

Linguistic relativism and selective adaptation for speech: A comparative study of English and Thai

KAREN G. FOREIT

University of Texas at Arlington, Arlington, Texas 76019

The effect of repeated exposure to prevoiced and voiced stimuli on the voiced/voiceless category boundary was compared for monolingual native English speakers and bilingual native speakers of Thai. The English speakers, for whom the two adaptors were members of the same phoneme class, showed equivalent shifts towards the adapting stimulus under both conditions. The Thai subjects, who perceived the two adaptors as belonging to two distinct phoneme classes, showed a shift of the voiced/voiceless boundary only for the voiced adaptation condition; exposure to the prevoiced adaptor had no effect.

Since Eimas and Corbit (1973) first demonstrated selective adaptation for the linguistic feature of voicing, their technique has been used with other categories and classes of speech (Cooper & Blumstein, 1974; Eimas, Cooper, & Corbit, 1973; Morse, Kass, & Turkienicz, 1976). The procedure is simple: repetition of a token taken from one end of a linguistic continuum leads to a shift in the category boundary, towards the adapting stimulus. In the original demonstration, subjects adapted with a /ba/ found stimuli previously lying on the /ba-pa/ boundary to be more /pa/-like, while those exposed to /pa/ reported just the reverse.

The first studies of selective adaptation used as adaptors stimuli chosen from the extreme ends of the continuum, stimuli which could be considered "good" exemplars of their phoneme class. It was, therefore, of both empirical and theoretical interest to test the effect of adaptors closer to the category boundary, which, although they were not as "good" exemplars acoustically, would still be unambiguously perceived as belonging to one or the other phoneme class. Using the voice onset time (VOT) continuum, Anderson (1976) and Miller (Note 1) found that stimuli close to the preadaptation category boundary produced less adaptation than more extreme stimuli. Miller (Note 1) interpreted these results to support the existence of feature detectors selectively tuned to overlapping distributions of VOT values. The closer the stimulus to the optimal values, the more strongly the feature detector fires, and the greater the

fatigue induced by that firing. Stimuli closer to the category boundary induce weaker firing and consequently weaker adaptation effects.

An alternative explanation for these data might be that the adapting stimulus acts as an "anchor" for the subject's perceptual judgment (Helson, 1964), so that stimuli farther from the category boundary exert a greater pull than stimuli closer to the boundary, hence the differential adaptation effects. The anchor hypothesis would predict that if moving the adapting stimulus closer to the category boundary produces less adaptation, then the reverse should also obtain: moving away from the boundary should increase the adaptation effect. A limiting factor to this prediction would be moving across a second phonemic boundary. In the latter case, no adaptation would be expected: using the feature of place of articulation, Cooper (1974) found that adapting with /bae/ shifted the adjacent /bae-dae/ boundary but did not affect the "far" /dae-gae/ boundary: similarly, adapting with /gae/ shifted the /dae-gae/ but not the /bae-dae/ boundary.

The prediction of the feature detector model in the case of stimuli farther from the category boundary depends on assumptions made about the shape of the sensitivity distributions of the detectors. If feature detectors which induce a dichotomous perception of the acoustic continuum show normally distributed, symmetrical sensitivity around their optimal stimulus value, then stimuli beyond the optimal region should produce less adaptation than stimuli in the optimal range, despite the greater distance from the category boundary of the former. On the other hand, if the sensitivity distribution is not symmetrical but plateaus at the optimal stimulus value instead, then moving beyond the optimal value should not change the adaptation effect. In cases where the continuum is categorized into three or more phonemic classes, the sensitivity distributions of the interior detectors would be presumed to be symmetrical, while the

This research was funded by Grant R03 MH 26168 from NIMH and FRAP 11184 from the Research Foundation of the City University of New York. The author wishes to thank Drs. F. S. Cooper and A. M. Liberman for making available the facilities of Haskins Laboratories, Ralph Bianca for gathering the data, and the Thai Student Association of Columbia University for providing volunteers. Above all, thanks are due to the Thai students themselves, who endured not only the experimental sessions, but a 3-h round-trip subway ride as well.

shape of the detectors lying at the extreme ends would still be open to question.

