THE THAI TONAL SPACE*

Arthur S. Abramson†

In the analysis of a tone language, the linguist normally thinks first of pitch levels and glides as the probable phonetic basis of phonologically relevant tones. This is true even though there may be other features, apparently secondary in importance, that go along with pitch. Of course, it is well known that in some languages, as in certain dialects of Vietnamese, a feature other than pitch may be dominant in one or more of the tones.

Against the background of earlier auditory (e.g., 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 Central 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., Gandour, 1978).

The present study is part of an ongoing 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 is constrained in production and perception. The paper makes use of unpublished or reanalyzed data obtained in Thailand from time to time at the old Central Institute of English Language at Mahidol University, the Faculty of Humanities of 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 in Abramson (1962) and was reaffirmed with much more extensive testing in 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

*To appear in the Proceedings of the 18th International Conference on Sino-Tibetan Languages and Linguistics, Bangkok, August 27-29, 1985.
†Also University of Connecticut
Acknowledgment. This work was supported NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. The American Council of Learned Societies and the University of Connecticut Research Foundation enabled the author to present an oral version in Bangkok at the 18th International Conference on Sino-Tibetan Languages and Linguistics.

[HASKINS LABORATORIES: Status Report on Speech Research SR-85 (1986)]

serious attention to the falling tone, which will have to be treated in another paper.

Experiment 1

The major physical correlate of the psychological feature pitch is **fundamental frequency** (F_0), which, for speech, varies with the vibration rate of the larynx. The speech synthesizer used in 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 [k^ha:] was chosen as the carrier for the five tones of Central Thai, yielding five tonally differentiated words. Each synthetic syllable was made 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 [a:], with formant transitions that yielded the percept of an initial dorso-velar stop. 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 F_0 contours. A slight upward tapering of the overall amplitude at the beginning and a slight downward one at the end made for greater naturalness.

For Experiment 1, the five F_0 contours (Figure 1) found in Abramson (1962) to be ideal for the synthesis of the tones were replicated as closely as possible with the newer synthesizer and imposed on tokens of the carrier syllable. These were played in a number of random orders, over the period of a month, to 37 native speakers of Central Thai, who wrote their responses as words in Thai script. The results, given in Figure 2, reveal rather robust identification functions. The two least satisfactory percepts are the mid and low tones, although both contours do achieve 88% identification. The falling, high, and rising tones are at least 10% higher. All three of them, including the allegedly static high tone, involve much F_0 movement.

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, the top of the lower third of the voice range, 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.)

Four hypotheses were put forth: (1) The beginning portion of this fanlike array is too low in the voice range for mid-tone responses. (2) The falls at the lower part of the array are too low and slow for the falling tone. (3) The upper variants rise too slowly for the rising tone. (4) The labels used for the set by the subjects should be mainly "low" and "high."

The responses to the stimuli are given in Figure 4. The first hypothesis is weakly 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

