

A cross-language study of range of voice onset time in the perception of initial stop voicing

Patricia A. Keating

36-521, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Michael J. Mikoś

Department of Slavic Languages, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

William F. Ganong, III

Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104
(Received 30 March 1981; accepted for publication 8 July 1981)

A series of experiments was carried out to compare the extent of range effects in the phonetic categorization of voice onset time (VOT) by speakers of Polish and of English, two languages which contrast different VOT categories. Results indicate that Poles are more prone to range effects than are Americans. For acoustic continua with appreciable numbers of prevoiced stimuli, monolingual Polish speakers' perceptual boundaries fall in the gap between their production categories. For ranges of VOT which include few prevoiced stimuli, their boundaries are substantially shifted. Americans show no shifts of this type, although they do show some small shifts. It was determined that the much smaller shifts shown by the American subjects were not due to expectations about the test. Results are interpreted in terms of the different VOT contrasts involved: their spacing along the VOT continuum, and their psychophysical basis.

PACS numbers: 43.70.Dn, 43.70.Ve

INTRODUCTION

Voice onset time, or VOT, the time between the release of a stop occlusion and the onset of glottal vibration, characterizes the voicing contrast for initial stop consonants in most languages. Lisker and Abramson (1964) proposed that languages which have stop voicing contrasts have chosen among three VOT categories. The categories are (1) voicing lead (with negative VOT values, also called prevoicing); (2) coincident and short-lag VOT (with zero or low positive VOT values); and (3) long-lag VOT (with high positive VOT values). The perception of VOT in English, which contrasts long-lag VOT with short-lag and prevoiced categories, has been studied intensively. Many languages, however, contrast lead with lag VOT, and further work on the perception of VOT in such languages would be of interest.

According to classical phonetics, Polish contrasts prevoiced stops with voiceless unaspirated or slightly aspirated stops, which corresponds to a contrast of voicing lead with short-lag VOT. In some pilot work, we therefore tested labeling of VOT by native Polish listeners residing or visiting in the United States. These subjects produced labeling boundaries like those which had been found for American listeners. This result was rather disconcerting, because acoustic measurement confirmed that the VOT categories of their Polish speech are quite different from those of English. In an early study (Mikoś *et al.*, 1978) we tried to encourage a boundary more appropriate to Polish by varying the range of VOT stimuli used in labeling tests. A range that was more representative of the VOT values actually used in Polish elicited a boundary more appropriate to the Polish production categories than did the Englishlike range of values that had been used before. This result was interesting, especially in light of other

work on stimulus range in speech perception (see below), but some caution was required at that point. All of this pilot work was done in the United States, with native Polish speakers who either spoke some English or were exposed to it constantly. We could not distinguish effects of bilingualism and language contact from effects of perceptual range *per se*. In this paper we report on a set of experiments conducted in Poland to resolve this question. We will show that Polish listeners are indeed susceptible to quite large range effects, and that American listeners categorize the same stimuli with much smaller range effects, if any.

By "range effect" we mean a difference in subjects' performance corresponding to a difference in the range of stimuli presented in, for example, a categorization task. The assumption is that, all things being equal, subjects will prefer to split the range into two equal halves. That is, subjects are assumed to make an implicit comparison of each test item with the set of test items, and to assign it to a response category relative to that set. An alternative to range effects would be "absolute perception," in which perception of a particular stimulus does not depend on the set of items it is presented with. Absolute perception is typically thought to characterize the perception of certain consonants, such as the voiced and voiceless stops.

However, it appears that consonant perception does depend on the set of stimuli used in a test. Contrast effects (Diehl *et al.*, 1978), frequency effects (Simon and Studdert-Kennedy, 1978), and range effects (Brady and Darwin, 1978; Rosen, 1979) have all been found. Of particular relevance here, Brady and Darwin (1978) looked at range effects in the perception of VOT. They used a continuum with a range from +5 to +55 ms VOT in 5-ms steps, and a series of five 20-ms subranges, e.g., 5-25 and 35-55 ms VOT. The small ranges and

small step sizes maximize the number of stimuli presented near the category boundary. Listeners' responses for the entire range were compared to those for the five subranges, and differed significantly. No boundary values were computed, but inspection of their figures suggests that average differences between ranges varied from 0 to about 7 ms VOT. Though reliable, such response shifts are small, on the order of typical selective adaptation effects.¹

In this type of study, stimulus range is used as an arbitrary experimental variable, as in psychophysical studies. Our approach is quite different: VOT range is viewed as a property of spoken languages and the categories they use in voicing contrasts. Our interest is not so much whether category boundaries can be slightly shifted around experimentally, but whether people can be induced to show perceptual categories that are inappropriate for their language. We addressed this question by comparing categorizations of VOT in Polish and English, which have different voicing contrasts. From acoustic measurements of natural speech, an overall range of VOT values for each language was determined. Test continua corresponding to each linguistic range were made, and listeners from each language were then tested on *both* ranges. Thus listeners always heard two linguistic voicing categories, but not necessarily their own.

One possible outcome of such an experiment is that perception would be absolute, i.e., the range of VOT used would be irrelevant, and listeners would show the same boundary (consistent with their production data) on both VOT ranges. This result would argue for a constant internal standard and would agree with the earlier results of Sawusch and Pisoni (1973). Another possible outcome is that listeners perform as expected only on their own language's range of VOT, but show an anomalous boundary when tested on the unfamiliar range. Such a range effect would suggest that phonetic categories are relative.

In this study we compare these predictions with results from experiments on perception in Polish and English. Before we describe these experiments, however, we describe the VOT categories used in the two languages.

I. PRELIMINARY TESTS OF ENGLISH AND POLISH VOICING CATEGORIES

In this section, two methods are used to determine the VOT categories of Polish and English: acoustic measurements of natural speech tokens, and discrimination. In later sections we will relate data on the labeling of VOT to the categories determined in these two independent ways.

A. Production measurements

The voicing contrast used in Polish for syllable-initial stops /bdg/ and /ptk/ is one between prevoiced and voiceless unaspirated (or slightly aspirated) stops. Polish allows a surface voicing contrast both word initially and medially. Prepausal stops are always voiceless, but word-final stops before an initial sonorant in

a following word may be either voiced or voiceless. Word-final stops before an initial obstruent in a following word are subject to cluster voicing-assimilation: obstruents in a cluster, whether word-internal or phrase-internal, must agree in voicing (Mikoś, 1977). English also contrasts a set of stops /bdg/ with a set /ptk/, in almost all environments. However, the stops are not phonetically identical to those of Polish, since in post-pausal position the voiced stops are often voiceless unaspirated, and the voiceless stops are usually aspirated.

