Pitch in the Perception of Voicing States in Thai: Diachronic Implications*

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

It is tempting for the experimental phonetician to believe that phonetic hypotheses on the causes of sound change should be testable in the laboratory. In the absence of any technological innovation that allows the resurrection of long-dead informants for ever so brief a stint of field work, perhaps the most we can hope to do is to test the phonetic plausibility of these hypotheses by using present-day speakers of one or more of the languages concerned. Even if little light is shed on the historical process, new information on the phonetic nature of the phonological categories of interest may be added to the literature. This paper represents just such an attempt. It examines changes in stop consonant voicing in the Tai family of languages by seeking new information on acoustic cues in modern Thai.

Changes in stop voicing must be viewed against the background of the putative emergence of tones in the Tai family (Gedney, 1974) as a function of initial consonants. Many scholars, e.g., Maspéro (1911), Li (1947), and Coedès (1949), have argued that for Tai and other families of Southeast Asia low tones have developed in word classes with ancient voiced initials, and high tones have developed in word classes with voiceless initials. Such an argument has at least indirect support from acoustic phonetic research, principally on English. House and Fairbanks (1953), as well as Lehisé and Peterson (1961), showed that the fundamental frequency ($f_0$) of phonation soon after the release of an English voiceless consonant is higher than after a voiced consonant. The phonetic rationale for the effect is that the unimpeded air flowing through the open glottis for the voiceless consonant momentarily perturbs the vibration rate of the vocal folds upward, once voicing begins, while the somewhat impeded air flow of the essentially closed glottis for voiced consonants may provide insufficient force to keep the vibration rate at the intended level and thus allow a slight drop in frequency during and just after the consonant closure or constriction. Recovery from the $f_0$ perturbation can take longer than the


†Also University of Connecticut, Storrs.

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.

[HASKINS LABORATORIES: Status Report on Speech Research SR-41 (1975)]
probable duration of the aerodynamic perturbing factor itself. We may suppose
that one transient is caused by the disturbing force, and a second transient is
manifested by the \( f_0 \) movement from its momentary excursion back toward the in-
tended contour of the syllable.

Modern Thai (Siamese) has three categories of stop consonants, usually
called voiced, voiceless unaspirated, and voiceless aspirated. In this study I
have taken the labial stops to represent the system and concentrated on them.
Recently Erickson (in press) has found that the voiced labial stop of Thai
typically shows a low \( f_0 \), while the two voiceless stops show high values. As
for the latter pair, the aspirated stop tends to have a higher value than the
unaspirated stop. [While agreeing with her general findings on the voiced-
voiceless distinction, in recent work Gandour (1974) surprisingly finds that the
aspirated stops show a smaller upward swing of \( f_0 \) than do the unaspirated stops.]
The notion is that in Proto-Tai such adjustments of \( f_0 \) were heard as pitch per-
turbations that were gradually enhanced in speech until they achieved phonemic
status. The argument would apply whether we are supposing a pristine state of
tonelessness in Proto-Tai or indeed a phonology that already included, say, two
tones, since the daughter languages today have five or more tones.

The Proto-Tai system of stop consonants, epitomized by the labials, is
shown in Table 1. These reconstructions and their subsequent changes to the
stop consonants of modern Thai represent the consensus of most scholars, except
that there are serious questions about the phonetic nature of the so-called
glottalized b. Haudricourt and Martinet (1946), incidentally, posit murmured
or voiced aspirated */bh/ as an intermediate stage between */b/ and */ph/.

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proto-Tai:</td>
</tr>
<tr>
<td>Thai (Siamese):</td>
</tr>
<tr>
<td>Script:</td>
</tr>
</tbody>
</table>

Using techniques of speech synthesis, other investigators—Fujimura (1971)
for Japanese and English, and Haggard, Ambler, and Callow (1970) for English—
have shown that pitch shifts can influence auditory judgments as to the voicing
assignments of syllable-initial stop consonants. In addition, Lea (1973) has
been very successful in using a \( f_0 \) criterion in making voicing identifications
of consonants in acoustic analysis of English utterances. To the best of my
knowledge, such experiments, particularly perceptual ones, have not been tried
for languages with more than two consonant categories distinguished by voicing
features.

