In B. Lindblom and S. Öhman (eds.), Frontiers of Speech Communication. London Academic Press, 1979.

The Noncategorical Perception of Tone Categories in Thai*

A. S. ABRAMSON

The University of Connecticut and Haskins Laboratories, Connecticut, USA

Background

Theories of the categorical perception of speech seek to explain how levels of acuity of discrimination vary with the type of phonetic segment involved. The now classical paradigm requires the preparation of a continuum of variants along some physical dimension susceptible to auditory division into phoneme categories. Zones of ambiguity between the groups of variants labelled as phonemes are the so-called phoneme boundaries. In the ideal case of categorical perception, subjects will not do much better than chance in discriminating variants within the labelled categories, but will show rather high peaks of discrimination in the regions of the phoneme boundaries. This is best shown in experiments with stop consonants (Liberman *et al.*, 1957). A rather different set of results prevails with steady-state isolated vowels (Fry *et al.*, 1962). Variants are indeed grouped into phoneme categories but with somewhat more overlap than for stops; however, discrimination is very good along the whole continuum without any special effects at the category boundaries. Two ways of explaining these differences have dominated the literature. The motor or articulatory-reference theory (Liberman *et al.*, 1967) points out that stop consonants and presumably certain other types of phonetic segments are essentially discontinuous in their mode of production, while vowels are continuously graded in production. Thus, it is claimed that the extent of categoricalness in perception is shaped at some psychological

* This is a slightly revised version of an oral paper given before the 93 Meeting of the Acoustical Society of America, The Pennsylvania State University, State College, Pennsylvania, 6–10 June, 1977. The stimuli for this research were made at Haskins Laboratories and the data were analysed there under Grant HD-01994 from the National Institute of Child Health and Human Development. The data themselves were gathered while I was on sabbatical leave in Thailand on research fellowships from the American Council of Learned Societies and the Ford Foundation Southeast Asia Fellowship Program.

level by the intervention of these radical differences in production. The other theory invokes short-term memory. Consonants, particularly stop sounds, have rapidly changing spectra, while vowels, certainly steady-state vowels, tend to be rather long in duration providing for more time to process the stimuli with less strain on memory (Fujisaki and Kawashima, 1969).

Arguments continue over the categorical perception of phonetic segments versus continuous perception. There is still a need for experiments on types of phonetic segments not previously examined and types of segments that have been insufficiently examined. For example, experiments with voice onset time (Abramson and Lisker, 1970) yield data very similar in categoricalness to those of data with stop consonants differing in place of articulation, while experiments with distinctive vowel length in Thai (Bastian and Abramson, 1962) yield data very similar to those for steady-state vowels. In an earlier study on the identification and discrimination of phonemic tones in Thai (Abramson, 1961), a set of five fundamental-frequency contours with final points moving upward incrementally from a level base, all with the same short drop at the end, was divided perceptually by native speakers into the mid and high tones. In ABX tests, the overall level of discrimination was very high with no convincing evidence of a discrimination peak at the category boundary, thus confirming a prediction of continuous perception. On the other hand, in a recent paper, Chan *et al.* (1975) presented contradictory results showing boundary effects for a fundamental-frequency continuum that yielded two of the four tones of Mandarin Chinese, namely the rising and level tones.* The present study is an attempt to explore the matter further.

Experiments

Central Thai or Siamese has five lexical tones characterized for the most part by different fundamental-frequency contours (Abramson, 1962; Erickson, 1976). They are conventionally divided into static tones—high, mid and low—and dynamic tones—falling and rising. There appears to be no single acoustic dimension along which all five of these tones lie. In my desire to run an experiment analogous to the study of the English stop consonants [bdg] (Liberman *et al.*, 1957), I sought a continuum that would yield three tonal categories, namely the low, mid and high tones. In addition, since I had a large number of subjects available to me, I wished to have a continuum with varying degrees of acceptance on the part of native speakers of Thai for division into those three tones. I thought that I might thus have a better chance to test the perceptual effects of phoneme categories. That is, if perception of tones is categorical rather than continuous, those subjects who showed good

* A summary is found in the more accessible paper by Wang (1976).

identification functions might also show high peaks of discrimination at the category boundaries, in contrast with those who did not readily sort the stimuli into tones. I did this by ignoring the small but noticeable movements of these "static" tones as normally produced and by using level fundamental frequencies instead.