VOT measurements were made for a corpus of stop consonants in each language. In Polish, VOT was measured for 42 disyllabic words containing all phonologically legal sequences of a stop consonant [bdgptk] followed by a vowel [ieəaəou] in initial position. For each labial stop, all eight C + vowel combinations occur; for each dental stop, only seven combinations occur, and for each velar stop, only six combinations occur. (We did not consider the palatalized allophones of /t/ and /d/, which occur before [i], nor did we consider those other cases of (underlying) /t/ and /d/ which do not surface as [t] or [d] (cf., Mikoś, 1977). In Polish, stress almost always falls on the penultimate syllable, so each first syllable of these disyllables was stressed. Five monolingual speakers in Łódź, Poland, recorded a list of these 42 words, ten times each, in a local radio station. The VOT of each initial stop was measured from a computer-implemented waveform display.

A few tokens were not measured because of omissions, phonetic substitutions, noise, or lack of a clear burst. In general these problems are minimal in list-type readings of absolute-initial stops, but some speakers at times tended to run words in the lists together, making measurement more difficult. In the data reported here, prevoicing was measured only for absolute-initial (post-pausal) stops. The number of measurements actually obtained for each stop was: [b], 340 out of 400; [d], 309 out of 350; [g], 259 out of 300; [p], 378 out of 400; [t], 338 out of 350; and [k], 282 out of 300.

The results of this acoustic analysis are summarized in Table I and illustrated in Fig. 1(a). The analysis confirms that Polish uses prevoicing for its voiced stops and short-lag VOT (voiceless unaspirated) for its voiceless stops, except that the [k] tokens have higher VOT's than we might expect. It can be seen that the productions of [k] are also more variable than

TABLE I. Summary data for measured VOT values of Polish stop consonants, with number of tokens, mean, and standard deviation. Means and standard deviations are in ms VOT.

Consonant	N	Mean	Standard deviation
b	340	-88.2	40.3
d	309	-89.9	33.1
g	259	-66.1	38.4
p	378	+21.5	10.2
t	338	+27.9	8.8
k	282	+52.7	20.0

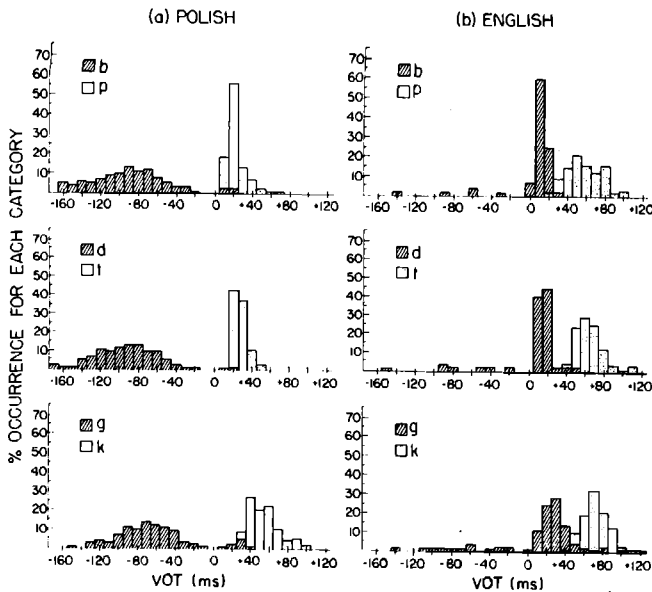


FIG. 1. (a) Distribution of measured VOT values for the six initial stops of Polish; (b) distribution of measured VOT values for the six initial stops of English. Summary statistics are given in Table I.

those of [p] and [t]. Otherwise, the well-known effect of place of articulation on VOT (Lisker and Abramson, 1964) is seen here for short-lag VOT (cf. Keating *et al.*, 1980), although not for prevoiced stops in this sample.

For the purpose of the comparisons we will be making in this study, we also present the distribution of VOT's measured just in initial [da-] and [ta-] tokens, those from this experiment, plus similar tokens from a different set of speakers in Wrocław, Poland (from Keating, 1979). A histogram showing measured VOT values for these tokens is shown in Fig. 2.

There is essentially no category overlap between voiced and voiceless stops, the few exceptions being cases of intrusion of voiced stops into the voiceless region (i.e., a failure to prevoice). Even if we combine values for all three places of articulation, the voiced and voiceless categories are well separated, and the separation is enhanced by a region from about -30 to +5 ms VOT, where few tokens are produced. Because the voicing contrast is between voicing lead and voicing lag, the only information needed to identify a stop as voiced or voiceless is whether the VOT value

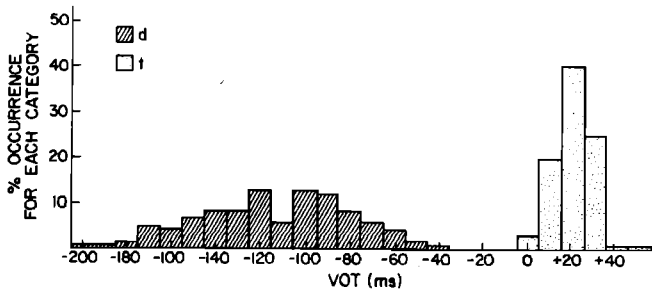


FIG. 2. Distribution of measured VOT values for those Polish tokens beginning with [da-] (152 tokens) and [ta-] (138 tokens).

is positive or negative, not its numerical value. This is not the case for English, where, as can be seen below, the two categories lie much closer together along the VOT continuum entirely on the positive side, and much more precise information is required for phonetic categorization. The gap between the two categories also allows room for variation in the category boundary. Since Polish children presumably do not hear VOT's around 0 ms (given our production data for Polish), they might be unsure of exactly where to place a category boundary.

B. Discrimination data

Discrimination of VOT was also assessed to determine the voicing categories in Polish. The stimuli were modified from a Lisker and Abramson series as described in the Appendix. Stimuli from -70 to +70 ms VOT in 10-ms steps were selected, and compared in a same-different (AX) format. Twelve repetitions of each pair were recorded in random order in blocks of ten. (Two stimuli for each VOT value were synthesized, and so there were actually six repetitions of each of the two pairs made. Differences between the two tokens of each VOT value were small and will not be discussed here.) The InterStimulus Interval (ISI) was 200 ms, the InterPair Interval (IPI) was 4 s, and the InterBlock Interval (IBI) was 6 s. Some pairs differed by 20 ms (two-steps): -70/-50, -50/-30, -30/-10, -20/0, -10/+10, 0/+20, +10/+30, +30/+50, +50/+70; others by 30 ms (three-step): all pairs from -70/-40 through +40/+70. Eight monolingual Poles in Łódź, Poland served as listeners. They responded "tak" (yes) or "nie" (no) to indicate whether or not the members of a pair were the same. Twelve repetitions of each of the 10 two-step and 12 three-step pairs, and of eight "same" pairs, were presented in random order.

