Tonal timing and vowel onset characteristics in Thai

David House and Jan-Olof Svantesson

Dept. of Linguistics and Phonetics, Lund University
Helgonabacken 12, S-223 62, Lund, Sweden.

1. Introduction

1.1 *Tonal timing in production and perception*

In many tone languages of Asia, fundamental frequency (F0) is seen as the most important correlate of tone (Gandour 1983). In the description of tones, a distinction is often made between contour tones (involving a substantial F0 movement) and level tones (involving a lesser degree of F0 movement). This distinction has usually been based on descriptions using auditory perception, although instrumental acoustic studies also corroborate these descriptions. A current issue in the description of tone involves the percept of a level tone versus the percept of a contour tone. When is tonal movement perceived as a level tone and when is it perceived as a contour tone? What are the perceptual constraints acting upon the course of F0 through the syllable?

Based on a series of perception experiments using Swedish listeners, House (1990) investigated the perceptual importance of tonal movement timing in relationship to vowel onset in simple CV syllables where the consonant was a voiced stop or a nasal. On the basis of the results and on theories of pitch perception, a model of optimal tonal perception was proposed. In the model, tonal movement early in the vowel (through areas of changing spectral and intensity characteristics) is recoded as level features, while movement later in the vowel (through areas of spectral stability) is coded as contour features. Perception studies in both Chinese and Swedish have also demonstrated the importance of tonal timing in relationship to vowel onset specifically concerning a falling tonal contour. In Chinese, a falling contour early in the vowel is perceived as tone 3 while the same contour later in the vowel is perceived as tone 4 (Gårding et al. 1986). In Swedish, a falling contour early in the vowel is perceived as accent I while the same contour later in the vowel is perceived as accent II (Bruce 1977).

In terms of production, the two categories in both languages above can be seen as comprising a similar tonal gesture differing in timing in relation to the onset of the vowel. In terms of perception, the two categories can be seen as a low tone (falling movement in an area of spectral change) and as a falling tone (falling movement in an area of spectral stability).

1.2 *Tonal timing in Thai*

As in Chinese, Thai has a tonal contrast between a tone described as low and one described as falling (Abramson 1962). The low tone is described as having a falling contour early in the vowel and ending at a low level while the falling contour is described as high during the first part of the vowel and then falling late in the vowel (Tumtavitikul 1995a).

In contrast to Chinese, however, Thai syllable onsets may consist of consonant clusters such as [kl] and [kw] with or without aspiration. This raises the question of the realization and perception of the tones and their timing in relationship to vowel onset. For example, if tonal movement is strictly timed in relation to vowel onset, the beginning of the falling component in the low tone (cf Gandour et al. 1991) could be delayed in relation to voice onset by the presence of a sonorant consonant in clusters such as [kl] and [kw]. If such a delay were long enough to cause the contour to occur in the vowel in an area of *relative* spectral stability the result could be potential confusion between the low tone and the falling tone. A similar delay in the falling tone (causing the fall to occur in an area of spectral change) could cause potential confusion between the falling tone and the high tone, although the high tone is often described as containing a final rise (Gandour et al. 1991, Tumtavitikul 1995a, 1995b). If, on the other hand, tonal movement is timed to the beginning of voicing, the falling contour could be perceived as falling too early in relationship to the onset of the actual vowel, thus causing confusion between the falling tone and the low tone.

1.3 *Goals of this study*

Perceptual aspects of these questions can be addressed by using perception tests whereby the tonal contours in minimal pairs are systematically manipulated using speech synthesis techniques. As a precursor to such tests, however, it seems advisable to investigate production aspects of tones occurring in cluster environments. This will provide more insight into the timing of tonal movement from a production standpoint. The goals of this study, therefore, are to ascertain if a consonant cluster and/or aspiration before vowel onset has an effect on the timing of a tonal falling movement in Thai. Is tonal movement synchronized with vowel onset, with syllable voicing onset, or simply variable?