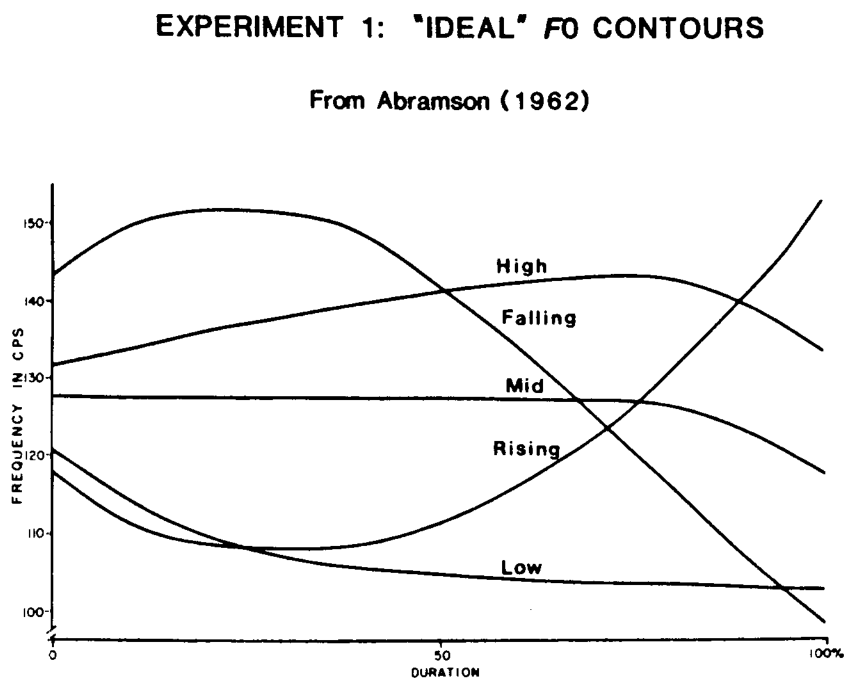

Figure 1. F_0 contours for the Thai tones of an adult male on long vowels resynthesized from Abramson (1962: Figure 3.6).

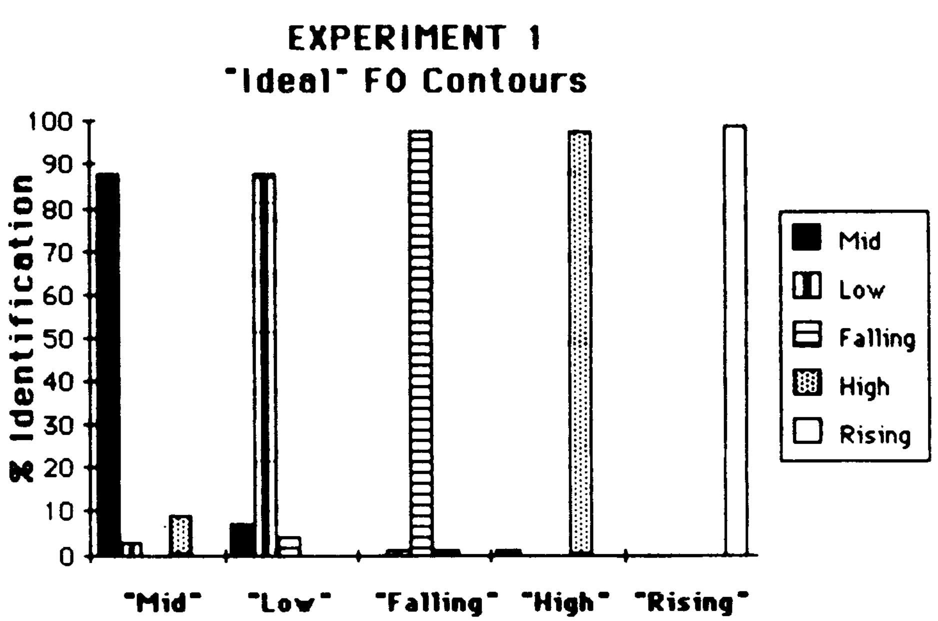

Figure 2. Experiment 1: Identification of the contours of Figure 1 by 37 subjects.