The results of this discrimination test are shown in Fig. 3, where average responses for the eight listeners are given for both the two-step and three-step comparisons. There is some bias toward "same" responses, but generally discrimination is heightened from about -15 to +25 ms VOT. Peak discrimination occurs at 0 ms VOT for the two-step and at +5 ms VOT for the three-step comparisons. While there is a broad region of discriminability, there is no evidence for a

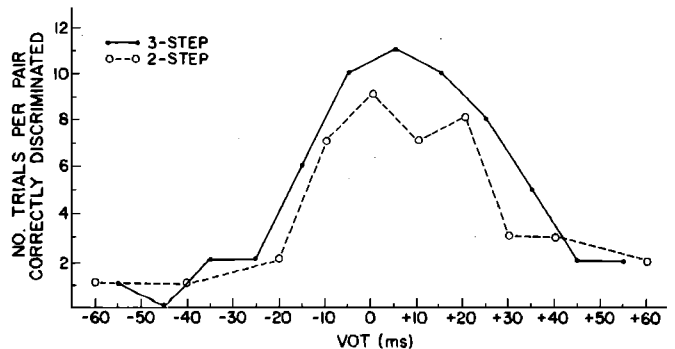


FIG. 3. Averaged discrimination function for eight Polish listeners.

particular peak corresponding to the usual English one, nor for one in between lead VOT stimuli, e.g., at -20 ms VOT.

Thus the data from production and discrimination independently point to a Polish category boundary at about 0 ms VOT. In addition, the production data indicate that the two categories are separated by a region of the VOT continuum, mostly lead VOT values. The discrimination data indicate that discrimination is heightened over a similar, but not identical, region of the VOT continuum, largely short-lag VOT values. That is, while the optimal boundary would seem to be at 0 ms VOT, there would be some variation within this region in the actual location of the boundary.

C. Comparisons with English

A great deal of data has accumulated on the production and perception of VOT in English, for example, Lisker and Abramson (1964), (1967), and (1970); Abramson and Lisker (1970); Zlatin (1974); Weismer (1979). For purposes of comparison with the Polish production data, Fig. 1(b) shows a distribution of VOT from our corpus of tokens of English /bdgptk/. Like the Polish tokens, these stops were word-initial in real-word disyllables with stress on the first syllable, before each of 12 vowels. Data are from two American speakers who also participated in experiment 4. Data on discrimination were also obtained from six American listeners using the same stimulus tapes and procedures described for the Polish discrimination test. When results from two-step and three-step comparisons are combined, there is a region of heightened discrimination from about +5 to +45 ms VOT, especially +20 to +30 ms. This region is about as broad as the Polish one, but it is to the right of the Polish one along the VOT continuum. Taken together, the production and discrimination data suggest a boundary at about +30 to +35 ms VOT, similar to findings of previous studies.

Thus the experimental evidence demonstrates that Polish and English use different VOT categories in their voicing distinctions, and have correspondingly different peaks in discrimination. The first experiment of the present study assesses Polish phonetic categorization directly.

II. EXPERIMENT 1

In this experiment we compare the performance of Polish and American listeners in phonetic categorizations of stimuli drawn from three different ranges of VOT stimuli, one Englishlike and two Polishlike.

A. Procedures

The stimuli were (a) a subset of those synthesized at Haskins Labs by Abramson and Lisker (1970), and (b) similar stimuli also made at Haskins (Blumstein *et al.*, 1977). Both sets represent apical stops followed by the vowel [a]. (Neither of these sets of stimuli is identical to that used in the preliminary tests.) Voicing lead is present as low-frequency harmonics of a buzz source. Voicing lag involves F_1 cutback (attenuation) and exci-

tation of F_2 and F_3 by a noise source. Details of stimulus construction are given in the original sources.

Three subsets of stimuli were chosen from this set, each forming a range of VOT values in 10-ms steps. From the Lisker and Abramson (1970) stimuli, two Polishlike ranges were made. The first, from -100 to +50 ms VOT, approximates the actual production distribution for apical stops in Polish. The second, from -100 to +20 ms VOT, represents an idealized Polish distribution limited to the prevoiced and true short-lag VOT categories. This range was included for two reasons: first, to see whether Polish listeners treat this range differently from the -100/+50 ms VOT one, and second, as a strong test of range effects in English. Because the English voicing boundary for apical stops has been found to fall at about +35 ms VOT, the range from -100 to +20 was expected to include only the voiced category for the English listeners.

The third stimulus subset ranged from -20 to +80 ms VOT, representing an Englishlike range. A tape which had been made previously for another study (Blumstein *et al.*, 1977) was used, containing one complete randomization of many repetitions of each stimulus. In the present study, listeners heard enough of the tape so that there were at least ten repetitions of each stimulus, requiring 146 test items to be used. Thus uneven numbers of repetitions of each stimulus were heard, from 10 to 17 repetitions per stimulus.

For the Polishlike ranges, block-by-block randomization was used. Each block consisted of one complete randomization of the continuum, with 10 blocks altogether. The ISI was 4 s and the IBI was 10 s. The three tests were originally recorded onto reel-to-reel tapes and were later copied onto a high-quality cassette tape for use in Poland.

Twenty-four students at the Institute of Telecommunications and Acoustics in Wrocław, Poland, and six Americans, most of whom were associated with the Brown University Phonetics Lab, participated as subjects in the experiment. An additional American subject had to be eliminated because he reported hearing no /t/'s on the test tapes. All subjects were native speakers of Polish and English, respectively, and had no speech or hearing problems. The Poles were paid volunteers, while the Americans were just volunteers.

For the Polish listeners, these tests formed only part of an experimental session (cf. Keating, 1979). All of the tests discussed here were run within this session. Half of the subjects heard the tests in the order -20/+80, -100/+50, -100/+20 ms VOT, and half heard them in the reverse order. The American listeners heard only these three tests, counterbalanced for order in the same way. Listeners' responses were forced-choice "D" or "T." The monosyllables used in the tests, /da/ and /ta/, are real Polish words, and Polish listeners were encouraged to hear them as such.

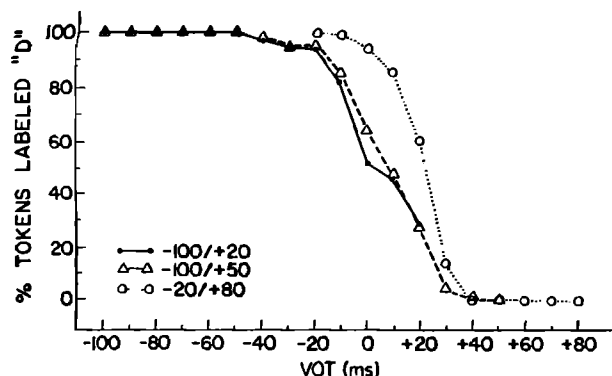


FIG. 4. Averaged labeling function for 21 Polish listeners on three VOT ranges: $-100/+20$, $-100/+50$, $-20/+80$. Summary statistics are given in Table II.