2. Method

2.1 Speech material

To investigate these questions, speech material containing 17 test words representing low and falling tones, initial aspirated and unaspirated stops and initial

clusters of stops and sonorants [l] and [w] were set into a sentence context adapted from Gandour et al. (1994). Sentence frames were used in which the test words were preceded by high, falling and rising tones and followed by mid tone to control for tonal coarticulation effects. Six of the test words in three different tonal contexts were used for a quantitative analysis and are presented in Table 1. An additional reason for using a high tone before the test words was to increase the falling component of the low tone through tonal coarticulation, thus producing an optimal context for further perception studies involving minimal pairs.

Table 1. Sentence frames for the three different tonal contexts and the six test words used for the statistical analysis. Each test word was used in all three contexts.

Sentence frames:

เธอซื้อ___มา	thəə sɨ́ɨ ___ maa.	'You bought ___.'
เธอได้___มา	thəə dâj ___ maa.	'You got ___.'
เธอขาย___มา	thəə khăaj ___ maa.	'You sold ___.'

Test words:

แก้ว	kɛ̂ɛw	'glass'
ไก่	kàj	'chicken'
ข้าว	khâaw	'rice'
ไข่	khàj	'egg'
กล้อง	klɔ̂ŋ	'pipe'
กล่อง	klɔ̀ŋ	'box'

2.2 Subjects

Three adult speakers of Standard Thai participated as subjects in this study. Subject 1 (female) was originally from southern Thailand, while Subjects 2 and 3 (male) were both from Bangkok. All three subjects are now living in southern Sweden.

2.3 Recording procedure

The sentences were printed in Thai script in random order on two sheets of paper. To avoid beginning and end effects, extra sentences similar to the speech material were placed before and after the first and last sentence of each page. The subjects were instructed to read the sentences at a normal conversational speech

rate. Each subject was instructed to read the set of sentences three times with a short break between each reading. The sentences were recorded in a sound studio on a Panasonic SV-3700 digital audio tape deck at the Department of Linguistics and Phonetics, Lund University.

2.4 Analysis and measurement procedures

The recordings were analyzed acoustically using the ESPS-Waves+ environment on a Sun workstation. The audio waveform, a wide-band spectrogram and a fundamental frequency curve were displayed for each utterance. The beginning and end of voicing for each test word in each utterance was labelled manually. An algorithm written by Marcus Filipsson was then applied to each test word in each utterance which first divided the word into six sections of equal duration. Depending on the duration of the word, section durations varied from about 50 to 70 ms. F0 mean was then calculated for each section and read into a separate text file.

The text files for six of the test words (Table 1) were then fed into a statistics program, Statistica, in which the mean F0 for each section was calculated over the three repetitions of each word in each tonal context for each speaker. The mean values were plotted in graphs to facilitate a comparison of the tonal contours for the two tones in the three different tonal contexts. Finally mean F0 values for all repetitions in all contexts for each section in each test word were plotted to ascertain the influence of the initial consonant clusters and aspirated stops on the timing of the falling contour for both the low and falling tone.

3. Results

3.1 Visual analysis

The visual analysis of the acoustic displays for all 17 test words in all tonal contexts for all three speakers showed no evidence of a systematic influence of consonant clusters or aspiration on the timing of tonal movement. Generally, the falling movement for the low tone was initiated at the onset of voicing while the falling tone maintained a relatively high tonal level well into the vowel falling more rapidly near the end of the syllable. The actual onset of the falling contour for the falling tone was, however, not always easy to establish in the individual utterances. Considerable variation was apparent ranging from a falling contour throughout the syllable in some cases to a clear onset of the falling contour in the final portion of the syllable in other cases.

Waveforms and F0 tracings of the minimal pair [klôŋ] and [klòŋ] in the context following a high tone are shown in Figure 1. In this example, the falling components of both tones are apparent as is the late onset of the fall near the end of the vowel for the falling contour.