In Anderson's and Miller's studies, acoustic manipulation was accomplished while the phonemics remained constant. The complement of their procedure would be to vary the phonemic value of the adapting stimulus while keeping the acoustic value constant. Such a design is impracticable for a single speaker, but it comes possible if listeners of different language backgrounds are compared. What is needed are stimuli which are allophonic in one language but which represent different phonemes in another language. This condition is precisely met by comparing the VOT continuum in English and Thai.

The VOT continuum is determined by the onset of voicing (vocal cord vibrations which produce periodic energy) relative to the release of sound. When voicing precedes the sound release, the segment is said to be prevoiced, and when voicing and release are simultaneous the segment is called voiced. If voicing is delayed relative to the release of sound, the segment is called unvoiced or voiceless. English distinguishes between only the voiced and voiceless categories (as in the distinction between /ba/ and /pa/), which means that voiced and prevoiced syllables are allophonic to the monolingual English speaker. Thai, on the other hand, distinguishes among all three categories—prevoiced, voiced, and voiceless (Lisker & Abramson, 1970).¹

The research design is now straightforward. The effect of prevoiced and voiced adaptors on the voiced/voiceless boundary will be compared for English and Thai speakers. Since the adapting stimuli are allophonic for English speakers, the anchor hypothesis predicts that the adaptor farthest from the category boundary—the prevoiced stimulus—should produce the greatest adaptation effect. The feature detector hypothesis predicts either equivalent or reduced adaptation following the prevoiced stimulus. Both models make the same prediction for Thai speakers: since the two adaptors represent different phonemic classes, only the adjacent category—the voiced stimulus—should produce adaptation. The prevoiced stimulus represents a "far" category and should be ineffective.

METHOD

Subjects

Five monolingual native speakers of English and six English-Thai bilingual native speakers of Thai served as paid subjects. The Thai speakers had been living in the United States a mean length of 2.7 years (range 5 months-6 years) and had first studied English in Thailand an average of 13 years (range 6-24 years) prior to the study. Four of the English speakers had had previous exposure to synthetic speech, but none of the subjects had ever participated in any study of speech labeling or selective adaptation.

Materials

A 12-item prevoiced-to-voiced continuum and a 10-item voiced-to-voiceless continuum were synthesized on the parallel resonance

synthesizer at the Haskins Laboratories from a program developed by Lisker and Abramson (1970). The stimuli, 450 msec in total duration, consisted of a variable-length prevoiced segment followed by a 45-msec formant transition, and ending with a variable-length three-formant steady-state vowel. VOT was defined as the onset of the first formant relative to the onset of the second and third formants. Prevoicing was accomplished by exciting the first formant with a 154-Hz steady-state periodic source prior to the onset of the second and third formants. The transition starting frequencies for the first, second, and third formants were 154, 543, and 1,190 Hz, respectively, terminating at 769, 1,232, and 2,525 Hz. The voicing lag was accomplished by cutting back the first formant and exciting the second and third formants by a noise source during the time the first formant was absent.

The prevoiced-to-voiced continuum progressed from -60 msec VOT to 0 msec in 5-msec steps, except for the first two tokens, which differed by 10 msec. The voiced-to-voiceless continuum went from 10 msec VOT lag to 60 msec in 5-msec steps, except that the last two tokens differed by 10 msec.

These two continua served to establish the prevoiced/voiced category boundary for the Thai speakers and the voiced/voiceless boundary for all speakers. In addition, the -60-msec VOT and 0-msec VOT stimuli were selected as adapting stimuli. Idealized spectrograms for the -60-, 0-, and 60-msec VOT stimuli are shown in Figure 1.