B. Analyses and results

The number of "D" responses to each stimulus was determined for each subject for each test. Labeling functions averaged for all subjects are shown in Figs. 4 (Polish) and 5 (English). Individual category boundaries were computed using Probit analysis (Finney, 1971), which fits a straight line to z -score transforms of the percentage of "D" responses per stimulus. The phoneme boundary (the point where the fitted line crosses 50%) and the slope of the fitted line were calculated for each function. In some cases category boundaries could not be computed—either because the identification function never crossed the 50% point,² or because the identification function was nonmonotonic (more than one labeling category for a single response).³ When subjects' phonetic identification functions did not cross the 50% point, their data were not excluded; rather, they were assumed to have legitimate category boundaries beyond the endpoint stimulus. To exclude these data would be to lose some important information about how boundaries shift with range. Therefore, medians and nonparametric statistical comparisons were used. Data on the computed boundaries are given in Table II.

For Polish, a Chi-square test on the distribution of boundaries for each range around the grand median

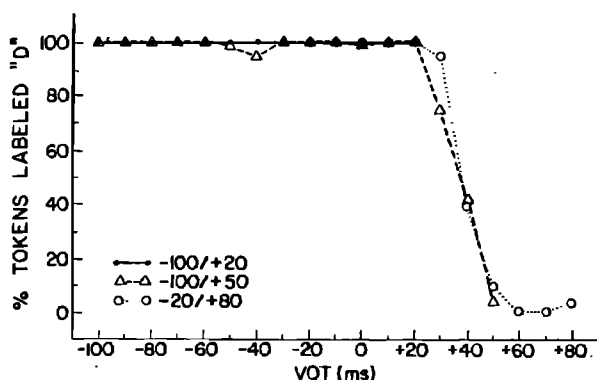


FIG. 5. Averaged labeling function for six English listeners on three VOT ranges: $-100/+20$, $-100/+50$, $-20/+80$. Summary statistics are given in Table II.

(12.5 ms VOT) showed a significant effect of range [$\chi^2(2)=26.63$, $p < 0.001$]. Examination of the data suggested that the size of the range effects was influenced by Task Order. To check this possibility, a two-way ANOVA on Range \times Task Order was computed. However, because of questions about the inclusion of subjects who had only one labeling category on the $-100/+20$ ms VOT range, several alternative tests were carried out, with and without these data.⁴ In all the analyses, there was a significant effect of Range, and a significant interaction of Range \times Task Order. Thus, we can be confident that the significant effects are not an artifact of any particular analysis technique or treatment of exceptional data. Polish listeners showed different boundaries depending on VOT range. *Post hoc* Newman-Keuls tests indicated that the effect of Range was due to the $-20/+80$ range being significantly different from both of the others. The interaction of Range \times Task Order reflects the fact that the boundary for a given range was higher when that range was heard first, and lower when it was heard last. This effect was particularly pronounced for the $-100/+20$ range, which had a lower boundary (and therefore a larger difference across ranges) when it was heard last. Perhaps after hearing two ranges with a fair number of both "T"'s and "D"'s, listeners lowered their boundaries so as to try to equalize the number of responses in each category.

The analysis of the American data included only the $-100/+50$ and $-20/+80$ ranges. A repeated-measures ANOVA for Range \times Task Order showed no significant effects or interactions (all $F < 1$). Although individuals showed large shifts, there was no systematic variation. The mean boundaries for the two ranges are identical, and on the $-100/+20$ ms VOT range, where the American listeners' boundary region is not included, they remained consistent as a group in their categorization and showed no boundary at all.

III. EXPERIMENT 2

This experiment was an attempt to replicate the previous results for Polish, by testing two new ranges of VOT. The stimuli were the same as those used in the

TABLE II. Summary data from statistical tests for labeling tests in experiment 1. For each VOT continuum used, the number of subjects included, the actual range of boundaries computed for those subjects, and the median boundary for that continuum are given for Polish and English tests. Boundary figures are in ms VOT.

Continuum	N	Range of boundaries	Median
Polish			
$-100/+20$	21	$-34 / >+20$	+5
$-100/+50$	21	$-20 / +29$	+6
$-20/+80$	21	$+10 / +30$	+20
English			
$-100/+20$	6	all $>+20$	none
$-100/+50$	6	$+28 / +44$	+37.5
$-20/+80$	6	$+29 / +46$	+37.5

preliminary discrimination tests. The continuum range used was from -70 to $+70$ ms VOT in 10-ms steps, plus stimuli at $+5$ and $+15$ ms VOT. The Polish range used was -70 to $+30$ ms VOT, and the experimental range was -30 to $+70$ ms VOT (13 stimuli each). There were 12 repetitions of each stimulus, with an ISI of 4 s and an IBI of 6 s. For each test, stimuli were randomized block-by-block.

Eight adult native Polish listeners in Bódz, Poland, the same listeners who served in the discrimination experiment, served as paid volunteer subjects. In the first session, subjects did the discrimination task as described above, and then labeled stimuli from one of the two ranges. In a second session four days later, they labeled stimuli from the other range. The order in which the two ranges were tested was counterbalanced across the eight listeners. They wrote "D" or "T" on answer sheets. All instructions were given in Polish by a native speaker.

A. Results and discussion of experiment 2

The results of this experiment are illustrated in Fig. 6, which shows the mean identification functions for the two ranges. Boundaries were computed using Probit analysis and were compared using a correlated t test. The mean boundary for the $-70/+30$ ms VOT range was -3.5 ms VOT; the mean boundary for the $-30/+70$ ms VOT range was $+8.4$ ms VOT. This difference was significant ($t = -5.502$, $df = 7$, $p < 0.001$). Thus these results confirm the tendency for a large (11 ms) boundary shift found for Polish in experiment 1. However, these results do not confirm the exact locus of the phoneme boundaries. The boundary for the $-30/+70$ ms VOT range, $+8.4$, is quite different from the $+20.9$ -ms mean for the similar range of $-20/+80$ ms VOT, and is not too different from the $+6.2$ ms VOT mean for the $-100/+50$ ms VOT range. On the other hand, the mean boundary for the $-70/+30$ ms VOT range is by far the lowest found, at -3.5 ms VOT.