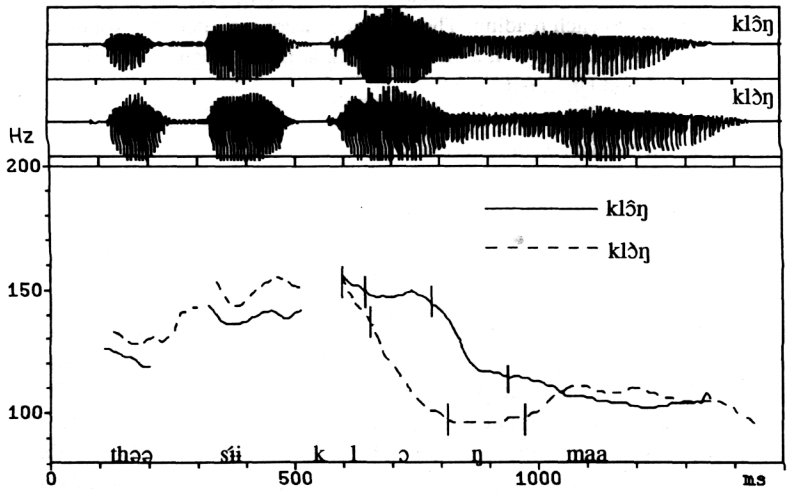

Figure 1. Waveforms and F0 tracings of the minimal pair [klɔ̂ŋ] (solid line) and [klɔ̀ŋ] (dashed line) in the context following a high tone. The vertical lines through the F0 contour indicate segment boundaries for the test words.

3.2 Quantitative analysis

A comparison of the mean F0 values for each analyzed segment for each of the six words in different tonal contexts showed no influence of aspiration or consonant cluster on the timing of the tonal fall. For both low tone and falling tone in the test words a clear influence of carry-over coarticulation is seen with the tonal onset of the test word generally being highest after a high tone and lowest after a falling tone (cf. Gandour et al. 1994). Mean F0 values for each word pair for speaker 3 are plotted in Figure 2 resulting in a time normalized average tonal contour for each word in each tonal context. The effects of carry-over coarticulation described above can be seen in the figure.

Mean F0 values over all tonal contexts are plotted for each word and each speaker in Figure 3. The falling movement for the low tone is confined to the first third of the voiced portion of the syllable while for the falling tone, the tonal movement has its steepest slope in the final part of the syllable. Here again, no influence of vowel onset characteristics can be seen on the timing of the tonal contour. For speaker 1, the somewhat variable nature of F0 values for the low tone can be explained by instances of creaky voice in her productions of the tone.

