The Thai Tonal Space

Arthur S. Abramson

The University of Connecticut and Haskins Laboratories

INTRODUCTION

In the analysis of a tone language (Gandour, 1994), the linguist normally thinks first of pitch levels and glides as the probable phonetic basis of phonologically relevant tones, even though there may be other factors, apparently secondary in importance, that go along with pitch. Of course, it is well known that in some languages something other than pitch, perhaps a voice quality, may be dominant in one or more of the tones.

Against the background of earlier auditory (Haas & Subhanka, 1945) and instrumental (Bradley, 1911) analysis, Abramson (1962) was apparently the first to combine techniques of acoustic analysis and speech synthesis to investigate the tones of Standard Thai (Siamese)—or, indeed, any tone language—both acoustically and perceptually. Since then, of course, other such treatments of Asian languages, including Thai, have appeared (e.g., Erickson, 1974; Gandour, 1978; Zee & Maddieson, 1980).

The present study is part of an ongoing but intermittent exploration (e.g., Abramson, 1975, 1976) of the Thai tonal “space.” This space is taken to be the set of articulatory and auditory dimensions by which the speaker of Thai is constrained in production and perception. The paper makes use of unpublished or reanalyzed data obtained in Thailand from time to time at the former Central Institute of English Language of Mahidol University, the Faculty of Humanities, Ramkhamhaeng University, and the Faculty of Arts of Chulalongkorn University. It has three broad goals: to revalidate earlier work on “ideal” contours for the tones on isolated monosyllables, to gain some insight into the latitudes of shifting levels and glides for the intelligibility of the tones, and to take another look (cf. Abramson, 1978) at the typological usefulness of the distinction between static and dynamic tones.

The identifiability of isolated natural Thai tones had been demonstrated (Abramson, 1962) and was reaffirmed with much more extensive testing later (Abramson, 1975). These findings were a necessary precursor to the five experiments with synthetic tones presented in this report. Aside from the baseline data for all five tones obtained in Experiment 1, the report gives no serious attention to the falling tone, which will have to be treated in another study.

PROCEDURE AND RESULTS

Experiment 1

The major physical correlate of the psychological feature pitch is fundamental frequency (F0), which for speech varies with the vibration rate of the larynx. The speech synthesizer (Intonator) used at first by Abramson (1962) has long since gone out of use. For this experiment, and the rest, the Haskins Laboratories computer-
controlled formant synthesizer was used. The syllable specified segmentally as [kʰːɐː] was chosen as the carrier for the five tones of Standard Thai, yielding five tonally differentiated words: /kʰːɐː/ ‘to be stuck,’ /kʰːɐː/ ‘galangal,’ /kʰːʔ/ ‘to deal,’ /kʰːʔ/ ‘worth,’ /kʰːʔ/ ‘leg.’ Each synthetic syllable was 450 ms long. The frequencies and amplitudes of three steady-state formants, simulating resonances of an adult male vocal tract, were made appropriate for a vowel of the type [ɐː] with formant transitions that yielded the percept of an initial dorso-velar stop. The timing of the source functions was set to produce a voiceless aspirated stop. This was done by turning on a turbulent source for the first 80 ms of the pattern (Lisker & Abramson, 1970), followed by a periodic buzz source to simulate glottal pulsing for the remaining 370 ms; the latter served as the carrier for the F0 contours. A slight upward tapering of the overall amplitude at the beginning and a slight downward one at the end made for greater naturalness. The foregoing factors were kept constant not only for Experiment 1 but for the following four experiments as well.

Figure 1. F0 contours for the Thai tones of an adult male on long vowels resynthesized from Abramson (1962, Figure 3.6).

For Experiment 1 the five F0 contours (Figure 1) found in Abramson (1962) to be ideal for the synthesis of the tones on isolated words for the voice of a typical adult male were replicated as closely as possible with the newer synthesizer and imposed on tokens of the carrier syllable. There were played through headphones in a number of random orders, over the period of a month, to 37 native speakers of Standard Thai, who wrote their responses as words in Thai script. The results, given in Figure 2, reveal rather robust identification functions. In the box at the right are shown the bar codes for the labels used by the subjects. The tone names in quotation marks stand for the tones intended in the synthesis. The two least satisfactory percepts are the mid and low tones, although both contours do achieve 88% identification. The falling, high,

---

1This synthesizer, too, is no longer in use. Our present formant synthesizer is entirely in software.
and rising tones are at least 10% higher. All three of them, including the allegedly static high tone, involve much F0 movement.

Figure 2. Experiment 1. Identification of the contours of Figure 1 by 37 subjects.

Figure 3. Sixteen F0 contours moving from 106 Hz to endpoints ranging from 90 to 152 Hz.
Experiment 2

In this experiment and in the remaining three, simple straight-line contours were used for a partial exploration of the tonal space. The 16 contours prepared for Experiment 2 are shown in Figure 3. These variants all start at 106 Hz, just below the top of the lower third of the voice range being used, and go to endpoints ranging from 90 to 152 Hz in 4-Hz steps. (An accidental exception is a 6-Hz step from 106 to 112 Hz.) These stimuli were played in several random orders to 38 native speakers for identification as Thai words.

Four hypotheses were put forth: (1) There will be no mid-tone responses, because the beginning portion of the fan-like array is too low in the voice range. (2) The falls at the lower part of the array are too low and slow for falling-tone responses. (3) The upper variants rise too slowly for the rising tone. (4) The labels used for the set by the subjects should be consequently mainly “low” and “high.”

![Diagram showing percent identification of contours across different F0 endpoints.]

Figure 4. Experiment 2. Identification of the contours of Figure 3 by 38 subjects.

The responses to the stimuli are given in Figure 4. The first hypothesis is fairly well confirmed in that the mid tone has a peak, just for the level variant at 106 Hz, of only 39%. The second hypothesis is confirmed; the word with the falling tone is not used as a label at all. The third hypothesis is not well supported since the highest variant is labeled “rising” 64% of the time; however, this peak, with only two variants above 50% is not very robust compared with the high tone, which has seven variants above 50%, and the low tone, which has five variants above 50% and a peak at 90%.

Experiment 5

Figure 5 shows the stimuli for this experiment. They are 17 F0 contours on tokens of the [kʰɔː] syllable. They all start at 90 Hz and go to endpoints ranging from 90 to 152 Hz in 4-Hz steps. (The exception is the first step, which is from 90 to 92 Hz.) The