The Haskins Laboratories parallel-resonance synthesizer was used to produce the stimuli. The basic pattern was a set of steady-state formant frequencies chosen to yield a vowel acceptable as Thai long /aa/ with initial formant transitions appropriate to the velar place of articulation. Voiceless aspiration was simulated by providing a voicing lag of 80 ms filled with turbulent noise in the regions of the upper formants, with the first formant absent during the lag. The voice source was turned on at the end of the voicing lag, and one of the set of fundamental-frequency levels was imposed. 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:]. Sixteen fundamental-frequency levels ranging from 92 Hz to 152 Hz in steps of 4 Hz were imposed on the basic pattern. These 16 synthetic syllables were presented in a number of random orders to 33 native speakers of Thai for identification as one of five possible words. In fact, it was expected that only three of the choices would be used: /khàa/ (low) "galangal, a rhizome", /khaa/ (mid) "a grass" and /khâa/ (high) "to engage in trade".*

For the discrimination experiment I used the 4-interval forced-choice test of pair similarity (4IAX) which, according to Pisoni (1971), has certain advantages over other methods. For each trial the subject is presented with two pairs of stimuli such that one pair contains identical members and the other pair contains members that differ along the stimulus dimension. All possible 4IAX arrangements to provide one-step and two-step comparisons along the 16-stimulus continuum were prepared in a number of randomizations on magnetic tape. The 33 subjects who participated in the identification tests were told to choose the pair containing different members in each trial. The stimuli within each pair were separated by 250 ms, the pairs by 1 s and the trials by 3 s. After every ten trials, there was an interval of 7 s.

Results

The identification data are displayed in Fig. 1. The stimuli are arranged along the abscissa, and the percentage identification along the ordinate with the functions showing, from left to right, responses as the low, mid or high tone. Although there is extensive overlap and the peaks do not exceed 92, 74 and

* The remaining two possibilities are /khaà/ (falling) "to kill" and /khaă/ (rising) "leg".

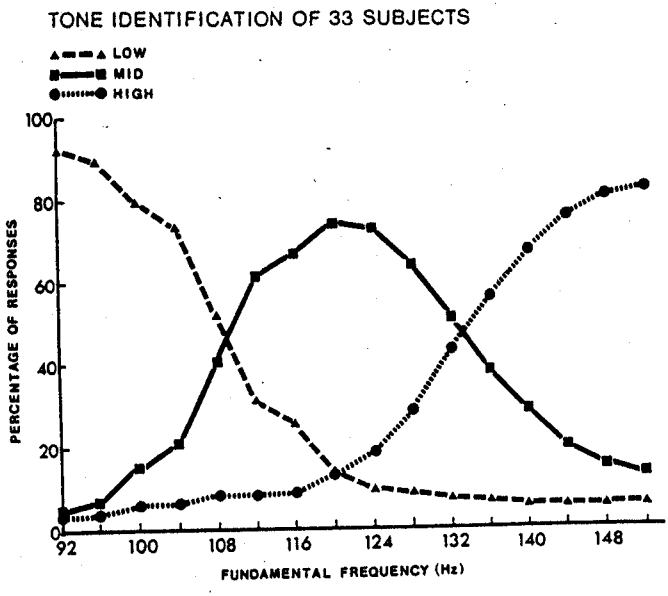

Fig. 1. Labelling of 16 F_0 levels as the low, mid and high tones by Thai subjects.

82 % respectively, the results are clearly systematic. There were 1647 responses to each stimulus. To avoid clutter, I have not plotted the infrequent and unsystematic choices of the falling and rising tones as response categories.

The 4IAX discrimination data for the 33 subjects are shown in Fig. 2. Once again, the abscissa shows the array of stimuli. The ordinate gives the percentage correct discrimination. The vertical lines at about 108 Hz and 133 Hz represent the crossover points of the identification curves in Fig. 1 that mark the boundaries between the tonal categories as indicated. There were 1024 responses to each comparison. The lower function, representing the one-step discrimination, fluctuates between 80 and 90 % with the lower part of that range favored at the high-frequency end. The two-step discrimination, represented by the solid line above, fluctuates between 92 and 96 %. Customarily, predictions of acuity of discrimination are calculated from the identification functions. To do so here, however, would have been a fruitless exercise. Both levels of discrimination are so far above 50 %, the chance level at the bottom of the graph, and so obviously unaffected by the category boundaries that there was really nothing to be demonstrated by such a calculation. Nevertheless, before coming to any final conclusion, I inspected the data more closely for signs of categoricalness.