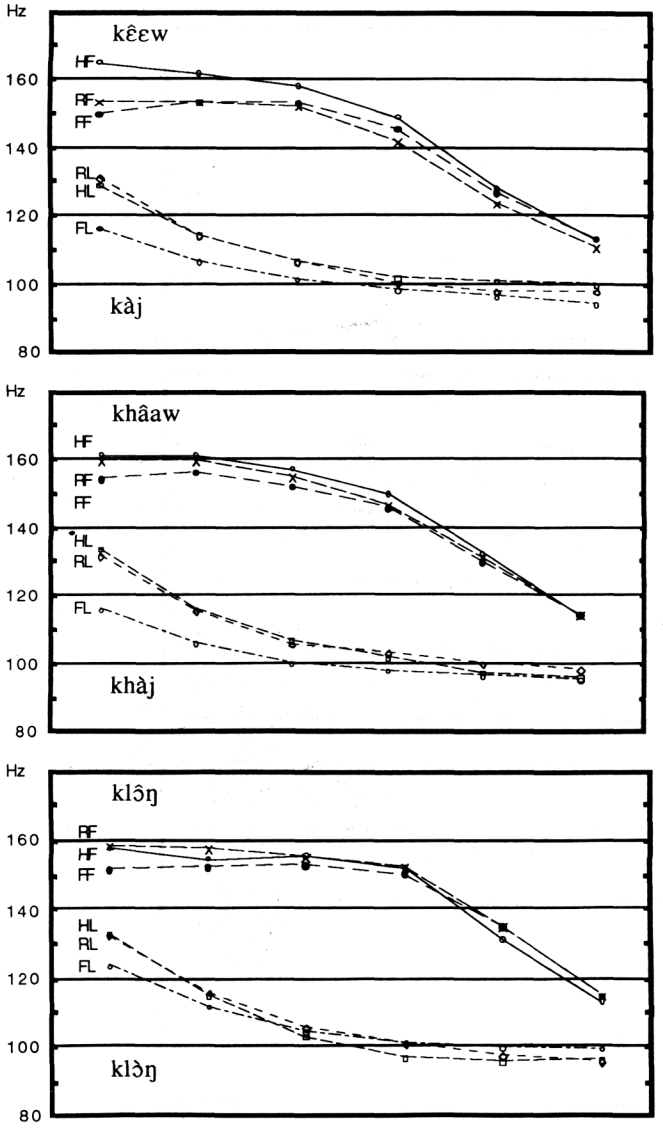

Figure 2. Mean F0 values for each word pair in each tonal context for speaker 3. Each point represents the average of three registrations of the mean frequency over one sixth of the voiced section of the test word.

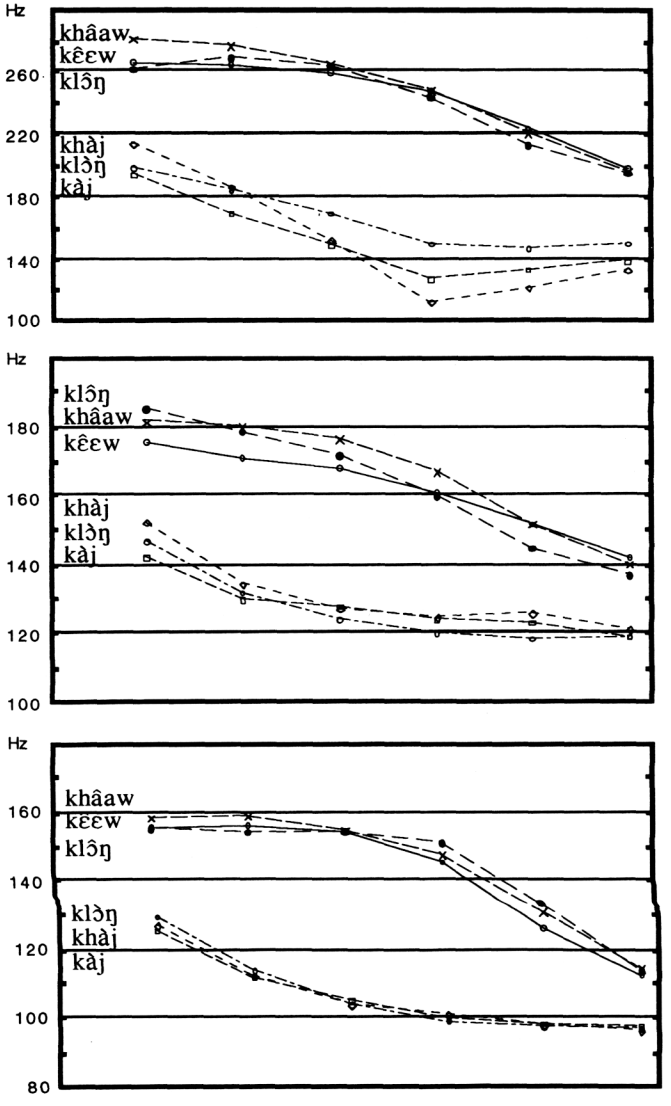

Figure 3. Mean F0 values for each word in all tonal contexts for speaker 1 (top), 2 (middle) and 3 (bottom).

4. Discussion

4.1 Onset of voicing as timing coordinator

The results of this study seem to indicate that for the low tone and the falling tone in Thai a consonant cluster and/or aspiration before vowel onset do not have an effect on the timing of the tonal contour. The timing of the tonal gesture in syllables beginning with a voiceless stop seems to be related more to the onset of voicing in the syllable than to the actual vowel onset.

Although the test material in this study does not include syllables beginning with a voiced sonorant consonant, we can speculate that in those instances the synchronization point would be the vowel onset since sonorants do not form consonant clusters in Thai. This would point toward a description where the area of maximum intensity and maximum spectral change in the syllable serves as a synchronization point for the tonal gesture. In syllables beginning with a consonant cluster having an initial stop, the area of maximum intensity and maximum spectral change coincides with voice onset, while in syllables beginning with a voiced sonorant consonant this area coincides with vowel onset.