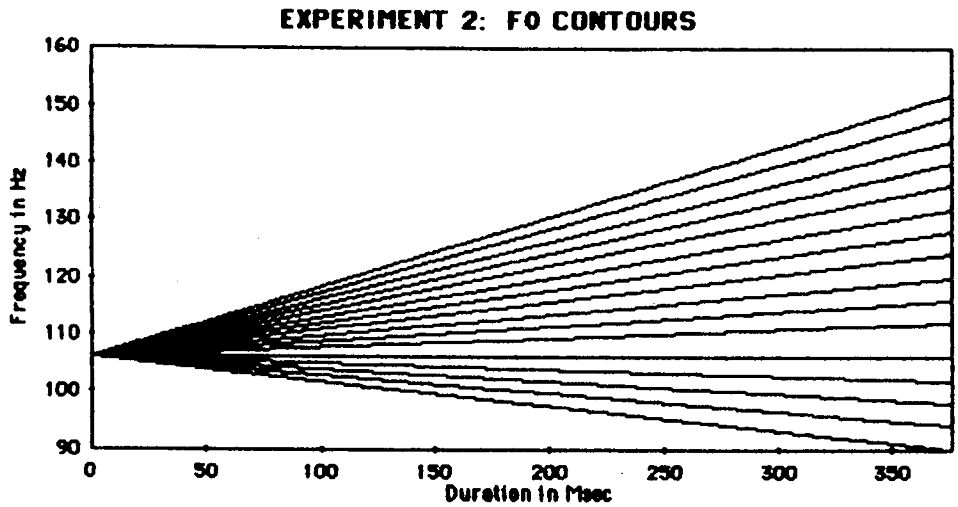

Figure 3. Sixteen F_0 contours moving from 106 Hz to endpoints ranging from 90 to 152 Hz.

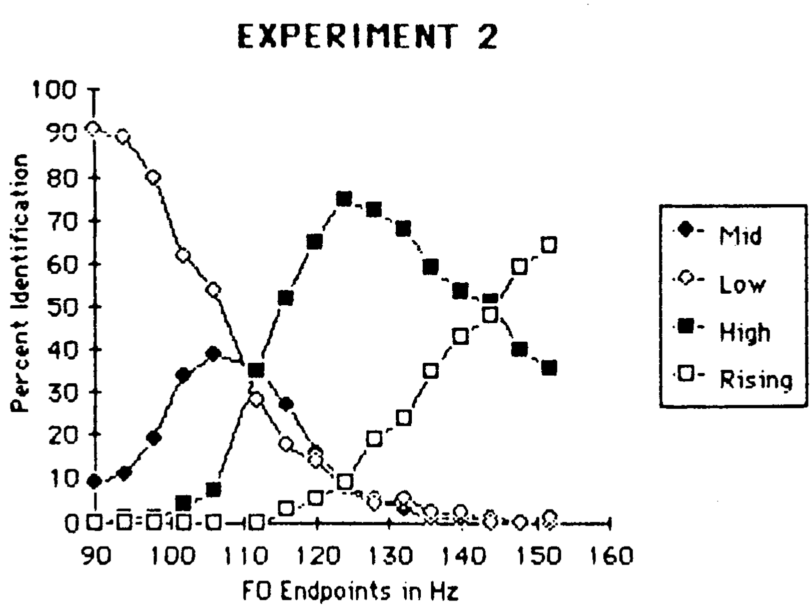

Figure 4. Experiment 2: Identification of the contours of Figure 3 by 38 subjects.

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%. As for the fourth hypothesis, it is true that the major peaks in the figure are for the low and high categories, but the labeling function for the rising tone is conspicuous too, while the mid tone, reaching a peak of 39%, is at least not negligible.

Experiment 3

Figure 5 shows the stimuli for this experiment. They are 17 F_0 contours on tokens of the [k^{h}a:] carrier syllable. The contours all start at 90 Hz, the bottom of the simulated voice range, and go to endpoints ranging, once again, from 90 to 152 Hz in 4-Hz steps. (The exception is the first step, which is from 90 to 92 Hz.) The original intent had been to make 92 Hz the bottom frequency.

This array was meant to explore four hypotheses: (1) The onsets are too low in the voice range to yield the mid tone. (2) The low onsets should give a much better rising category than in Experiment 2. (3) There should be no high-tone responses. (4) The first two or three contours at the bottom ought to be heard mainly as the low tone.

The results of Experiment 3 are given in Figure 6. With the labeling function of the mid tone hovering around 10% over the first half of the stimulus array and then dropping to nothing, the first hypothesis is well supported.