Thus the magnitude of the shift found is nearly as great as that of experiment 1, but the boundaries for both ranges are lower than those obtained in experiment 1 (Sec. III). There are three possible causes for this discrepancy. The first is the slight differences in the various Lisker and Abramson VOT stimuli used

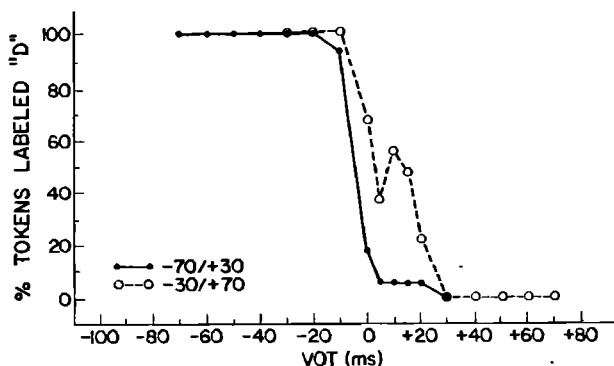


FIG. 6. Averaged labeling function for eight Polish listeners on two VOT ranges: $-70/+30$, $-30/+70$.

in these ranges. The second is the differences in stimulus frequency in the test tapes. The $-20/+80$ ms tape contained unequal numbers of each stimulus, and was not randomized into blocks. The $-30/+70$ - and $-70/+30$ -ms ranges included stimuli at $+5$ and $+15$ ms as well as 10-ms steps. If there were frequency as well as range effects, these stimuli could have helped weight the continua towards the short-lag category. The third is the differences in subjects and test equipment, since these two experiments were conducted in different locations. However, the results of this experiment do confirm the susceptibility of the Polish voicing boundary to substantial range effects.

IV. EXPERIMENT 3

In this section we address a potential problem in our cross-language comparison, and further extend our results for English. The question is how listener expectations about the test itself (not about the linguistic system) influence performance. We consider how expectations may have been implicitly manipulated in previous experiments, and we manipulate them explicitly here.

It could be the case that the Polish and American listeners had quite different expectations in experiment 1, especially as the Americans were somewhat more experienced in this type of perceptual task. As the test proceeds, the subject typically becomes more and more uneasy about the "failure" to hear approximately equal proportions of, here, "D" and "T." Possibly the relatively test-naive Polish listeners consciously tried to equalize the proportions of responses, while the relatively test-wise Americans realized that three labeling tasks were likely to differ in response proportion (for any of a number of reasons). The cross-language difference in range effects found in experiment 1 could be caused by this difference. To eliminate this possibility, new tests were carried out with naive American undergraduates. Some were told that the response proportions would be unequal, on the hypothesis that they would not show range effects since they would not be uneasy about their responses. Others were incorrectly told to expect equal response proportions, on the hypothesis that their resulting uneasiness would encourage range effects.

A new test tape with a range of $+20/+80$ ms VOT was constructed from the old $-20/+80$ -ms test tape. On the new tape, block-by-block randomization was used, with 10 blocks (repetitions) of seven stimuli each. The ISI was 3 s, and the IBI was 6 s. Twenty students at Wellesley College and 10 students and employees at MIT participated as paid volunteers. All were naive native speakers of American English who reported no speech or hearing problems and had not participated in speech perception experiments before.

Subjects were tested alone or in small groups in quiet testing rooms at Wellesley or MIT. They wrote "D" or "T" on answer sheets. The 10 subjects at MIT (controls) were given the same instructions used in previous experiments; no mention was made of how

TABLE III. Summary data for experiment 3. For each experimental group, the mean boundary and standard deviation are given in ms VOT.

Group	Mean	Standard deviation
Control	38.7	5.1
1 (D = T)	34.4	1.6
2 (D < T)	34.6	2.6

many "D"'s and "T"'s there would be. Ten of the subjects at Wellesley (condition 1) were explicitly told (incorrectly) that there would be an equal number of "D"'s and "T"'s. The last 10 subjects (condition 2) were told (correctly) that there would be more "T"'s than "D"'s.

A. Results and discussion of experiment 3

An estimate of the phoneme boundary was computed for each subject using Probit analysis. The mean results by group are given in Table III. An ANOVA with subjects nested under condition gave a significant overall effect [$F(2, 27) = 4.9557, p < 0.02$]. *Post hoc* Newman-Keuls tests failed to show any significant differences between any groups, and so *post hoc* uncorrelated *t* tests, which are much less conservative (more sensitive) were also done. These indicated that the significant main effect of the analysis of variance was due to the two treatment groups each being significantly different from the control group, but not from each other. [For the controls versus condition 1, $t = 2.527 (p = 0.02)$; for controls versus condition 2, $t = 2.251 (p < 0.03)$; for condition 2 versus condition 1, $t = -0.209, (p < 0.8)$.]

The result that the two experimental groups were not significantly different indicates that differences in subject expectations are not the cause of the cross-language differences. The effect of the two sets of explicit instructions was the same whichever expectations they encouraged. However, the control group was different from either experimental group, an unexpected result. The difference reached statistical significance because the variances within the experimental groups were extremely small. Why this should be so is unclear. One possibility is that calling subjects' attention to response proportions somehow inhibits range effects. Another is that differences in subjects or equipment caused the effect. Whatever the cause all conditions resulted in boundaries similar to those reported for other experiments in this study, and in other studies. The small difference obtained (4 ms VOT) is like those obtained by Brady and Darwin (1978). Thus these results support the conclusion that American listeners cannot easily be induced to show large range effects, although small effects may occur. We consider this same question in a different way in the next experiment.

V. EXPERIMENT 4

In this experiment we tested the effect of a massive difference in range for American listeners. Two ranges were used, $-100/+20$ and $+20/+80$ ms VOT, which share only the $+20$ ms VOT stimulus. Normally,

no "T" responses would be expected for the $-100/+20$ range. If any do occur, the calculated boundary could fall at about $+20$ ms VOT, 15 ms VOT lower than the expected boundary. Thus this experiment is designed to make it more likely for American listeners to produce a range effect of the size found for Polish listeners. It also allows a comparison of the $+20/+80$ range with the $-20/+80$ range from experiment 1, both of which cover the entire English voiceless category.

The $-100/+20$ ms VOT test tape used in experiment 1, and the $+20/+80$ ms VOT test tape used in experiment 3, were used again. Similarly, a new tape for the $-20/+80$ ms VOT continuum was made, with block-by-block randomization of the eleven stimuli.

Twelve students and employees at MIT participated in this experiment as paid volunteers. All were naive native speakers of American English who reported no speech or hearing problems and had not been in psychology or speech experiments before. Three other subjects from the same population did just the new $-20/+80$ ms VOT test.

Subjects were tested alone or in small groups in a sound-treated room at MIT. Twelve listeners heard both the $-100/+20$ and the $+20/+80$ ms VOT range tests, in counterbalanced orders, in two 15-min sessions a week apart.

A. Results

Mean results for this experiment are shown in Fig. 7. For the $-100/+20$ -ms range, only one of the 12 subjects had a function for which a category boundary could be computed.⁵ For the $+20/+80$ -ms range, all subjects had the usual type of labeling function, with a group mean boundary at 41.8 ms VOT. A statistical comparison was done of the number of "D" responses for each range to the shared $+20$ ms VOT stimulus. The three-way ANOVA (Range \times Task Order \times Subjects nested under Task Order) showed only a significant effect of Range [$F(1, 10) = 8.897, p < 0.01$]. Thus there is a reliable effect of stimulus range on responses with this extreme design, but it does not generally result in boundaries of $+20$ ms VOT or lower.

