The Tones of Central Thai: Some Perceptual Experiments*

Arthur S. Abramson†
Haskins Laboratories, New Haven, Conn.

INTRODUCTION

In recent years, research into prosodic features as part of the effort to understand the nature of speech communication has become an increasingly conspicuous aspect of experimental phonetics (Fry, 1968; Lieberman, 1974). Within prosody, such features as phonologically distinctive stress, segment length, and tone are clearly central to the sound pattern of a language in that they differentiate linguistic expressions. These are usually of more immediate interest to the linguist than are other types of prosodic features, and it is usually easier to design experiments testing hypotheses concerning them.1 Prosodic research of this kind should make contributions to general phonology and, more narrowly, to our understanding of the phonology of a particular language, in this instance Thai.

The research reported here forms part of a larger project; other aspects of the project will be presented elsewhere. Taken together, these studies should furnish much information on the perceptually tolerable ranges of the tones of Thai. The work takes as its starting point the usual assumption that the major phonetic features of phonemic tone are found in the domain of pitch. The primary acoustic correlate of pitch is, of course, frequency. In instrumental analyses of tonal features, then, we measure the fundamental frequency of the voice as determined by the repetition rate of glottal pulsing.

*This is a slightly revised version of a chapter to be published in Studies in Tai Linguistics, ed. by Jimmy G. Harris and James Chamberlain. (Bangkok: Central Institute of English Language, in press).

† Also University of Connecticut, Storrs.

1 This is not to assert that other realms of prosody, such as sentence intonation and rhythm, are irrelevant to the enterprise of phonetic research. Indeed, some important psycholinguistic questions have been raised in work on intonation; see, e.g., Lieberman (1967) and Studdert-Kennedy and Hadding (1973).

Acknowledgment: This research was conducted while the author was on sabbatical leave in Thailand on research fellowships from the American Council of Learned Societies and the Ford Foundation Southeast Asia Fellowship Program. I gratefully acknowledge the hospitality of the Faculty of Humanities, Ramkhamhaeng University, and the Central Institute of English Language, both in Bangkok.

The language under analysis here is Central Thai (Siamese), the regional dialect that serves as the official language of Thailand. All native speakers used as subjects were from Bangkok or its close environs. The question of the tonal homogeneity of the central area of the kingdom has not been well explored and thus cannot be categorically ruled out as a perturbing factor in some of the data presented; nevertheless, the speech of the test subjects gave an impression of general uniformity in tonal behavior.

One aim of this study was to determine how well the five tones of the language could be identified in isolation. It is at least conceivable that the identifiability of one or more of the tones would suffer without the benefit of an immediate context. Another aim was to reaffirm earlier work (Abramson, 1961, 1962) on the sufficiency of certain ideal fundamental frequency contours for the identification of the tones, using synthetic speech in which it would be possible to make frequency contours the only variable. In the expectation that less-than-perfect identification would be achieved with fundamental frequency alone, we also planned to learn whether the addition of variations in the amplitude of the speech signal would enhance the identifiability of the tones. Finally, we proposed to test the strong hypothesis that absolute fundamental frequency heights contribute nothing to the identification of the tones, while the shapes of the frequency contours carry all the information.

**BACKGROUND WORK**

Much of the present work is a continuation of earlier work done by the author (Abramson, 1962) with fewer informants and test subjects. That study showed (p. 128) that sets of tonally differentiated monosyllabic words, as produced by a single speaker, could be correctly identified nearly 100 percent of the time. In addition, fundamental frequency \(f_0\) measurements were taken from a large sampling of monosyllabic words with both short and long vowels, yielding average contours for the five tones (pp. 112-127).

These average \(f_0\) contours were then synthesized to see if Thai speakers could indeed identify each of the tones on the basis of \(f_0\) alone. The synthesizer used was the Haskins Laboratories' Intonator (Abramson, 1962:20), which enabled the experimenter to analyze the spectrum of the speech signal and then resynthesize it on the machine's own "voice" source with new \(f_0\) contours. Thus, most of the phonetic features of the original utterance are kept even when an \(f_0\) contour is imposed. The five average tonal curves were thus imposed upon syllables that had originally carried the mid tone, namely /khaj/ 'dried sweat' and /looo/ 'unstable.' The two perception tests prepared in this way exposed native Thai listeners to the tonal contours on both short and long vowels. The perceptual labeling of the randomized stimuli of these two tests showed clearly that the isolated \(f_0\) contours provided sufficient cues for identifying the tones (pp. 131-132).