Procedure

The adapting stimuli and identification sequences were recorded on a single channel of a Crown SP822 tape deck at the Haskins Laboratories, and diotically presented to the subjects from two Crown tape decks via a TEAC mixer and Koss Pro 4AA headphones. The mixer arrangement allowed the adaptation loop to play continuously; switching between the adaptation loop and the

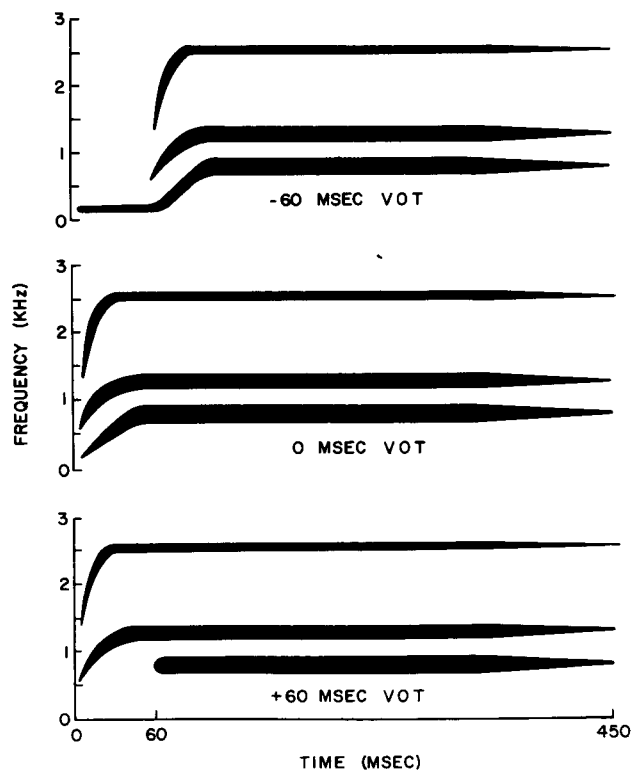


Figure 1. Idealized spectrograms of the prevoiced -60 msec VOT and the voiced 0 msec VOT adaptors, plus the voiceless 60 msec VOT stimulus.

test sequences was done manually by the experimenter, who also recorded the subject's responses.

The subjects were individually tested in a quiet room. All instructions were given in English; the response keys were labeled with the letters "B" and "P" for the English speakers, and the appropriate Thai letters for the Thai subjects (prior to their first session, each Thai speaker was asked to write out the letters corresponding to the -60-, 0-, and 60-msec VOT tokens; these labels were then used on the response console so that each subject was responding to his/her own handwriting). The subjects were encouraged to listen to the stimuli as speech sounds in their native language.

Preadaptation. The purpose of the experiment was briefly explained at the start of the first session and the voiced/voiceless continuum played in order of increasing VOT. The subjects were instructed to label each sound by pressing the appropriate button. Ten tokens of each of the 10 stimuli were then presented randomly at 5-sec intervals. The Thai speakers similarly labeled the 12-item prevoiced/voiceless continuum prior to their second session.

Adaptation. The two adaptation sessions, held at least 24 h apart, followed the procedure of Eimas and Corbit (1973). The adaptor was presented at the rate of 1/sec on a continuous loop. Following 2 min of adaptation, the subjects alternated 1 min of adaptation and four randomly selected identification trials. The identification trials were signaled by a brief 500-Hz tone burst; 5 sec separated each trial, and adaptation resumed within 10 sec after the last identification trial. The subjects were allowed to draw or play cards during the adaptation loops and were given two rest periods during the session. After each rest period, they were again given 3 min of adaptation prior to the first identification trial. In all, the subjects labeled 12 tokens of each stimulus. Order of adaptation conditions (-60 or 0 msec VOT) was counter-balanced across subjects.

