

The perception of stop consonant voicing by Spanish-English bilinguals

LEE WILLIAMS

Harvard University, Cambridge, Massachusetts 02138

The performance of Spanish-English bilinguals in two perception tasks, using a synthetic speech continuum varying in voice onset time, was compared with the performance of Spanish and English monolinguals. Voice onset time in speech production was also compared between these groups. Results in perception of bilinguals differed from that of both monolingual groups. Results of bilingual production in their two languages conformed with results obtained from each monolingual group. The perceptual results are interpreted in terms of differences in the use of available acoustic cues by bilingual and monolingual listeners of English and Spanish.

In recent years there has been a growing research interest in bilingualism. In much of this work the emphasis has been on the semantic and syntactic factors underlying bilingual performance. Very little attention has been paid to bilinguals' perception of speech. However, the ability to correctly perceive the acoustic-phonetic material of their two languages should directly influence whether or not bilinguals speak either language with or without an accent as well as their ability to comprehend running speech.

The primary purpose of the present study was to investigate bilingual ability to perceive phonetic material in the context of each language and to compare performance in perceptual tasks with the ability to produce the same phonetic material in speech. Implicit in this focus is a comparison between the bilingual's performance in their two languages with the performance of monolingual speakers of each. A secondary purpose of the study was to determine whether or not bilingual performance shows the effects of a language-specific perceptual set during a speech perception task, specifically if task performance differs according to which language the bilingual listener is prepared to hear. This question is related to the notion that bilinguals have two separate coding systems, one for each language, and that they are "switched" into the system of the

language they are currently processing (Kolers, 1966; Macnamara, 1967; Macnamara, Krauthamer, & Bolgar, 1968).

EXPERIMENTAL PARADIGM

The specific area of investigation in this study was concerned with patterns of perception by Spanish-English bilinguals of two phonemic classes which differ in the phonetic feature of "voicing." The choice of the voiced vs. voiceless contrast was made, not only because there is some knowledge of how this contrast is realized in speech production as well as in perception (Abramson & Lisker, 1970, 1973; Lisker & Abramson, 1964, 1970), but also because it is a contrast that is known to cause difficulty for the second language learner moving from either Spanish to English, or vice versa (Stockwell & Bowen, 1970), and may therefore continue to present problems for the Spanish-English bilingual.

Based on spectrographic analyses of utterances produced by native speakers of Spanish and English, it has been demonstrated that an acoustic variable called voice onset time, or VOT, provides a fairly reliable measure to allow for the separation of voiced and voiceless word-initial homorganic stops in spoken utterances within both Spanish and English (Dent, 1976; Lisker & Abramson, 1964; Williams, 1977). Voice onset time is defined as the duration between the onset of voicing, or vocal cord vibration, and the release of articulatory closure, e.g., the release of the lips in producing a [b] or a [p] (Lisker & Abramson, 1964). Zero in the voice onset time scale corresponds to the point of release, with negative values representing voicing onset preceding release (prevoicing) and positive values voicing onset following release (voicing lag).

There is also evidence that voice onset time can provide a sufficient perceptual cue for Spanish and English monolinguals, enabling them to distinguish between voiced and voiceless stops (Abramson &

The research reported here was supported by a research fellowship awarded by the National Institutes of Health (MTLH Branch No. 1 FO1 MH 48740) and a fellowship awarded by the National Institutes of Health (NICHD Award 1 F22 H00 1912-01). I would like to acknowledge my gratitude to Dr. Arthur Abramson and Dr. Leigh Lisker, whose own research was seminal to mine. I am also indebted to Dr. Alvin Liberman and Dr. Franklin S. Cooper, who made the facilities of the Haskins Laboratories available for the production of synthetic speech stimuli used in the experiment and offered their help and encouragement. I am deeply grateful to Dr. Kenneth S. Stevens, whose insightful comments were helpful at several stages in the completion of this project. The author's current address: Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