---

2 The topic is being studied by Dr. Udom Warotamasikkhat.

3 It is gratifying to note that in data obtained some 14 years later (most of the tonal data for the 1962 publication were collected before 1960) Erickson (1974) provides general verification of the old contours while adding important information on the perturbing effects of initial consonants.

4 In /loolee/. 92
In another experimental condition, five monosyllabic words minimally differentiated by tone were manipulated with the Intonator to yield 25 new syllables. That is, five new syllables were derived from each spoken syllable by removing the original tonal contour and replacing it with the synthetic contours, one by one, on this syllable carrier. This was done to see whether the curves carried enough information to override the effects of other features found in association with the pitch movements of the tones. The results (pp. 131-134) included a small number of confusions apparently attributable to such concomitant features as variations in amplitude and duration, which would be likely to survive the analysis and resynthesis, although by and large the $f_0$ curves were heard as intended. In general, then, the data supported the conclusion that the $f_0$ contours isolated by means of acoustic analysis furnished sufficient cues by themselves for the identification of the five tones of Central Thai.

**Experiment 1: Isolated Monosyllabic Words**

Earlier work (Abramson, 1962:128) indicated that Thai listeners can identify the tones of monosyllabic utterances nearly perfectly. Although four sets of tonally differentiated words were used in that study, the productions of only one speaker furnished the stimuli. R. B. Noss (personal communication) has argued that generalizing from these results is, perhaps, not warranted and that normally the mid and low tones, at least, are difficult to identify unless they are embedded in a context. We thought that this objection might be handled by replicating the experiment with a somewhat larger number of speakers and listeners. In addition, data from responses to real speech were needed to furnish a standard for the later evaluation of results obtained with synthetic speech.

The following set of words was chosen for all the experiments to be described:

<table>
<thead>
<tr>
<th>Tone</th>
<th>Thai Script</th>
<th>Transcription</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid</td>
<td>งกก</td>
<td>/khaa/</td>
<td>'a grass (Imperata cylindrica)'</td>
</tr>
<tr>
<td>Low</td>
<td>งก</td>
<td>/khaa/</td>
<td>'galangal, a rhizome'</td>
</tr>
<tr>
<td>Falling</td>
<td>งก</td>
<td>/khaa/</td>
<td>'slave, servant'</td>
</tr>
<tr>
<td>High</td>
<td>งก</td>
<td>/khaa/</td>
<td>'to engage in trade'</td>
</tr>
<tr>
<td>Rising</td>
<td>งก</td>
<td>/khaa/</td>
<td>'leg'</td>
</tr>
</tbody>
</table>

The tone names are conventional but not fully descriptive. Ten native speakers of Central Thai recorded three or four randomizations each of the list with short pauses such that there were five tokens of each word in each randomization. The speakers, five men and five women, included nine university instructors and one clerk.

The tests were played to 25 native speakers of Central Thai through headphones in the language laboratory (Tandberg Teaching System IS 6 with Beyer

---

5This kind of information can preserve the tonal distinctions to a severely limited extent in whispered speech in which no $f_0$ is present (Abramson, 1972).
DT98B Dynamic Headphones) of Ramkhamhaeng University, Bangkok, twice a week for a month. Only one test order was used for each of the ten speakers. Here and in later experiments the subjects were instructed to write the numerals 0, 1, 2, 3, or 4 for the mid, low, falling, high, and rising tones, respectively. These numbers are appropriate to the nomenclature of traditional Thai grammar and facilitated later scoring. Familiar as the subjects were with this convention, they were nevertheless provided with a sheet at each session showing the five words in Thai script with their numerical equivalents.6

The results are shown in Table 1, which is arranged as a confusion matrix with the tonal stimuli in the first column and the perceptual responses to them in the next five columns. The sixth column shows the total number of responses to each stimulus word. A perfect response to the tones as intended would yield