The rising-tone category is clearly more robust here than in Experiment 2, thus confirming the second hypothesis. More abrupt rises to the same endpoints produce more convincing tokens of the rising tone. Although the labeling function for the high tone is rather poor, with a plateau at about 40% for four of the stimuli, this result does not bear out the very categorical prediction of the third hypothesis. Of course, this should be compared with Experiment 2 in which the higher starting point led to a much more robust high-tone percept. In agreement with the fourth hypothesis, the first few contours are heard predominantly as the low tone; however, the greater area under the "low" curve in Experiment 2 (see Figure 4) suggests that a slight fall enhances the acceptability of those stimuli.

Experiment 4

This time, the full voice range furnishes the set of beginning points and the top of the range, the endpoint. Thus, as shown in Figure 7, the beginnings of the 16 contours range from 90 to 152 Hz in 4-Hz steps, except for a 5-Hz step at the bottom (90 to 95) and a 3-Hz step at the top (149 to 152). All the contours end at 152 Hz.

The hypothesis here is that only the high and rising tones should be heard. This portion of the tonal space seems utterly unsuitable for any other tone.

In fact, aside from the essentially negligible "low" labels along the bottom of the graph in Figure 8, the two categories that emerge are the high and rising tones. Interestingly enough, the stronger of the two categories is the high tone. Apparently, these less abrupt rises, compared with those of Experiment 3 (Figures 5 and 6), bias the response toward the high tone.

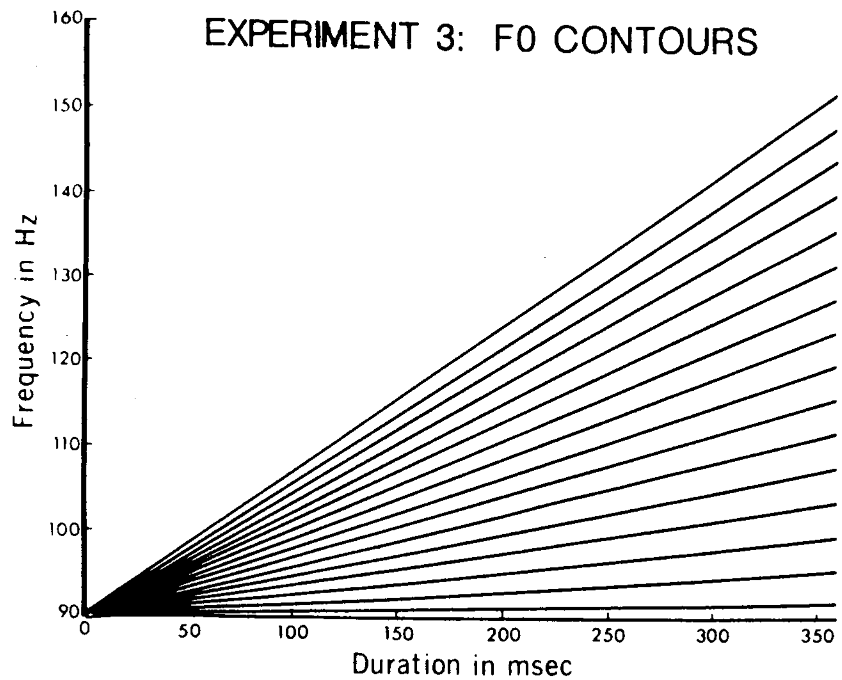

Figure 5. Seventeen F_0 contours moving from 90 Hz to endpoints ranging from 90 to 152 Hz.