RESULTS AND DISCUSSION

Category boundaries for each subject were calculated using the method of least squares (Eimas & Corbit, 1973). The data from two Thai speakers were discarded for failure to hear the stimuli as speech sounds; they failed to show consistent categorization of both the prevoiced/voiced and the voiced/voiceless continua. The mean category boundary for the prevoiced continuum for the four remaining Thai speakers was -21 msec VOT, which agrees with the boundary of -20 msec VOT found by Lisker and Abramson (1970). The prevoiced identification of the Thai speakers clearly establishes that they perceived the -60- and 0-msec VOT adap-

tors as different phonemes. The English speakers expressed no awareness of a difference between the adapting stimuli. Therefore, it is appropriate to compare the effects of the two adaptors, in terms of both their acoustic and phonemic differences.

The performances of the individual subjects are presented in Table 1, which lists the voiced/voiceless category boundaries in the preadaptation situation and following -60- and 0-msec VOT adaptation. A two-way analysis of variance comparing boundary shifts following prevoiced and voiced adaptation for the English and Thai speakers showed a significant main effect of adaptor type [$F(1,7) = 10.38, p < .05$] and a significant interaction between adaptor type and language experience [$F(1,7) = 33.22, p < .01$]. Exposure to the -60-msec VOT stimulus produced less adaptation than exposure to the 0-msec stimulus, because the Thai speakers as a group showed no significant shift in that condition. All four Thai speakers showed a larger shift toward the adapting stimulus following 0-msec VOT adaptation than following -60-msec VOT adaptation; a follow-up *t* test revealed that the resulting mean difference was statistically reliable [$t(3) = 8.06, p < .01$]. The performance of a typical Thai speaker is presented graphically on the upper half of Figure 2. The English speakers, on the other hand, showed category shifts under both adaptation conditions. Three of the five showed larger shifts following -60-msec adaptation than following 0-msec adaptation, but this difference was not statistically reliable [$t(4) 1.05, p > .30$]. The performance of an English speaker is presented on the lower half of Figure 2.

In summary, the Thai and English speakers showed quite different reactions to the voiced and prevoiced adaptors. The Thai speakers exhibited a significantly greater category shift following the voiced than following the prevoiced adaptation, while the English speakers exhibited essentially the same response to both.

These results clearly indicate that the effects of acoustic manipulations are constrained by linguistically determined phoneme boundaries. Thus the

Table 1
Voiced/Voiceless (ba/pa) Category Boundary in Milliseconds VOT for English and Thai Speakers Before Adaptation and Following -60 msec VOT and 0 msec VOT Adaptation

| Subject | English Speakers | | | Thai Speakers | | |
|------------|------------------|----------------|--------|---------------|----------------|--------|
| | Preadaptation | Postadaptation | | Preadaptation | Postadaptation | |
| | | -60 msec | 0 msec | | -60 msec | 0 msec |
| 1 | 25.0 | 22.0 | 22.0 | 30.0 | 26.9 | 20.0 |
| 2 | 31.3 | 23.2 | 23.5 | 41.4 | 42.1 | 35.0 |
| 3 | 31.6 | 22.8 | 24.1 | 26.0 | 21.6 | 17.5 |
| 4 | 36.4 | 34.0 | 33.6 | 37.0 | 37.0 | 31.9 |
| 5 | 25.0 | 15.4 | 20.0 | | | |
| Mean Shift | | 6.38 | 5.22 | | 1.70 | 7.50 |
| SD | | 3.40 | 2.38 | | 2.44 | 2.17 |