We compared these results with those for nine other

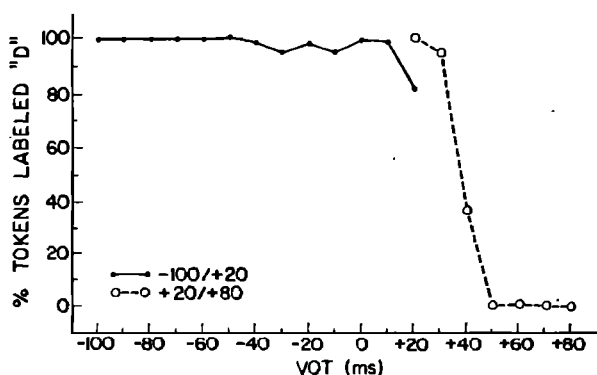


FIG. 7. Averaged labeling function for 12 English listeners on two VOT ranges: $-100/+20$, $+20/+80$.

listeners on the +20/+80 ms VOT range, the six subjects from experiment 3 and the three new ones. The group mean boundaries of 35 vs 40 ms VOT were significantly different [uncorrelated, unequal-*N* one-way ANOVA, $F(1, 19) = 5.7818, p < 0.03$]. Although such a *post hoc* comparison is not entirely valid, it does suggest that even with an entire voicing category present in each range, differences in overall range do affect Americans' performance.

VI. GENERAL DISCUSSION

In this paper we have reported on several experiments on the nature and extent of range effects for VOT in Polish and American English. A summary of the mean boundaries, standard deviations for those means, and median boundaries found for the various ranges used is given in Table IV. It can be seen that the English boundaries are uniformly higher than any of the Polish boundaries, reflecting the fact that Polish and English use different VOT contrasts. This difference is reliable; for example, an uncorrelated one-way ANOVA for the -20/+80 ms VOT range ($N = 33$) showed a significant difference between the two language-groups [$F(1, 31) = 35.3067, p < 0.0001$].

It can also be seen that the Polish boundaries have higher standard deviations than the English boundaries, that is, more variation across individual listeners. In addition, the Poles showed more variation across experiments. The differences in the means across ranges are 23 ms VOT for the Poles, and only 5 ms VOT for the Americans.

Thus, one result of these studies (including the acoustic measurements and the discrimination test) is that the Polish categories and boundary are different from those of English. However, the major result is that Polish and English differ in their susceptibility to range effects. Polish listeners' boundaries vary widely with different ranges, while Americans' vary little or not at all. A focus of our discussion, then, must be an account of why speakers of some languages should be so much more sensitive to stimulus range.

The results of the discrimination and labeling tasks

TABLE IV. Summary of all data from experiments 1, 2, and 4, plus control condition from experiment 3. For each continuum, the mean boundary and standard deviation, and the median boundary, are given in ms VOT.

Continuum	<i>N</i>	Mean	Standard deviation	Median
Polish				
-100/+20	21	+4.2	15.1	+5.0
	16	-1.0	13.4	+1.5
-100/+50	24	+5.1	11.6	+5.0
-20/+80	24	+20.9	5.8	+20.0
-70/+30	8	-3.5	2.7	-2.5
-30/+70	8	+8.4	7.3	+10.0
English				
-100/+20	18	none	none	none
-100/+50	6	+37.3	6.4	+37.5
-20/+80	9	+35.0	6.9	+34.0
+20/+80	22	+40.4	5.1	+42.0

for Polish indicate that Poles are quite sensitive to differences in VOT around 0-ms VOT. The fact that their boundaries shifted as a function of stimulus range also shows that they can resolve relatively fine differences in VOT in their boundary region. Such a finding is surprising in view of evidence that these VOT values are generally less discriminable than others. In a series of experiments (Eimas *et al.*, 1971; Eimas, 1975), infants discriminated VOT pairs spanning the English boundary and extreme lead pairs, but consistently failed to discriminate pairs around 0 ms VOT as well as extreme lag pairs. Infants in a Spanish environment show a similar pattern (Lasky *et al.*, 1975). These results suggest that the Polish type of VOT boundary, around 0 ms, is not due to predispositions of the auditory system. Rather, Polish listeners must start with one set of discrimination functions as infants, and acquire another set as they learn Polish. Such a readjustment of VOT boundaries was postulated by Pisoni and his colleagues (e.g., Aslin and Pisoni, 1980) as part of a model in which auditory predispositions are attributed to a constraint on temporal resolution of components of complex stimuli (Pisoni, 1977). This constraint cannot account for the widespread occurrence of VOT contrasts like that of Polish (e.g., French, Spanish, Kikuyu),⁶ and in fact by itself seems to rule them out. However, the observation that some adult data do not correspond to infant data, and that therefore some process of adjustment must occur during acquisition of some languages, is supported by data such as ours.

Given our evidence of perceptual sensitivity in just that VOT region where the infant data least predicts such sensitivity, one could ask whether in fact Polish is using VOT to distinguish voiced and voiceless stops. In Polish, the initial voicing contrast can be seen as one of voicing during closure versus absence of closure voicing, and no actual VOT value need be computed by a listener. That is, our proposal is that Poles do not use VOT as English speakers do, in the sense of a temporal interval between burst and voicing onset. Given this type of contrast, and the gap in production values between the categories (cf. Fig. 2), Poles may never need to establish a precise VOT category boundary. Poles are thus at a disadvantage in a VOT labeling experiment, since they need not have a criterion for categorizing boundary stimuli in normal speech. The criterion they normally use may be highly correlated with VOT, but it may not require making fine temporal distinctions along that dimension.

While the Polish boundary may be inherently more variable because of the gap between the production categories, we still must account for the tendency of the perceptual boundaries to cluster in the short-lag region, rather than the lead region. A possible basis for boundaries in the short-lag VOT region is the psychoacoustic salience of certain spectral cues correlated with VOT. The articulatory dimension of VOT has several acoustic consequences, as is well known (see, for example, the references cited in Diehl, 1981, p. 12). Cues such as *F1* onset frequency or the presence of

aspiration noise after the burst, which depend on whether there is voicing in the immediate vicinity of the burst, distinguish stops with voicing lead or coincident onset from stops with some voicing lag (+15 to +25 ms VOT). Thus, these dimensions may underlie the preference of Polish listeners for voicing boundaries in the short-lag VOT region. There is ample evidence that lag region boundaries are more salient than lead region boundaries (Abramson and Lisker, 1972; Williams, 1977; Streeter, 1976a, b; Streeter and Landauer, 1976). Therefore we propose that Polish listeners, under conditions of experimental uncertainty, will shift boundaries towards psychoacoustically salient short-lag VOT values.