<table>
<thead>
<tr>
<th>Labels:</th>
<th>Mid</th>
<th>Low</th>
<th>Falling</th>
<th>High</th>
<th>Rising</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid</td>
<td>97.9</td>
<td>2.1</td>
<td></td>
<td></td>
<td></td>
<td>1220</td>
</tr>
<tr>
<td>Low</td>
<td>3.4</td>
<td>96.6</td>
<td></td>
<td></td>
<td></td>
<td>1220</td>
</tr>
<tr>
<td>Falling</td>
<td>0.2</td>
<td>99.1</td>
<td>0.4</td>
<td></td>
<td>0.3</td>
<td>1220</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td>100.</td>
<td></td>
<td></td>
<td>1220</td>
</tr>
<tr>
<td>Rising</td>
<td>0.1</td>
<td>0.4</td>
<td>99.5</td>
<td></td>
<td></td>
<td>1220</td>
</tr>
</tbody>
</table>

Total = 6100
Subjects = 25
% Correct = 98.6

100 percent in each cell along the diagonal from the upper left to the lower right. The overall intelligibility of 98.6 percent, 85 errors out of 6100 responses, is high. Inspection of the responses to the mid and low tones indicates that some confusion between them accounts for most of the small number of errors.

The data were also examined for individual differences among speakers and listeners. One of the ten speakers caused 45.9 percent of the errors, another speaker, 16.5 percent, and a third speaker, 12.9 percent. The recordings of only one speaker produced no errors at all. As for intersubject differences, the worst listener made 12.9 percent of the errors, followed by two who each made 8.2 percent of the errors. Three of the 25 subjects made no errors at all.

6The capable and efficient selection and supervision of the test subjects by Miss Panit Chotibut of the Faculty of Humanities, Ramkhamhaeng University, is much appreciated.
This larger sampling of speakers and listeners supports the earlier conclusion that Thai words minimally differentiated by tones can nearly always be identified correctly even as isolated forms. Some speakers appear to provide minimal perceptual cues for words out of context, especially for the contrast between mid and low tones. Some listeners seem to require more than these minimal cues for the identification task. Finally, the data in Table 1 provide a baseline for the other experiments now to be discussed.

SYNTHETIC SPEECH

Experiment 2: Perception of \( f_0 \) Contours

To test once again the perceptual efficacy of the \( f_0 \) contours derived from speech measurements in the earlier study (Abramson, 1962:112-127), a different speech synthesizer, the Haskins Laboratories' formant synthesizer, was used under control of a computer. For the present experiment, the parameters specified were the frequency and amplitude values of the first three formants, the timing of source functions for voicing and voicelessness, the overall amplitude of the signal, and the fundamental frequency (\( f_0 \)). Steady-state formant frequencies were chosen to yield a vowel acceptable as Thai /aa/; formant transitions (Liberman, Delattre, Cooper, and Gerstman, 1954) appropriate to the velar place of articulation were used, and voiceless aspiration was simulated by providing a long voicing lag (Lisker and Abramson, 1964, 1970) filled with turbulent noise in the regions of the formants. The overall amplitude was kept flat throughout the syllable except for a slight rise at the beginning and a slight fall at the end. These specifications yielded syllables of the type [kha:] which, it was hoped, with suitable \( f_0 \) contours would be heard as the five Thai words listed in the section on Isolated Monosyllabic Words. The five \( f_0 \) contours of the 1962 study (Abramson, 1962:127, Fig. 3.6) were retained with slight adaptations required by the nature of the computer program and imposed one-by-one upon this syllable. The contours, shown in the upper part of Figure 1, covered a range from 92 Hz to 152 Hz, which was reasonable for an adult male voice. Three tokens of each stimulus type thus produced were randomized into six test orders and played to 38 native speakers of Central Thai over a period of a month, together with other tests in the same sessions.

The results of these listening tests are shown in Table 2. Note that the tonal names in the first column are written with quotation marks. This is meant to convey that these \( f_0 \) contours were intended as those tones but can be so labeled only to the extent that the subjects accept them as such. In the same spirit, the word correct at the bottom of the table is also printed with quotation marks. Otherwise, the form of the confusion matrix is the same as that of Table 1.