4.2 Syllable structure

The above description of tonal gesture timing could have implications for an analysis of syllable structure in tone languages. If the area of maximum intensity and spectral change is crucial for tonal timing, the onset of a syllable may then be defined as the portion of the syllable occurring prior to that point. Thus the voiced sonorant consonants of the initial clusters can be said to belong to the syllable nucleus and function as tone-bearing segments. The fact that the falling contour for the falling tone occurs late in the syllable, often in the final sonorant consonant, points to the coda as an important tone-bearing portion of the syllable.

4.3 Perceptual implications

The onset of voicing as a tonal timing coordinator for the early falling contour for the low tone is consistent with the tonal perception model presented in House (1990) where a falling contour through an area of spectral change will be coded as a low tone. The production data for the falling tone is, however, not consistent with the model (see also Tumtavitikul 1995a, 1995b). For a falling tone to be optimally perceived as a fall, in terms of the model, the falling contour should occur in an area of maximum spectral stability. In our data, the falling contour often occurs through areas of spectral change at the beginning of the syllable coda. A considerable part of the fall does, however, occur in the spectrally stable portion of the coda. This gives rise to speculation that syllable structure, especially in terms

of *maximum* intensity and spectral change, may function as an important perceptual element in the perception of tone. In this sense, an increase in intensity together with spectral change (at the beginning of the syllable nucleus) could play a different perceptual role than a decrease in intensity together with spectral change (at the beginning of the syllable coda). Thus a falling contour at the beginning of the nucleus would be coded as low while a falling contour at the beginning of the coda would be coded as falling. Perceptual testing is needed to confirm this. An important task in this regard would be to investigate the perceptual crossover point using synthetic stimuli with Thai word pairs such as [klôŋ] and [klòŋ] where the tonal contour is successively manipulated.

4.4 Description of Thai tones

We can speculate that for the low and falling tones in Thai, the same falling gesture is produced with a timing difference to maintain the contrast. This appears to have certain similarities with the low versus falling tone contrast in Chinese and the word accent contrast in Swedish (Gårding et al. 1986, Bruce 1977). Thai is different, however, in that the timing difference is maximized in the syllable leading to a much later falling contour for the falling tone than in either Chinese or Swedish. The initial falling element of the low tone may help to contrast this tone with the mid tone which in our material lacks an initial falling element. If this is the case, it may be more important in Thai to maximize the timing difference between the early fall and the late fall for the low and falling tones than it is in either Chinese or Swedish. Here, again, perceptual studies are needed to confirm this.

5. Acknowledgement

We would like to thank Damrong Tayanin for arranging the recording sessions with our subjects and also for advice concerning the test material.

6. References

Abramson, Arthur S. 1962 *The vowels and tones of standard Thai: acoustical measurements and experiments.* Bloomington, IN: Indiana University Research Center in Anthropology, Folklore, and Linguistics (Publication Twenty). Also Part II, *International Journal of American Linguistics* Vol. 28:2.

Bruce, Gösta. 1977 *Swedish word accent in sentence perspective.* Lund: Gleerup.

Gandour, Jack. 1983 Tone perception in Far Eastern languages. *Journal of Phonetics* 11:149-75.

Gandour, Jack, Siripong Potisuk and Sumalee Dechongkit. 1994 Tonal coarticulation in Thai. *Journal of Phonetics* 22: 477-492.

Gandour, Jack, Siripong Potisuk, Suvit Ponglorpisit and Sumalee Dechongkit.

1991 Inter- and intraspeaker variability in fundamental frequency of Thai tones. *Speech Communication.* 10: 355-372.

Gårding, Eva, Paul Kratochvíl, Jan-Olof Svantesson and Zhang Jialu. 1986 Tone 4 and Tone 3 discrimination in Modern Standard Chinese. *Language and Speech* 29:281-93.

House, David. 1990 *Tonal Perception in Speech.* Lund: Lund University Press.

Tumtavitikul, Appiluck. 1995a Tonal movements in Thai. *Proc. ICPhS 95.* Vol. 1:118-121. Stockholm.

Tumtavitikul, Appiluck. 1995b Toward a model of tonal recognition in Thai. *Proc. 2nd Symposium on Natural Language Processing*: 387-395. Bangkok.