Within my expectation that some subjects would find the dimension more acceptable than other subjects did, I looked for a subset of people with

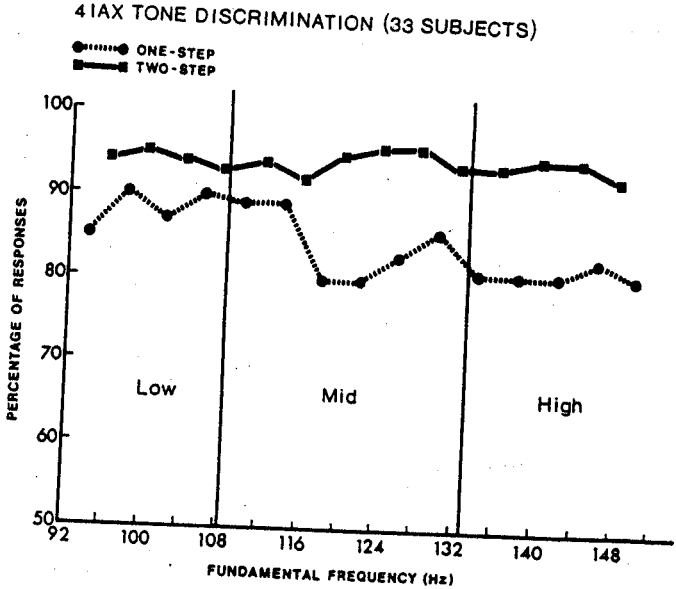

Fig. 2. Discrimination of level-F_0 tonal variants by Thai subjects.

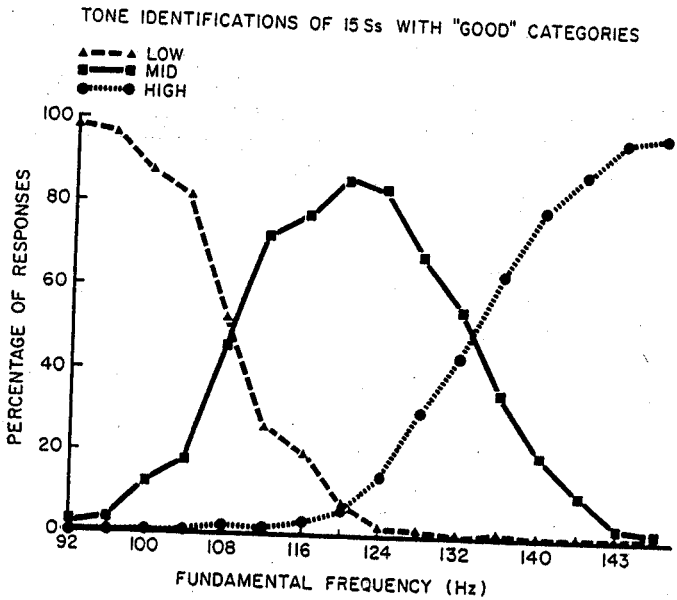

Fig. 3. Labelling of F_0 levels by Thai subjects with "good" categories reaching at least 80%.

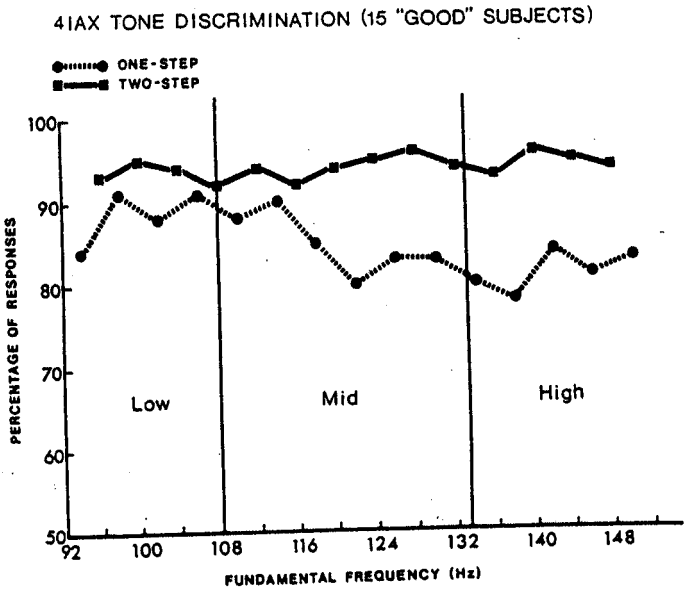

Fig. 4. Discrimination of level-F_0 tonal variants by Thai subjects with "good" categories.

