterpart on secondary- and unstressed syllables.

Regarding tonal contrast under changes in speech rate and stress, as seen in Fig. 1 and Fig. 2, contrary to what was hypothesized, there does not seem to be any case where the five-way contrast is reduced to three-way. However, it appears that no single acoustic property investigated here can be used to distinguish all the five tones. To achieve the five-way tonal distinction, one may try using a combination of the four acoustic properties. Mid tones are marked by their small excursion size. Low and rising tones can be distinguished from the other three tones by their low average F0 values. Fig. 1 and Fig. 2 show that Low reaches the lowest F0 value around the end of the syllable while the lowest F0 location is around the middle part in the case of Rising. The two tones differ from each other with respect to the lowest F0 location. Unlike others, high and falling tonal contours reach the highest F0 value around the end of the syllable. Between them, falling tones are higher in F0 value.

For contour tones, [17] shows that compared to their long counterparts, shorter rising tones still display a rising portion at the end while shorter falling tones do not have a falling portion. In the present study, as seen in Fig. 1 and Fig. 2, rising tones falls to the lowest F0 value around 60% of the rhyme duration where the contours start rising and falling tones fall slightly at the very end. Compared to [17], rising tones here appear to have a larger rising portion and falling tones do not lack a falling portion. The source of the discrepancies seems to be the different tokens used in the experiments. In [17], F0 was extracted from the vowel portions of syllables ending in a voiceless unaspirated stop /p/ while the current study obtained F0 values from the rhyme portion of CV:, CVN, and CV:N syllables.

In sum, this investigation of Thai tonal contours under

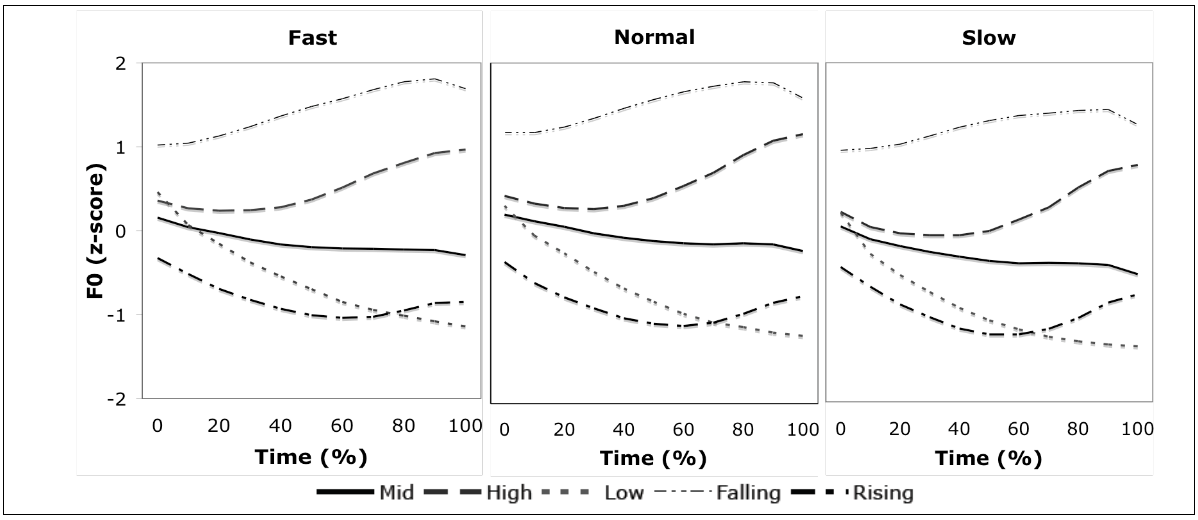

Figure 1: *Average F0-time-normalized contours of all five tones at fast, normal, and slow rate*