My plan was to see whether pitch shifts, brought about by control of the \( f_0 \)
parameter of a speech synthesizer, would affect listeners' judgments as to the
voicing-class membership of initial stops. By way of background, it must be
said that some years ago Lisker and I (Lisker and Abramson, 1964; Abramson and
Lisker, 1965) showed, both acoustically and perceptually, that the three stop
categories of Thai lie along the dimension of voice onset time, namely, the
temporal relation between the closing of the glottis for audible pulsing and the
release of the occlusion of the initial stop. To furnish a baseline for the
present research, it was necessary to replicate the perceptual part of this old study with the new subjects who were to be used for the experiments on the efficacy of fundamental frequency perturbations. I used the Haskins Laboratories' parallel resonance synthesizer to produce a syllable of the type labial stop plus \( [a:] \). Thirty-seven variants of the syllable were made to form a continuum of voice onset time ranging from a voicing lead of 150 msec, before the release of the stop, to a voicing lag of 150 msec, after the release. The range was divided into 10-msec steps except for the portion from a lead of 10 msec to a lag of 50 msec, which was divided into 5-msec steps. For voicing lead, I simply had low-frequency harmonics during the simulated stop occlusion. For voicing lag, during the interval after the release when no voicing is present, the second and third formants were filled with noise to simulate aspiration and the first formant was simply omitted to simulate the extreme consequence of an open glottis. I also tapered the overall amplitude in ways roughly appropriate to the effects of laryngeal timing. I restricted the experiment to the mid tone of the five tones of Thai by providing a flat fundamental frequency contour except for a slight dip at the end. The identification data for 48 native speakers of Thai are presented in Figure 1. These subjects were presented with the stimuli randomized into eight test orders for labeling as initial stop consonants. The ordinate shows percent identification. The abscissa shows values of voice onset time. Voicing lead is indicated in negative numbers, voicing lag in positive numbers, while zero means voice onset at the moment of release. The three expected categories emerge, although the middle one, unaspirated \( p \), loses responses to the two categories on either side and does not get as close to 100 percent. The 50 percent crossover values between categories fall at -7 and +26 msec.

With the sufficiency of voice onset time as a cue once again demonstrated for Thai, I went on to new experiments. Unlike Haggard et al. (1970), who used \( f_0 \) excursions far greater than any observed in the literature, I restricted my range to 20 Hz above a reference level and 20 Hz below. This choice is well in accord with Erickson's (in press) values for nine speakers of Thai. She found five male and four female adults to produce \( f_0 \) perturbations for stops well within a range of 40 Hz and just one female with a range of 52 Hz. I set the level portion of my mid tone at 120 Hz and shifted upward to it from 110 and 100 Hz and downward to it from 130 and 140 Hz. For these \( f_0 \) shifts I used three time spans: 50, 100, and 150 msec. I also made variants with no \( f_0 \) shift, that is, a level \( f_0 \) onset. Finally, for all these conditions I provided 13 voice onset time variants, the ones shown along the bottom of the graph in Figure 2. These values were chosen by pretests and inspection of the data of Figure 1. Thus for each voice onset time value there were 13 \( f_0 \) variants, a flat one plus 12 perturbations, yielding 117 stimuli that were randomized eight times with a sample at the beginning of each tape.

The labeling responses of 46 subjects (two having dropped out) are given for the flat \( f_0 \) onsets in Figure 2. This set is presented alone to make it easier to look at the rest of the graphs. The voice onset time values are arrayed along the bottom, while the percentages are at the left end. The bars are coded to show responses in terms of the three initial consonants. If we extrapolate from these bars, the results accord well with the perceptual crossover points in the preceding graph, falling at -10 msec and +22 msec. The responses to all of the \( f_0 \) perturbations for each of the three time spans, 50, 100, and 150 msec, are given in Figures 3-5, respectively. From top to bottom on each page there are graphs for the four frequency shifts.
THAI LABIAL STOPS