Lisker, 1970, 1973; Lisker & Abramson, 1970). This evidence comes from results in tasks in which Spanish and English listeners are presented with randomly ordered synthetic consonant-vowel syllables drawn from a series which varies in a step-wise fashion in terms of the voice onset time associated with the initial consonant. When they are asked to label each stimulus as beginning with either a voiced (e.g., /b/) or voiceless (e.g., /p/) consonant, they typically divide the series into voiced and voiceless domains which correspond to those domains derived from the spectrographic measurements of voice onset time from a spoken sample of their language. Different language-specific results are obtained from Spanish and English monolingual groups. Furthermore, when Spanish and English monolinguals are required to discriminate pairs of consonant-vowel syllables which are separated by equal physical intervals of voice onset time, they are, on the average, able to distinguish well only those pairs drawn from across the VOT category boundary, as defined by the division of these categories in spectrographic and labeling data. Again, this results in language-specific performance. Figures 1 and 2 summarize such performance in the perception and production of the voicing contrast taken from a study of syllable-initial labial stops by eight Puerto Rican Spanish and eight American English monolinguals (Williams, 1974). Data from the labeling task are represented as a function of percent of "/p/" judgments. The 50% point, where half of the responses are "voiced" and half "voiceless," defines the perceptual crossover between voiced and voiceless domains of the

VOT continuum. Performance in discrimination is represented as a function describing the percent of "different" judgments. All stimulus pairs being judged in fact differed by 20 msec of VOT. Production of voice onset time in Figures 1 and 2 are presented as a histogram representing the percent occurrence of VOT values within the voiced or voiceless categories.

There are a number of possible outcomes of a perceptual labeling task with bilingual listeners using a VOT series of synthetic speech stimuli. Some of these are illustrated in Figure 3.

(1) Bilinguals might have a *double standard* for perceiving speech which would show up as a sensitivity to acoustic properties which provide perceptual cues for the voicing contrast in each language. This would appear as a nonmonotonic function of the kind shown in Figure 3a. This function represents an increasing probability of labeling members of the continuum as /p/s as the stimuli move away from extreme prevoicing towards the region of VOT near the Spanish crossover location followed by a decreasing probability of /p/ judgments in the area of VOT approaching the English crossover location. A double standard in labeling performance might be considered to reflect "bilingual" perception. Accordingly, bilinguals would be expected to have four phonetic categories in terms of VOT, a Spanish /b/ and /p/ pair and an English /b/ and /p/ pair, with two distinct boundaries dividing the members of each pair. The presence of two perceptual category boundaries might not be tapped in a single two-category forced choice test. However, it might be