Our explanation thus depends on two factors. One is the gap between production categories in Polish, which makes an exact VOT boundary relatively less clear to Polish listeners. The other is the psychoacoustic salience of short-lag VOT cues which encourage a boundary in the short-lag region. We therefore predict that any language with the same distribution of VOT categories as Polish should show similar range effects. English does not show such effects for two reasons. First, its VOT categories are not separated by a wide gap; rather, English listeners must have a fairly precise boundary. Second, the English VOT boundary already corresponds to a psychoacoustically salient discontinuity along the VOT dimension. The evidence for this claim is the performance of chinchillas in VOT labeling (Kuhl and Miller, 1978), where the chinchillas have nearly identical category boundaries as humans. Our claim is that the English linguistic boundary is aligned with a psychoacoustic boundary (whatever its basis may be) and so is more easily maintained by listeners in the face of experimental manipulation.

Support for this claim can be found in a comparison of the slopes of the identification functions given by Probit analysis for the 24 Polish listeners in experiment 1. Since steepness of slope is a measure of consistency of categorization, we might have expected the linguistically appropriate categorizations to have the steepest slopes, while the shifted categorizations for the -20/+80 range would be less reliable and have shallower slopes. However, the opposite pattern is found; in addition, the standard deviation is smallest for the shifted range. These observations suggest that the shift condition provides the Polish listeners with a clearer standard for categorization; the most plausible interpretation is that such a clearer standard would be psychoacoustic.

The experiments presented here were originally undertaken in an attempt to reconcile incompatible data from Polish production and perception pilot studies. Our results here show less of a discrepancy, since all but one of the obtained perceptual boundaries fall more or less between the production categories. The exception is the +20 ms one for the -20/+80-ms range. The range corresponding least to normal Polish production produced a shifted boundary that did not match the production data. Thus our earlier results seem to have been due to the general susceptibility of Polish listen-

ers to range effects for VOT, and to language contact effects. Poles in Poland do best if a test range includes stimuli with substantial prevoicing; Poles in the United States will show somewhat higher boundaries even on Polishlike ranges, and English boundaries on an Englishlike range.

We were also interested in whether phonetic categories are absolute entities, or relative properties. For American listeners, small boundary shifts with VOT range can be seen in our results, as in Brady and Darwin's (1978) results. Their subjects showed more of a tendency to range effects than our English subjects did; presumably this difference is due to differences in our methodologies. The small size of these range effects shows that English VOT categories are quite stable.

In contrast, our Polish subjects showed larger shifts, which would appear to indicate more fluid categories. Even so, these shifts seem to be constrained by the categories being labeled. For example, though we find effects of range on categorization, we never find that a boundary shifts to the midpoint of a range (Diehl *et al.*, 1980). (For example, the midpoint of the -20/+80 range is +30 ms VOT, but even the extreme boundary found for this range is only +20 ms VOT.) That is, the categorizations are not completely fluid. Rather, as we have noted, there is a region of high discriminability within which the various boundaries fall and which seems to constrain possible boundary shifts.

In conclusion, the experiments we have presented demonstrate that languages can differ in their sensitivity to simple experimental manipulations such as range effects. We attribute these differences to differences in the internal composition of their phonemic categories. Little is known about how the use of certain linguistic categories rather than others affects speech production, perception, or acquisition. Such research would lead to a better understanding of the nature and basis of the factors underlying linguistic phonetic contrasts, and the preferences languages show in choosing among them.

ACKNOWLEDGMENTS

This research was supported in part by postdoctoral fellowships to P. Keating (from NIH) and to W. F. Ganong (from NSF and NIH). Earlier versions of this work were presented at Acoustical Society of America, Linguistic Society of America, and American Association of Teachers of Slavic and East European Languages meetings. Experiment 1 is included in Chap. 2 of P. Keating's 1979 Brown University dissertation; Sheila Blumstein and Philip Lieberman of the Department of Linguistics were valuable committee members. We also acknowledge the help of Leigh Lisker, Arthur Abramson, and Terry Halwes at Haskins Laboratories in making stimuli available to us, with support from NIH. For help on conducting our experiments, we thank Dr. Wojciech Majewski of the Instytut Telekomunikacji i Akustyki in Wrocław, Poland; the Mikoś family in Bódz, Poland; and the Wellesley College Psychology Department. We wish to thank Kenneth Stevens, Bruno Repp, and Robert Zatorre for helpful comments.

APPENDIX

This appendix contains a description of the parameters used to synthesize the stimuli in the VOT continuum used in the preliminary tests and experiment 2. Other stimuli used were similar in construction (cf., Blumstein *et al.*, 1977; Abramson and Lisker, 1970).

Stimuli all represented apical stops plus the vowel [a]. Each had three formants. Frequencies were set the same for all stimuli. Before and during the burst, $F_1=151$ Hz, $F_2=1611$ Hz, and $F_3=3697$ Hz. Five ms after the burst, F_3 started to fall, reaching 2527 Hz 25 ms later. Ten ms after the burst, F_2 started to fall, reaching 1230 Hz 40 ms later, and F_1 started to rise, reaching 765 Hz 50 ms later. The VOT distinctions were made by varying excitation sources and amplitude settings of the three formants. For stimuli with voiced bursts (lead and 0-ms values), F_1 came on with a low amplitude at voice onset, then jumped to its peak value at 5 ms after the burst. F_2 came on at the burst, with its amplitude starting to rise 10 ms later, and reaching its peak value 50 ms after the burst (following its frequency course). F_3 came on at its peak value at the burst. For stimuli with voiceless bursts, F_1 came on at its peak value at voice onset. F_2 amplitude was like that for voiced stimuli. F_3 had a higher amplitude just at the burst (thus making the burst) and otherwise was like that for voiced tokens. Voice onset for lag stimuli consisted of buzz source excitation (rather than hiss), and an abrupt onset of F_1 . Prevoicing consisted of F_1 excited by the buzz source with a low amplitude. Stimuli were 350-ms long from the burst; any prevoicing thus represents added duration. (The Lisker and Abramson originals were longer than our versions.) The fundamental frequency was 114 Hz but fell gradually to 70 Hz at the end of the stimulus. Formant amplitudes also fell at the end.

test is thus the parametric equivalent of the chi-square test. In both cases there was a significant effect of Range at the 0.0001 level. (2) A repeated-measures Range \times Task Order ANOVA, once with and once without subjects having any estimated values. Both Range and the interaction of Range \times Task Order were significant at the 0.01 level for both tests [with 21 subjects, for Range, $F(2, 38) = 21.74$, $p < 0.0001$; for Task Order, $F(2, 38) < 1$; for Range \times Task Order, $F(2, 38) = 9.55$, $p < 0.001$]. (3) An ANOVA using not boundaries, but rather the proportion of "D" responses to the stimulus at +20 ms VOT, for Range \times Task Order \times Subjects nested under Task Order, with the data for all subjects included. As before, both Range and Range \times Task Order were significant at the 0.0001 level [for Range, $F(2, 44) = 15.0238$; for Range \times Task Order, $F(2, 44) = 13.0104$].

⁵Six used exclusively "D" responses, and five either used "T" responses randomly across the entire range, or else did not have enough "T" responses to meet the criterion for two separate labeling categories.