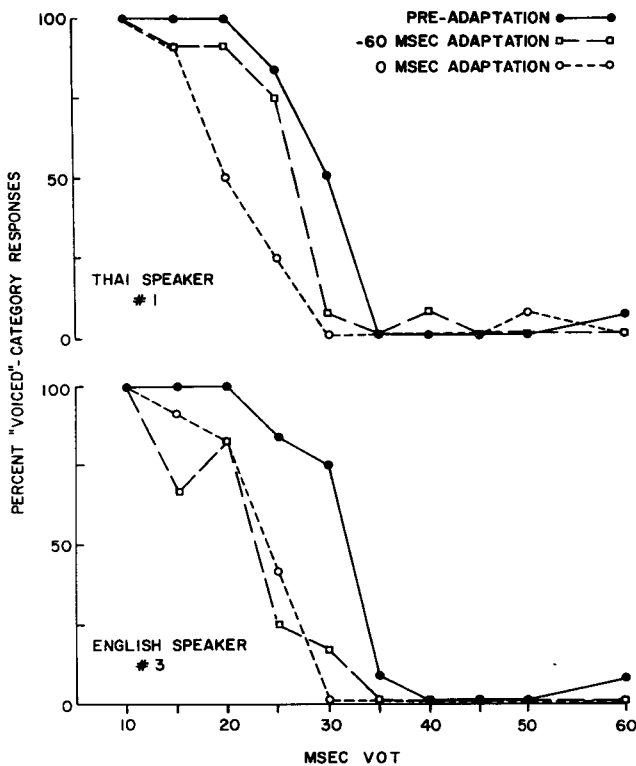


Figure 2. Percentages of "voiced" category responses on the voiced/voiceless continuum for an individual Thai-speaking and an individual English-speaking subject, prior to adaptation and following adaptation with the -60 msec VOT and 0 msec VOT adaptors.

English speakers, for whom a 0 -msec VOT stimulus represents a "good" exemplar of the category /ba/, continue to show adaptation effects at the same level following exposure to the -60 -msec VOT stimulus, which they also perceive as /ba/. For the Thai speakers, however, moving from 0 -msec to -60 -msec VOT means crossing a second category boundary. Exposure to the 0 -msec adaptor, a "good" voiced category token, effectively shifted the adjacent voiced/voiceless category boundary, while adaptation with the farther-removed -60 -msec VOT stimulus, a member of the prevoiced category, had little effect on the nonadjacent voiced/voiceless boundary.

The simple anchor model received no support from the present data, since the English speakers showed no greater effects following adaptation with prevoiced than with voiced adaptors. Assuming that linguistically tuned feature detectors underlie the selective adaptation effect, we may then speculate on possible differences in the feature detector systems for speakers of English and Thai. The Thai case is straightforward: three feature detectors, at the least the middle one of which shows a symmetrical sensitivity distribution, divide the VOT continuum into three categories. Adaptation of the voiced/voiceless bound-

ary requires fatigue of the voiced category detector, which was unaffected by the prevoiced adaptor.

The case of the English speakers, who showed equivalent adaptation of the voiced/voiceless boundary by both the prevoiced and the voiced adaptors, is theoretically more complex. At least two possibilities present themselves: (1) that English speakers lack the prevoiced feature detector, or (2) that English speakers possess a prevoiced detector, but, since English does not make a prevoiced/voiced distinction, they interpret its output as signaling voicing information. If the former is true, that English speakers lack the prevoiced detector, they must have lost it, possibly through lack of stimulation by the appropriate contrasts. Studies of infants growing up in monolingual environments (Eimas, 1975; Lasky, Syrdal-Lasky, & Klein, 1975; Streeter, 1976) have shown that infants are sensitive to voicing contrasts that their parents do not use and are often unable to discriminate. Although the infants' perception seems very nearly categorical, their category boundaries often lie at different points along the continuum from the adult speakers in their linguistic environment. Presumably, the sensitivity of the feature detectors may be modified by experience or even lost altogether, as seems to be the case with certain feature detectors in the visual system (cf. Stryker & Sherik, 1975).