Experiment 3: \( f_0 \) Plus Amplitude

Changes in the contraction of certain laryngeal muscles\(^7\) and in subglottal air pressure can separately or together produce variations in the fundamental

\(^7\) A University of Connecticut doctoral dissertation by Donna Erickson, soon to be completed, Laryngeal Mechanisms and Coarticulation Effects in the Tones of Thai, explores the role of intrinsic and extrinsic laryngeal muscles in the production of the tones of Thai.
Figure 1: Upper part: $f_0$ contours used in Experiment 2. Lower part: amplitude contours used in Experiment 3 in conjunction with the $f_0$ contours.
TABLE 2: Synthetic speech: responses to five \( f_0 \) contours.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Mid</th>
<th>Low</th>
<th>Falling</th>
<th>High</th>
<th>Rising</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Mid&quot;</td>
<td>82.0</td>
<td>3.5</td>
<td>0.2</td>
<td>14.2</td>
<td></td>
<td>906</td>
</tr>
<tr>
<td>&quot;Low&quot;</td>
<td>7.2</td>
<td>87.3</td>
<td>5.2</td>
<td>0.3</td>
<td></td>
<td>906</td>
</tr>
<tr>
<td>&quot;Falling&quot;</td>
<td>0.4</td>
<td>0.7</td>
<td>97.8</td>
<td>0.6</td>
<td>0.6</td>
<td>906</td>
</tr>
<tr>
<td>&quot;High&quot;</td>
<td>2.0</td>
<td>0.2</td>
<td>97.7</td>
<td>0.1</td>
<td></td>
<td>906</td>
</tr>
<tr>
<td>&quot;Rising&quot;</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>99.1</td>
<td>906</td>
</tr>
</tbody>
</table>

Total = 4530
Subjects = 38
% "Correct" = 92.8

The frequency of the voice. These mechanisms are also available for controlling the intensity of phonation and thus variations in the overall amplitude of the speech signal. To a certain extent, then, the two acoustic features, \( f_0 \) and amplitude, may covary. Since the major psychological correlate of amplitude or intensity is loudness, just as that of \( f_0 \) is pitch, it is not unreasonable to suppose that the ear may detect shifts in loudness in conjunction with large pitch excursions of one or more of the tones. If this is so, even though it has already been demonstrated with synthetic speech that certain ideal \( f_0 \) contours carry sufficient information for the identification of the five tones, the perceptual processing of some of the tones in real speech may include awareness of relative amplitude as a concomitant feature which, under certain conditions, may actually contribute to tonal identification. In the present study the question arises as to whether the discrepancies in intelligibility between Experiments 1 and 2 will be removed simply by the addition of appropriate amplitude contours.

To answer this question, we added the amplitude contours of the lower part of Figure 1 to the corresponding \( f_0 \) contours of the upper part. Amplitude is indicated in decibels (db) as decrements from the maximum output of the synthesizer at zero db. These amplitude contours are approximations derived from inspection of a small sample of amplitude displays made with sound spectrograms of Thai words in isolation. The new synthetic stimuli were randomized five times each into three test orders and played to 40 native speakers of Central Thai over a period of a month, together with other tests. The results are shown in the confusion matrix of Table 3.

Comparison of the Three Conditions

Inspection of Tables 1-3 shows that the overall identifiability of the stimuli moves from 98.6 percent for real speech through 92.8 percent for fundamental frequency alone to 96.1 percent for \( f_0 \) plus amplitude, thus suggesting that while \( f_0 \) alone is by and large a sufficient cue, its efficacy is enhanced
by the addition of amplitude information. A comparison of the correct mean scores in the diagonals across the matrices is more to the point than just the overall scores. As we move from real speech (Table 1) to f₀ alone (Table 2), the most striking changes are in the cells for the mid and low tones which lose, respectively, 15.9 percent and 9.3 percent. In Table 3, for f₀ combined with amplitude information, the entries in the corresponding two cells move back in the direction of real speech, although the improvement is considerably greater for the mid tone.