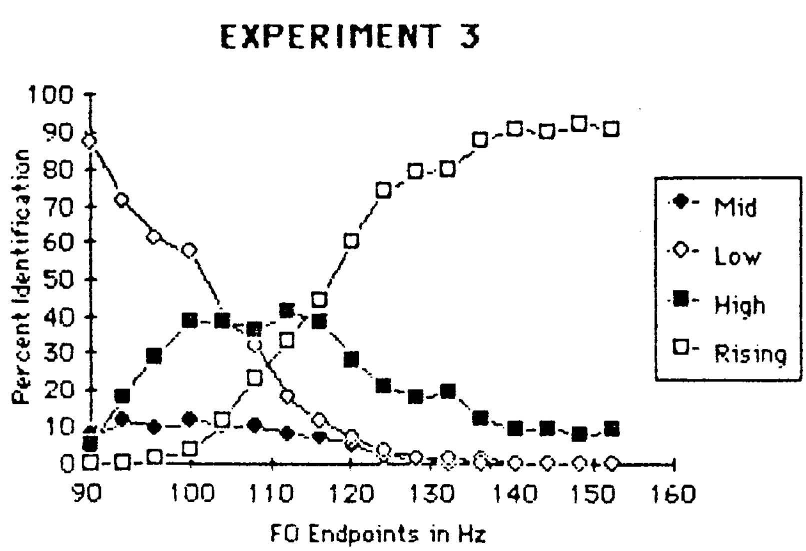

Figure 6. Experiment 3: Identification of the contours of Figure 5 by 38 subjects.

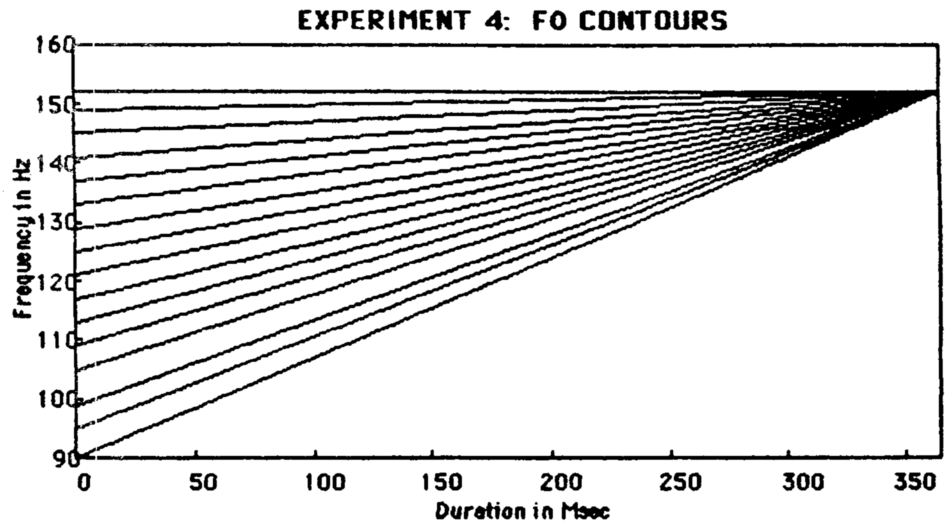

Figure 7. Sixteen F_0 contours starting at points ranging from 90 to 152 Hz, all ending at 152 Hz.

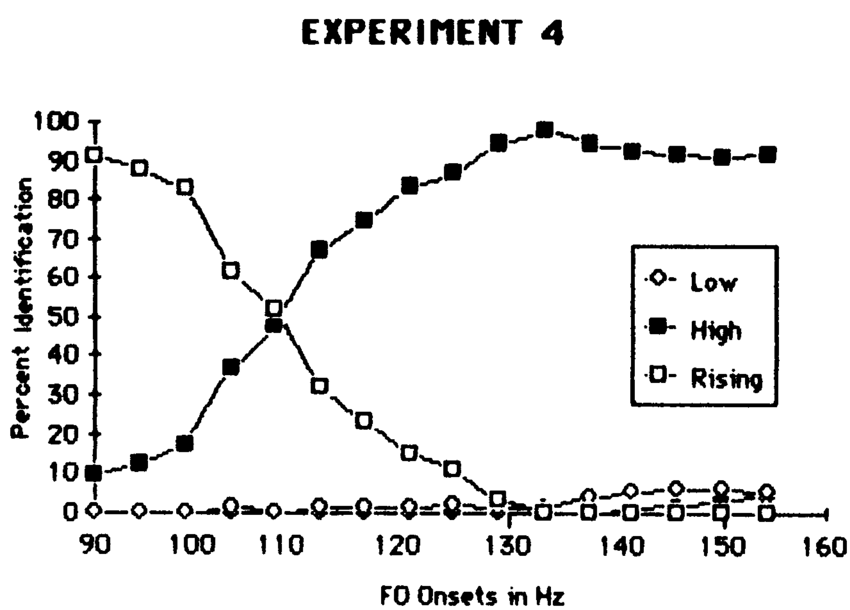

Figure 8. Experiment 4: Identification of the contours of Figure 7 by 38 subjects.