especially good identification categories. I defined a "good" category as having an identification peak of at least 80 %. The identification data for the 15 "good" subjects out of the original group of 33 are displayed in Fig. 3 in which the peaks for the three tones are seen to achieve 98, 86 and 98 %, respectively. There were 771 responses to each stimulus. There is still considerable overlap, but it is less striking than before. The discrimination data for the same subjects are given in Fig. 4, with 448 responses to each comparison. These curves differ only in minor and unsystematic ways from the curves for all 33 subjects in Fig. 2. Although their data are not displayed separately, I did find five subjects who identified the tonal continuum in terms of just two categories, high and low, or mid and low. Their discrimination data did not differ significantly from any of the functions shown. Indeed, not one subject in all 33 showed boundary effects.

Discussion

The results of my experiments are consistent with the view that the perception of Thai tonal categories is not categorical.* My results of 1961, based on a

* The title of a recent study by Siegel and Siegel (1977) implies that the perception of musical intervals, at least by musicians, is categorical, just as in certain phonetic dimensions. The article itself, however, makes it clear that there was no psychoacoustic testing of hearers' ability to discriminate variants of categories.

continuum of changing fundamental-frequency shapes yielding judgments of mid and high tone are reaffirmed here by work with a continuum of flat variants yielding three tones as labelled by a much larger number of subjects, who also showed no effects of the phoneme categories in their discrimination performance. The discrepancy between Chan *et al.* and myself remains to be explained. They cite Klatt (1973), who found that auditory sensitivity is greater to differences between unchanging fundamental-frequency contours than to differences between changing contours. Ten of the 11 variants in the Chan *et al.* study were rising contours,† while the eleventh, at the upper end, was flat. This might be a partial explanation of the difference. Differences between Mandarin Chinese and Thai may be relevant. Another difference is that Chan *et al.* synthesized the isolated vowel [i] to yield two tonally differentiated words, while I used the somewhat more complex syllable [kha:] to yield three words. In any event, I have presented a continuum of rising and falling ramps elsewhere (Abramson, 1978) on the same syllable type that Thai subjects divided into the three tones of the present study. Perhaps a discrimination experiment with those variants will help in reconciling the two studies.

References

Abramson, A. S. (1961). Identification and discrimination of phonetic tones, *J. Acoust. Soc. Am.* 33, 842(A).

Abramson, A. S. (1962). "The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments." Indiana U. Res. Center in Anthropology, Folklore and Linguistics, Pub. 20, Bloomington, Ind.

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

Abramson, A. S. and Lisker, L. (1970). Discriminability along the voicing continuum: cross-language tests. Proc. 6 Int. Congr. Phon. Sci., Prague, 1967. Academia, Prague, pp. 569–573.

Bastian, J. B. and Abramson, A. S. (1962). Identification and discrimination of phonemic vowel duration. *J. Acoust. Soc. Am.* **34**, 743–744(A).

Chan, S. W., Chuan, C. K. and Wang, W. S-Y. (1975). Cross-language study of categorical perception for lexical tone. *J. Acoust. Soc. Am.* 58, S119(A).

Erickson, D. M. (1976). "A Physiological Analysis of the Tones of Thai." Ph.D. Dissertation, The University of Connecticut.

Fry, D. B., Abramson, A. S., Eimas, P. D. and Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Lang. Speech* 5, 171–189.

Fujisaki, H. and Kawashima, T. (1969). "On the Modes and Mechanisms of Speech Perception." Annual Report of the Engineering Research Institute (Univ. of Tokyo) No. 28, pp. 67–73.

† In fact, their rising variants were a bit more complex in pattern. They remained flat at the starting F_0 for 100 ms and then rose to the final frequency value.

Klatt, D. H. (1973). Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception. *J. Acoust. Soc. Am.* **53**, 8–16.
Liberman, A. M., Cooper, F. S., Shankweiler, D. S. and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* **74**, 431–461.
Liberman, A. M., Harris, K. S., Hoffman, H. S. and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. exp. Psychol.* **54**, 358–368.
Pisoni, D. B. (1971). "On the Nature of Categorical Perception of Speech Sounds." Ph.D. Dissertation, The University of Michigan.
Siegel, A. and Siegel, W. (1977). Categorical Perception of Tonal Intervals: Musicians can't tell *sharp* from *flat*. *Percep. Psychophys.* **21**, 399–407.
Wang, W. S-Y. (1976). Language change. *Ann. N.Y. Acad. Sci.* **280**, 61–72.