Tonal timing and vowel onset characteristics in Thai

David House and Jan-Olof Svantesson

Dept. of Linguistics and Phonetics, Lund University

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 (Tuntavitikul 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, Tuntavitikul 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:

ก้าว__มา theo sii __ maa. ‘You bought ___’

ก้าว__มา theo daj __ maa. ‘You got ___’

ก้าว__มา theo khàaj __ maa. ‘You sold ___’

Test words:

ก้าว kēew ‘glass’

ก้า kàj ‘chicken’

ก้าว khåaw ‘rice’

ก้า kàj ‘egg’

ก้าว kløŋ ‘pipe’

ก้าว kløŋ ‘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 [klɔŋ] and [klɔŋ] 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.