Turning to the confusions in the matrices, we note a much greater scattering of errors in the two synthetic speech experiments as compared with real speech. Specifically, in Table 2 there is some confusion between the mid and high tones. The intended mid tone is called high 14.2 percent of the time. There is also some confusion in the other direction, high heard as mid, but only 2 percent of the time. Table 3 shows that the addition of amplitude information eliminates most of the confusion between these two tones. The only notable confusion in the real-speech test of Table 1 is between the mid and low tones. This confusion is even worse in Table 2. Under both conditions the hearing of the intended low tone as mid accounts for most of the confusion. This latter effect is not eliminated by the addition of amplitude information in Experiment 3, although the identification of the intended mid tone itself is now improved to the level of real speech or slightly better. The intended low tone is heard as falling 5.2 percent of the time in Table 2 but is nearly back to the level of real speech in Table 3. One small but puzzling distortion apparently caused by

---

8An experiment not performed as part of this research would be to try amplitude contours alone and then supplement them with f₀ information. Previous research (Abramson, 1972) implied that amplitude alone would be not nearly sufficient for perception of the tones.
the addition of amplitude information is the hearing of the intended falling tone as high 2.5 percent of the time in Table 3.\(^9\)

It seems safe to infer from the results of the preceding experiments that \(f_0\) contours carry most of the information for tonal identification in Thai; that is, they carry sufficient information most of the time to identify words that are minimally distinguished by tone. A concomitant feature of some relevance for at least part of the tone system is contained in changes in the overall amplitude of the utterance. The confusions in the various matrices indicate the need for further information to improve the synthesis of Thai tones by rule. The improvements needed may be small refinements of the \(f_0\) contours and better amplitude specification. In addition, simulating glottal tension in the voice source might make the high tone more natural and acceptable in utterance-final position.

**Experiment 4: Perception of \(f_0\) Levels**

The five tones of Central Thai can be viewed as falling into two groups, the dynamic tones and the static tones (Abramson, 1962:9-11). In this scheme, the sharp upward and downward movements of the rising and falling tones place them in the dynamic category. Since the high, mid, and low tones often sound as if they simply occupy three levels, they are classified as static. Of course, the acoustical measurements, as reflected in the upper part of Figure 1, show that even the static tones undergo some \(f_0\) movement. Kenneth Pike (1948:5) speaks of level tonemes and gliding tonemes: "A LEVEL toneme is one in which, within the limits of perception, the pitch of the syllable does not rise or fall during its production. A GLIDING toneme is one in which during the pronunciation of the syllable in which it occurs there is a perceptible rise or fall, or some combination of rise and fall, such as rising-falling or falling-rising."

It may be the case, then, that speakers of the language, not to mention field phoneticians, hear the static tones of Thai as simple levels. A lengthy quotation from Pike (1948:4) is of considerable interest here:

Tone languages have a major characteristic in common: It is the relative height of their tonemes, not their actual pitch, which is pertinent to their linguistic analysis. It is immaterial to know the number of vibrations per second of a certain syllable. The important feature is the relative height of a syllable in relation to preceding and following syllables. It is even immaterial, on this level of analysis (but not in the analysis of the linguistic expression of emotion), to know the height of a specific syllable in proportion to the general average pitch which the speaker uses. Rather, one must

---

\(^9\)The effects set forth here seem obvious from simple inspection of the tables. Indeed, statistical analysis (t-tests for differences between correlated means) performed with the kind help of Dr. Lyle Bachman of the Central Institute of English Language and the Ford Foundation, Bangkok, shows them by and large to be significant at the 5 percent level of confidence or better. More refined statistical procedures might lead to further observations, but such effects would probably be so subtle as to be uninteresting for our understanding of the perception of tones.
know the relationship of one specific syllable to the other syllables in the specific context in the particular utterance. A man and a woman may both use the same tonemes, even though they speak on different general levels of pitch. Either of them may retain the same tonemes while lowering or raising the voice in general, since it is the relative pitch of syllables within the immediate context that constitutes the essence of tonemic contrast.