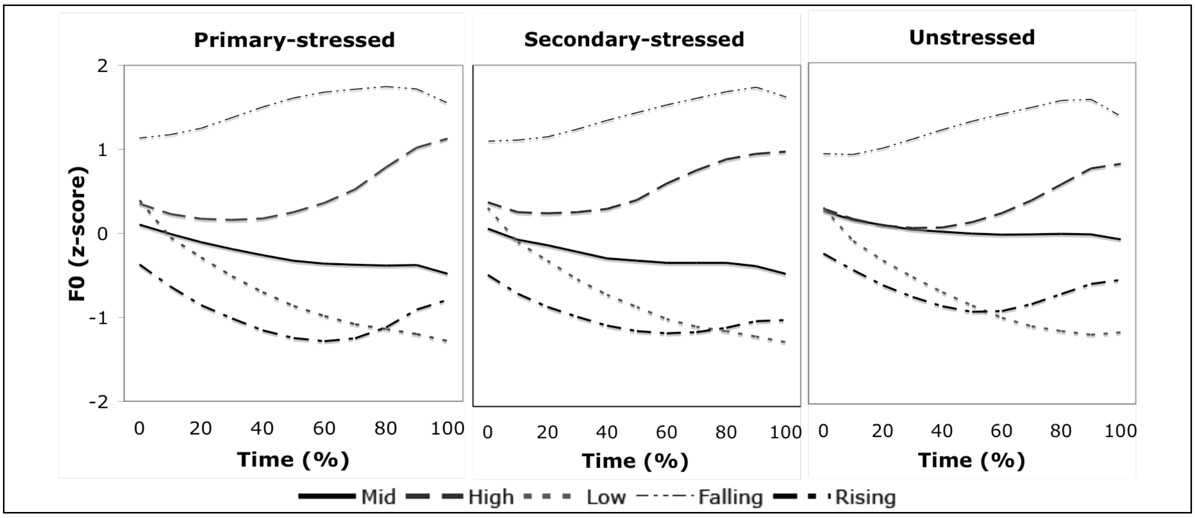

Figure 2: *Average F0-time-normalized contours of all five tones on primary-, secondary-, and unstressed syllables*

changes in speech rate and stress shows that speech rate did not have any significant effects on the height and shape of Thai tones while stress played a role in determining the excursion size of tonal contours. Next, the five-way contrast is always maintained. The five tones can be distinguished by using the properties of F0 height, excursion size, and F0 peak or valley location.

5. Acknowledgement

I wish to thank Dr. Elizabeth Zsiga for her valuable advice and comments.

6. References

[1] Abramson, A.S., 1979. The coarticulation of tones: an acoustic study of Thai. In *Studies in Tai and Mon-Khmer Phonetics and Phonology in Honor of Eugénie J.A. Handerson,* T.L. Thongkum, V. Panupong, P. Kullavanijaya, M.R.K. Tingsabadh (eds.), Bangkok: Chulalongkorn UP, 1-9.

[2] Boersma, P., Weenink, D., 2005. Praat: Doing phonetics by computer (Version 4.3.01) [Computer program]. Retrieved from http://www.praat.org/.

[3] Caspers, J., van Heuven, V.J., 1993. Effects of time pressure on the phonetic realization of Dutch accent-lending pitch rise and fall. *Phonetica* 50, 161-171.

[4] Gandour, J., 1975. On the representation of tone in Siamese. In *Studies in Tai Linguistics in Honor of William J. Gedney,* J.G. Harris, J.R. Chamberlain (eds.), Bangkok: Central Institute of English Language, 170-195.

[5] Gandour, J.; Potisuk, S.; Dechongkit, S., 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22, 477-492.

[6] Gandour, J., Tumtavitikul, A., Satthamnuwong, N., 1999. Effects of speaking rate on Thai tones. *Phonetica* 56, 123-134.

[7] Hiranburana, S., 1972. Changes in the pitch contours of unaccented syllables in spoken Thai. In *Tai Phonetics and Phonology,* J.G. Harris, R.B. Noss (eds.), Bangkok: Central Institute of English Language, 23-27.

[8] Igarashi, Y., 2004. “Segmental anchoring” of F0 under changes in speech rate: Evidence from Russian. *Speech Prosody 2004,* International Conference, B. Bel, I. Marlien (eds.), 25-28.

[9] Jassem, W., 1975. Normalization of F0 curves. In *Auditory Analysis and Perception of Speech,* G. Fant, M. Tatham (eds.), London: Academic Press, 523-530.

[10] Kallayanamit, S., 2004. *Thai Intonation: Contours, Registers, and Boundary Tones.* Ph.D. diss., Georgetown University

[11] Kohler, K.J., 1983. F0 in speech timing. In *Studies in Speech Timing,* K.J. Kohler, Ch.E. Hoequist (eds.), Kiel: Institut für Phonetik der Universität Kiel, 55-97.

[12] Ladd, R.D., Faulkner, D., Faulkner, H., Schepman, A., 1999. Constant segmental anchoring of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106, 1543-1554.

[13] Luksaneeyanawin, S., 1983. *Intonation in Thai.* Ph.D. diss., University of Edinburgh.

[14] Morén, B., Zsiga, E., 2006. The lexical and post-lexical phonology of Thai tones. *Natural Language and Linguistic Theory* 24, 113-178.

[15] Potisuk, S., Gandour, J., Harper, M.P., 1994. F0 correlates of stress in Thai. *Linguistics of the Tibeto-Burman Area* 17(2), 1-27.

[16] Potisuk, S., Gandour, J., Harper, M.P., 1996. Acoustic correlates of stress in Thai. *Phonetica* 53, 200-220.

[17] Roengpitya, R., 2003. (End) truncation of Thai contour tones on different durations of TBUs. *Proceedings of the 15th International Congress of Phonetic Sciences* (ICPhS 03)*,* Barcelona, 1109-1112.

[18] Rose, P., 1987. Considerations in the normalization of the fundamental frequency of linguistic tone. *Speech Communication* 6, 343-351.

[19] Steppling, M.L., Montgomery A.A., 2002. Perception and production of rise-fall intonation in American English. *Perception and Psychophysics* 64, 451-461.

[20] Xu, Y., 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55, 179-203.