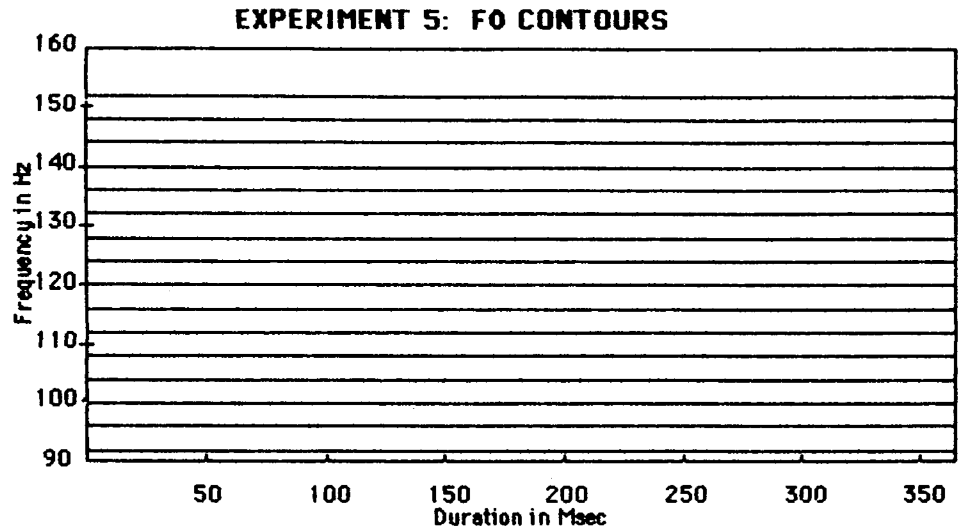

Figure 9. Sixteen level F_0 contours ranging from 92 to 152 Hz.

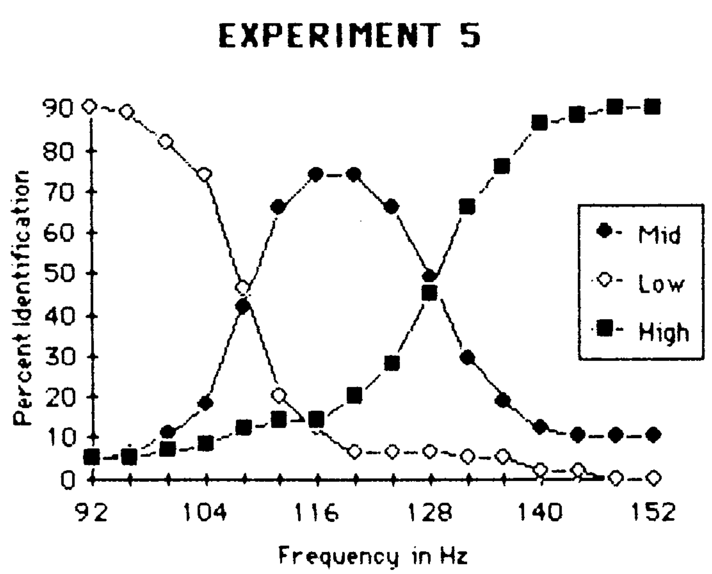

Figure 10. Experiment 5: Identification of the contours of Figure 9 by 37 subjects (adapted from Figure 2 in Abramson 1978).