Ilse Lehiste (1970: Chap. 3) provides a useful survey of these matters and related questions. 10

How likely is it that absolute values of f₀ levels provide sufficient cues for the recognition of the tones of a given speaker of a tone language in a certain context? The relativity of the pitches of tones (Cook, 1972) is usually taken for granted. Indeed, the few studies that have yielded acoustical measurements of tones, e.g., for Mandarin Chinese (Howie, 1972) and Thai (Abramson, 1962; Erickson, 1974), show that the tones are characterized by at least some movement and, in some cases, much movement. That is, they tend not to be uttered with a flat, unchanging f₀. The one tone in Thai that appears most likely to have a flat f₀ in certain nonfinal environments is the mid tone. In an interesting experiment, Victor Zue (discussed in Klatt, 1973) demonstrated that Howie's Mandarin contours still showed rather high intelligibility even when the range of absolute f₀ is severely compressed, thus indicating that the pitch movement still available to the listeners carried sufficient information.

Of course, most of the foregoing observations are derived from tones in isolated words or at least in very short utterances. It seems likely that at least for some of the tones of Thai, presumably the high, mid, and low, the perturbations of their f₀ contours occasioned by the many coarticulations of running speech should produce flat variants here and there (Abramson, in preparation). If so, are such words understood by virtue of contextual redundancy or, to return to the question raised in the preceding paragraph, do the absolute levels furnish sufficient cues? That is, are "level" pitches assigned to tones only when they are the perceptual responses to small f₀ movements, or will true acoustic levels suffice? Experiment 4 was designed to provide some answers to the question.

As shown in Figure 1, the "voice" of the synthesizer in the experiments reported so far was set to range from 152 Hz down to 92 Hz. In this experiment the range was divided to produce 16 flat fundamental frequencies in steps of 4 Hz; the amplitudes were flat too, except for a slight rise at the beginning and a slight fall at the end. These variants were imposed upon the same basic syllable, randomized, and played to 37 native speakers of Thai for identification as members of the same set of five words as before. Table 4 reveals that the falling and rising tones, which are characterized by very dynamic movements, elicited practically no responses. Indeed, the 0.1 percent response to the 100 Hz level as the rising tone is so improbable as to suggest momentary inattention. From top to bottom, we have here a gradual crossover from the high

---

10 An important background article for the phonological treatment of tone is Wang (1967).
TABLE 4: Synthetic speech: responses to 16 level \( f_o \)'s.

<table>
<thead>
<tr>
<th>Hz</th>
<th>Mid</th>
<th>Low</th>
<th>Falling</th>
<th>High</th>
<th>Rising</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>152</td>
<td>8.0</td>
<td>4.1</td>
<td>0.2</td>
<td>87.7</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>148</td>
<td>7.9</td>
<td>4.2</td>
<td>0.3</td>
<td>87.6</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>144</td>
<td>8.6</td>
<td>4.1</td>
<td>0.2</td>
<td>87.0</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>140</td>
<td>12.1</td>
<td>4.3</td>
<td>0.1</td>
<td>83.5</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>136</td>
<td>18.6</td>
<td>5.4</td>
<td>0.1</td>
<td>75.9</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>132</td>
<td>29.2</td>
<td>5.4</td>
<td></td>
<td>65.3</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>128</td>
<td>49.3</td>
<td>6.2</td>
<td></td>
<td>44.5</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>124</td>
<td>65.3</td>
<td>5.5</td>
<td>0.1</td>
<td>29.0</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>120</td>
<td>72.6</td>
<td>6.3</td>
<td>0.1</td>
<td>20.9</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>116</td>
<td>73.0</td>
<td>12.2</td>
<td></td>
<td>14.7</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>112</td>
<td>66.4</td>
<td>19.7</td>
<td></td>
<td>13.8</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>108</td>
<td>42.1</td>
<td>45.7</td>
<td></td>
<td>12.2</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>104</td>
<td>18.4</td>
<td>73.8</td>
<td></td>
<td>7.9</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>100</td>
<td>11.0</td>
<td>81.5</td>
<td>0.3</td>
<td>7.1</td>
<td>0.1</td>
<td>903</td>
</tr>
<tr>
<td>96</td>
<td>5.5</td>
<td>88.7</td>
<td>0.1</td>
<td>5.6</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>92</td>
<td>4.8</td>
<td>90.1</td>
<td>0.1</td>
<td>5.1</td>
<td></td>
<td>903</td>
</tr>
</tbody>
</table>