⁶Simon and Fourcin, 1978; Williams, 1977b; Steeter, 1976a.

Abramson, A. S., and Lisker, L. (1970). "Discriminability along the voicing continuum: Cross-Language tests," in *Proc. 6th Int. Congr. Phon. Sci.*, Prague (1967) (Academia, Prague), pp. 569-73.

Abramson, A. S., and Lisker, L. (1972). "Voice-timing perception in Spanish word-initial stops," *J. Phon.* 1, 1-8.

Aslin, R. N., and Pisoni, D. B. (1980). "Some developmental processes in speech perception," in *Child Phonology, Vol. 2: Perception*, edited by G. H. Yeni-Komshian, J. F. Kavanaugh, and C. A. Ferguson (Academic, New York).

Blumstein, S. E., Cooper, W. E., Zurif, E. B., and Caramazza, A. (1977). "The perception and production of voice-onset time in aphasia," *Neuropsychologia* 15, 371-383.

Brady, S. A., and Darwin, C. J. (1978). "A range effect in the perception of voicing," *J. Acoust. Soc. Am.* 63, 1556-58.

Diehl, R. L. (1981). "Features detectors for speech: A critical reappraisal," *Psychol. Bull.* 89, 1-18.

Diehl, R. L., Elman, J. L., and McCusker, S. B. (1978). "Contrast effects on stop consonant identification," *J. Exp. Psychol.* 4, 599-609.

Diehl, R. L., Lang, M., and Parker, E. M. (1980). "A further parallel between selective adaptation and contrast," *J. Exp. Psychol.* 6, 24-44.

Eimas, P. D. (1975). "Speech perception in early infancy," in *Infant Perception*, Vol. II, edited by L. B. Cohen and P. Salapatek (Academic, New York), Chap. 6.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). "Speech perception in infants," *Science* 171, 303-306.

Eimas, P. D., and Corbit, J. D. (1973). "Selective adaptation of linguistic feature detectors," *Cog. Psychol.* 4, 99-109.

Finney, D. J. (1971). *Probit Analysis* (Cambridge U. P., Cambridge).

Keating, P. A. (1979). "A phonetic study of a voicing contrast in Polish," unpublished Ph.D. dissertation, Brown University.

Keating, P. A., Westbury, J. R., and Stevens, K. N. (1980). "Mechanisms of stop-consonant release for different places of articulation," *J. Acoust. Soc. Am. Suppl.* 1, 67, S93.

Kuhl, P., and Miller, J. D. (1978). "Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli," *J. Acoust. Soc. Am.* 63, 905-917.

Lasky, R. E., Syrdal-Lasky, A., and Klein, R. E. (1975). "VOT discrimination by four and six and a half month old infants from Spanish environments," *J. Exp. Child Psychol.* 20, 215-225.

Lisker, L., and Abramson, A. S. (1964). "A cross-language study of voicing in initial stops: Acoustical measurements," *Word* 20, 384-422.

Lisker, L., and Abramson, A. S. (1967). "Some effects of context on voice onset time in English stops," *Lang. Speech*

¹For example, Ohde and Sharf (1979) looked at adaptation of a VOT continuum in several experimental conditions, and found boundary shifts of from 1 to 6 ms VOT. Tartter and Eimas (1975) obtained a VOT boundary shift of 3 ms. [However, Eimas and Corbit (1973) obtained a 10-ms shift with a voiceless adaptor, much larger than the usual size effect.]

²Three Polish listeners' responses on the -100/+20 ms VOT range had to be eliminated. The criterion adopted was that their percent "D" responses fell to below 50% at some VOT value, but at the next higher VOT value there were at least an additional 50% "D" responses, e.g., 10% followed by 60%.

³Five Polish listeners had only one labeling category on the -100/+20 ms VOT range, as did all six Americans. The criterion adopted for boundaries was that the last stimulus along the continuum had to have fewer than 50% "D" responses.

⁴Subjects who had no boundary on the -100/+20 range were assigned an "estimated" boundary of +21 ms VOT. The statistical tests were (1) a one-way, three-level ANOVA using only range, in an uncorrelated, unequal numbers design, once with and once without the estimated values. This

- Lisker, L., and Abramson, A. S. (1970). "The voicing dimension: some experiments in comparative phonetics," in *Proc. 6th Int. Congr. Phon. Sci.*, Prague (1967) (Academia, Prague).
- Mikoš, M. J. (1977). "Problems in Polish phonology," unpublished Ph.D. dissertation, Brown University.
- Mikoš, M. J., Keating, P. A., and Moslin, B. J. (1978). "The perception of voice onset time in Polish," *J. Acoust. Soc. Am. Suppl.* 1 63, S19.
- Ohde, R. N., and Sharf, D. J. (1979). "Relationship between adaptation and the percept and transformations of stop consonant voicing: Effects of the number of repetitions and intensity of adaptors," *J. Acoust. Soc. Am.* 66, 30-45.
- Pisoni, D. B. (1977). "Identification and discrimination of the relative onset of two component tones: Implications for the perception of voicing in stops," *J. Acoust. Soc. Am.* 61, 1352-1361.
- Rosen, S. M. (1979). "Range and frequency effects in consonant categorization," *J. Phon.* 7, 393-402.
- Sawusch, J. R., and Pisoni, D. B. (1973). "Category boundaries for speech and nonspeech sounds," 86th meeting ASA, Los Angeles.
- Simon, C., and Fourcin, A. (1978). "Cross-language study of speech-pattern learning," *J. Acoust. Soc. Am.* 63, 925-935.
- Simon, H. J., and Studdert-Kennedy, M. (1978). "Selective anchoring and adaptation of phonetic and nonphonetic continua," *J. Acoust. Soc. Am.* 64, 1338-1357.
- Streeter, L. A. (1976a). "Kikuyu labial and apical stop discrimination," *J. Phon.* 4, 43-49.
- Streeter, L. A. (1976b). "Language perception of 2-month-old infants shows effects of both innate mechanisms and experience," *Nature* 259, 39-41.
- Streeter, L. A., and Landauer, T. K. (1976). "Effects of learning English as a second language on the acquisition of a new phonemic contrast," *J. Acoust. Soc. Am.* 59, 448-61.
- Tartter, V. C., and Eimas, P. D. (1975). "The role of auditory feature detectors in the perception of speech," *Percept. Psychophys.* 18, 293-298.
- Weismer, G. (1979). "Sensitivity of voice-onset time (VOT) measures to certain segmental features in speech production," *J. Phon.* 7, 197-204.
- Williams, L. (1977). "The perception of stop consonant voicing by Spanish-English bilinguals," *Percept. Psychophys.* 21, 289-297.
- Williams, L. (1977). "The voicing contrast in Spanish," *J. Phon.* 5, 169-184.
- Zlatin, M. A. (1974). "Voicing contrast: perceptual and productive voice onset time characteristics of adults," *J. Acoust. Soc. Am.* 56, 981-994.