Experiment 5

Here, on tokens of synthetic [k^{h}a:], there are 16 level contours, ranging from 92 to 152 Hz in 4-Hz steps, as seen in Figure 9. These are undoubtedly a greater deviation from natural speech than any of the foregoing contours; nevertheless, given the frequent assumption of "level" tones in the linguistic literature, it was important to see what the perceptual response to such stimuli would be. Indeed, the hypothesis expected only static tones, that is, the mid, low, and high tones.

The results, first presented in Abramson (1978), are given in Figure 10. Only the mid, low, and high categories appear. There is much overlap, resulting in a lower peak for the mid tone than for the other two.

Conclusion

This study continues to support the primacy of the fundamental frequency of the voice as the carrier of tonal information in Thai, although some concomitant features may, in certain contexts, have at least secondary cue value. The "ideal" contours found in earlier work (Abramson, 1962; Erickson, 1974; Gandour, 1975) are still quite acceptable for isolated Thai words.

The new work has yielded some information on the perceptual latitudes of four of the tones. Level contours are fairly good for the static tones. For absolute levels to be so identified in citation forms of words in natural speech, there must be some auditory accommodation to the speaker's voice range (Abramson, 1976; Leather, 1983), as well as to the immediate tonal context. A comparison of Figures 8 and 10 does reveal, however, that the high-tone percept is improved by F_0 movement. (Similar observations were made for the mid and low tones in Abramson, 1978.)

Fairly rapid movements are needed for the dynamic tones. This conclusion is supported here only for the rising tone, although data not presented here show the same effect for the falling tone. While the dichotomy between static and dynamic tones is thus not categorical, it does have some perceptual support.

There is more work to be done on the tonal space for Thai and other languages. The present findings seem compatible with the pitch features isolated by Gandour (1978) and the emphasis on the importance of the onset frequency values of the contours for Thai by Saravari and Imai (1983). Of course, in running speech, all this is further complicated by interactions between sentence intonation and the tonal space (Abramson & Svastikula, 1983).

References

Abramson, A. S. (1962). _The vowels and tones of standard Thai: Acoustical measurements and experiments_. Bloomington, IN: Indiana Univ. Research Center in Anthropology, Folklore, and Linguistics, Pub. 20.

Abramson, A. S. (1975). The tones of Central Thai: Some perceptual experiments. In J. G. Harris & J. R. Chamberlain (Eds). _Studies in Tai linguistics in honor of William J. Gedney_ (pp. 1-16). Bangkok: Central Institute of English Language.

Abramson, A. S. (1976). Thai tones as a reference system. In T. W. Gething, J. G. Harris, & P. Kullavanijaya (Eds.), Tai linguistics in honor of Fang-Kuei Li (pp.1-12). Bangkok: Chulalongkorn University Press.

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. Language and Speech, 21, 319-325.

Abramson, A. S., & Svastikula, K. (1983). Intersections of tone and intonation in Thai. Haskins Laboratories Status Report on Speech Research, SR-74/75, 143-67.

Bradley, C. B. (1911). Graphic analysis of the tone-accents of the Siamese language. Journal of the American Oriental Society, 32, 282-289.

Erickson, D. (1974). Fundamental frequency contours of the tones of standard Thai. Pasaa, 4, 1-25.

Gandour, J. (1975). On the representation of tone in Siamese. In J. G. Harris & J. R. Chamberlain (Eds), Studies in Tai linguistics in honor of William J. Gedney (pp. 170-195). Bangkok: Central Institute of English Language.

Gandour, J. (1978). The perception of tone. In V. A. Fromkin (Ed.), Tone: A linguistic survey (pp. 41-76). New York: Academic Press.

Haas, M. R., & Subhanka, H. R. (1945). Spoken Thai. New York: Henry Holt.

Leather, J. (1983). Speaker normalization in perception of lexical tone. Journal of Phonetics, 11, 373-382.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress Phonetic Sciences Prague 1967 (pp. 563-567).

Saravari, C., & Imai, S. (1983). Perception of tone and short-long judgement of vowel of variants of a Thai monosyllabic sound. Journal of Phonetics, 11, 231-242.