Alternatively, it is possible that English speakers possess all three VOT detectors but interpret the output of the prevoiced detector as voiced information. Preservation of unused detectors would greatly simplify second-language learning. Streeter and Landauer (Note 2) have reported that formal English training improved the voiced-voiceless discrimination ability of Kenyan speakers whose native language, Kikuyu, does not include a voiced/voiceless contrast. Furthermore, combining the output of the prevoiced and voiced detectors, each with a symmetrical sensitivity distribution, into a single voiced percept would produce an asymmetrical sensitivity distribution to voicing within the range of VOT values used in the present study, which reaches asymptote at the optimal stimulus value for the voiced detector. The present data cannot discriminate between the asymmetrical sensitivity hypothesis, which implies that English speakers have lost the prevoiced detector, and the model of combined detectors with symmetrical sensitivity distributions. Since the hookup between the two detectors must be made at an automatic, if not neural, level to show equivalent adaptation throughout the prevoiced range, it is difficult to imagine an empirical test which could choose between the two alternatives.

In an independent study conducted at the same time as the present research, Donald (Note 3) reported virtually identical results: English speakers showed reliable shifts in their voiced/voiceless category boundary following exposure to both -80 -msec VOT and $+5$ -msec adaptors, while Thai subjects produced

reliable shifts following +5-msec adaptation alone. Taken together, these two studies conclusively demonstrate that the effects of acoustic manipulations on selective adaptation are strongly limited by their linguistic implications.

REFERENCE NOTES

1. Miller, J. L. *Properties of feature detectors for VOT*. Paper presented at Acoustical Society of America, Washington, D.C., April 1976.

2. Streeter, L. A., & Landauer, T. K. *The effects of learning English as a second language on the acquisition of a new phonemic contrast*. Paper presented at Acoustical Society of America, Austin, Texas, April 1975.

3. Donald, L. The effects of selective adaptation on voicing in Thai and English. Haskins Laboratories *Status Report on Speech Research*, 1976, SR-47, 129-136.

REFERENCES

- ANDERSON, F. *Some implications for the operation of feature detectors in speech perception: Use of identification response time as a converging operation*. Unpublished doctoral dissertation, Brown University, 1976.
- COOPER, W. E. Adaptation of phonetic feature analyzer for place of articulation. *Journal of the Acoustical Society of America*, 1974, **56**, 617-627.
- COOPER, W. E., & BLUMSTEIN, S. E. A labial feature analyzer in speech perception. *Perception & Psychophysics*, 1974, **15**, 591-600.
- EIMAS, P. D. Speech perception in early infancy. In L. B. Cohen & P. Salapatek (Eds.), *Infant perception*. New York: Academic Press, 1975.
- EIMAS, P. D., COOPER, W. E., & CORBIT, J. D. Some properties of linguistic feature detectors. *Perception & Psychophysics*, 1973, **13**, 247-252.
- EIMAS, P. D., & CORBIT, J. D. Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 1973, **4**, 99-109.
- HELSON, H. *Adaptation-level theory: An experimental and systematic approach to behavior*. New York: Harper, 1964.
- LASKY, R. E., SYRDAL-LASKY, A., & KLEIN, R. E. VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, 1975, **20**, 213-225.
- LISKER, L., & ABRAMSON, A. S. The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences*, Prague, 1967. Prague: Academia, 1970.
- MORSE, P. A., KASS, J. E., & TURKIENICZ, P. Selective adaptation of vowels. *Perception & Psychophysics*, 1976, **19**, 137-143.
- STREETER, L. A. Language perception of two-month-old infants shows effects of both innate mechanisms and experience. *Nature*, 1976, **259**, 38-41.
- STRYKER, M. P., & SHERK, H. Modification of cortical orientation selectivity in the cat by restricted visual experience: A reexamination. *Science*, 1975, **190**, 904-906.

NOTE

1. These working definitions are meant to delineate the three voicing categories in a word-initial prevocalic position. The English categories referred to as voiced and voiceless correspond to the IPA conventions voiced and voiceless aspirated; the Thai categories referred to as prevoiced, voiced, and voiceless correspond to prevoiced, voiceless unaspirated, and voiceless aspirated.

(Received for publication October 15, 1976;
revision accepted January 18, 1977.)