N = 440
Ss = 48

PERCENT IDENTIFICATION

VOICE ONSET TIME IN MSEC

FIGURE 1
THAI STOPs: LEVEL F₀ AT 120 Hz
N = 224
S₅ = 4.6

V O I C E  O N S E T  T I M E  I N  M S E C

P E R C E N T  I D E N T I F I C A T I O N

Figure 2
THAI STOPS: 50 MSEC SHIFT OF $F_0$ TO 120Hz

Figure 3
THAI STOPS: 100 MSEC SHIFT OF F₀ TO 120 Hz

Figure 4
THAI STOPS: 150 MSEC SHIFT OF F₀ TO 120 Hz

ONSET: 110Hz

PERCENT IDENTIFICATION

100Hz

130Hz

140Hz

VOICE ONSET TIME IN MSEC

Figure 5
If we look only at the two ends of the voice onset time continuum, \(-100\) and \(+80\) msec, we cannot see that the Thai subjects are influenced by pitch perturbations of any duration they may hear. As a matter of fact, as we look over all the bar graphs, it becomes clear that voice timing is a far more powerful cue than pitch shifts. In general, the perceptual crossover points between categories are not moved; instead, the distribution of values within each category is pushed and pulled in both directions. If, however, you scan down the \(-20\)-msec columns in Figure 4 for 100 msec and Figure 5 for 150 msec, you will find that for onset values of 130 and 140 Hz we do have a boundary shift; for the distinction between the voiced and voiceless unaspirated stops, the boundary shifts leftward from \(-7\) msec to about \(-20\) msec with more stimuli assigned to the voiceless category under the influence of a long duration of fundamental frequency fall. Although the 50-msec shift seems to be too short to provide for a boundary change, at \(+5\) msec we see the voiced category succumbing to the voiceless aspirate. The boundary between the voiceless unaspirated and aspirated stops does not shift at all, remaining at about \(+22\) msec. Indeed in the region of this boundary it is hard to see a consistent trend. One exception would appear to be on the 50-msec display where we see that at this boundary for the 100-Hz onset, responses are pulled from the aspirate to the inaspirate, as might be expected. A refined statistical analysis, yet to be performed, may yield a few more subtle tendencies.

We may conclude then that perturbations of fundamental frequency at the beginnings of syllables with initial stops can influence voicing judgments in Thai. The effect is enhanced with greater durations of frequency shift. It seems to favor the boundary region between the voiced stop and the voiceless inaspirate. There are some effects at the boundary between the two voiceless categories, but they are less consistent and not easy to interpret. Shifts of fundamental frequency, to be ascribed, along with the feature of voice onset time, to states of the larynx have some cue value in Thai, although they are clearly subordinate to voice onset time. That is, the effects are most marked in zones of perceptual ambiguity along the voice timing continuum.

The effects found in this study do lend some support to the argument that the emergence of tones in Proto-Tai or, perhaps, the increase in the number of tones, could have been a conditioning factor in the shifting and merging of consonant voicing categories. We can imagine something like the following situation. As the pitch perturbations associated with the voicing states of initial consonants became apparent to speakers of the language and gradually moved toward phonemic status as tones, the vowel allophones with their concomitant pitch characteristics became more and more differentiable. The pitch coloring must have taken up increasing amounts of time as it became more noticeable; this is implied by my effect with the longer durations of fundamental frequency shift. Brown (1965) has suggested that speakers of the language concentrated perceptually more on the central portion of the word at the expense of attention to the initial consonant when arriving at a lexical decision. In this way, the syllable initial became less and less important. We may speculate that children learning these lexical classes with a shifted perceptual set must have begun to rearticulate the initials, as a deviation from the practice of their elders, more in conformity with what they heard, namely, a shift in the voicing boundary conditioned by pitch. Finally, my data make the unaspirated stop at least as reasonable as an intermediate stage for the change from voiced to voiceless aspirated stop in modern Thai as is the murmured stop posited by Haudricourt and Martinet (1946).
REFERENCES