Total = 14,448
Subjects = 37
tone through the mid tone to the low tone. Nowhere is 100 percent identification as a particular tone achieved. The closest is a peak of 90.1 percent for the low tone, which is comparable with the peak of 87.3 percent for the low tone in Experiment 2 (Table 2). The other two peaks here, 73 percent for the mid tone and 87.7 percent for the high tone, compare somewhat less well with their counterparts, 82 percent and 97.7 percent respectively, among the ideal contours of Experiment 2. Despite these peaks, it is important to note that all three tones persist in eliciting responses throughout their ranges. Most of this is accounted for not by the sporadic responses of all the subjects but rather the deviant response behavior of three of them. One subject agreed with the main group in calling the upper part of the range the high tone, but at 120 Hz she started crossing over to the low tone and remained there the rest of the way down with only scattered responses in the mid-tone column. The second subject of the three called the upper part of the range the mid tone and crossed over to the low tone at 108 Hz. The third subject deviated in the most surprising way from the performance of the main group: she assigned the upper part of the range to the low tone and, crossing over at about 120 Hz, the rest of the range to the high tone! (This subject's correct use of labeling conventions in other tests taken during the same sessions shows that she is not guilty of misuse of labels here.) Apparently, however, her psychological set shifts from time to time, because in some of the test sessions she assigned variations in the lower part of the range to the mid tone.

We may infer from the results of Experiment 4 that even in isolated monosyllabic words unchanging levels of fundamental frequency can carry considerable information as to the identity of the high, mid, and low tones. We suppose that in such a situation some accommodation to the speaker's pitch range is necessary. The subjects who took these tests were quite used to the "voice" of the synthesizer, and care was taken to confine the absolute levels of $f_0$ to the range already in use in other tonal experiments. At the same time, the fact that there is no gliding movement at all in the stimuli seems to cause a certain amount of confusion across the three categories that the subjects accepted. Indeed, for three of the 37 subjects this factor was perceptually very disrupting. They may represent a population of Thai speakers for whom the small gliding movements normally found in productions of the static tones in isolation are essential for correct identification. In the absence of any $f_0$ movement at all, it is not surprising that the falling and rising tones were not used as response categories.

Conclusion

The experiments described here lead to a few general conclusions about some aspects of the perception of the tones of Thai. First of all, it is more evident than heretofore that the intelligibility of tonally differentiated monosyllables presented in isolation is quite high. In addition, the average fundamental frequency contours obtained some years ago (Abramson, 1961, 1962), when applied as instructions to the parameters of a different synthesizer (and therefore presumably other synthesizers), still carry enough information for acceptable synthesis of Thai words. A slight reduction of the discrepancy in intelligibility between these contours and those of real speech can be effected by adding rough approximations of natural amplitude movements often found in correlation with shifts in fundamental frequency. To eliminate the small remaining discrepancy further work is needed. Finally, a continuum of level fundamental
frequencies can be divided perceptually by native speakers of Thai into the high, mid, and low tones, but with very gradual transitions and with rather aberrant response patterns on the part of some test subjects; these data suggest that while levels, even in isolated syllables, can carry much information about these tones, there is still a fair amount of interference from the abnormal lack of change in fundamental frequency.

Given the paucity of similar perceptual data for other tone languages, it is hard to say how we may generalize these findings beyond the Thai language. In fact, even within Thai there is little information about other major regional dialects. Such phonetic features as 'creaky voice' or 'glottal tension' prominent in some tonal systems will probably require parameters other than simple control of fundamental frequency or amplitude for experimental investigation. Glottal tension is found in the high tone of Central Thai but appears to be unstable. Reports on more complicated manipulations of fundamental frequency to elucidate further the nature of tonal perception in Thai will be forthcoming.

REFERENCES


11 The most comparable study that comes to mind is Howie's (1972) work on Mandarin; for more detail see his doctoral dissertation (1970). For a study of Cantonese, see Chan (1971).

12 Field work about to be started in Thailand by Jack Gandour of the University of California at Los Angeles should yield information along these lines.