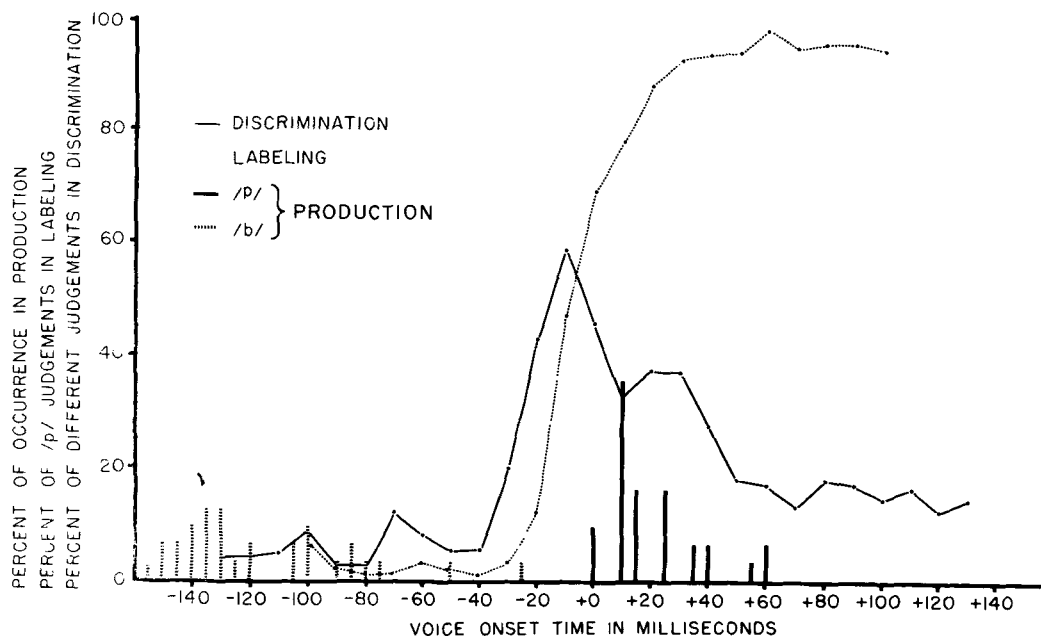


Figure 1. Measurements of voice onset time in word-initial voiced and voiceless labial stops and the results of labeling and discrimination of a synthetic series of speech stimuli varying in voice onset time by Spanish monolinguals. A 20-msec VOT step size was used in discrimination.

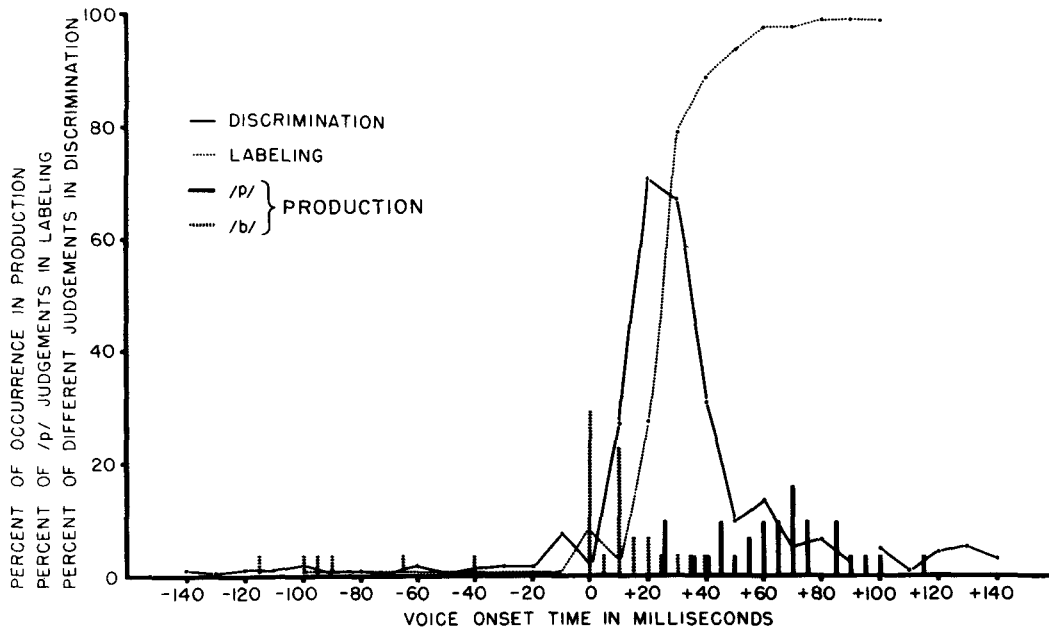


Figure 2. Measurements of voice onset time in word-initial voiced and voiceless labial stops and the results of labeling and discrimination of a synthetic series of speech stimuli varying in voice onset time by English monolinguals. A 20-msec VOT step size was used in discrimination.

demonstrated under the conditions of this experiment, in which there are two testing situations, each of which is designed to bring out a perceptual set for one of the bilingual's languages. Combining the data from two such tests might result in a function of the type shown in Figure 3a.

(2) Bilinguals might show *uncertainty* in labeling stimuli from the critical area between Spanish and English crossover locations. This could be reflected in a flattening of the function in the area lying between the Spanish and English crossover locations shown in Figure 3b. Uncertainty might also appear as a very gradual shift in labeling VOT continuum from the /b/ to the /p/ categories shown in Figure 3c. Uncertainty in labeling the stimuli in the critical area would presumably be the result of sensitivity to acoustic-phonetic information appropriate for a label of a Spanish /p/ as well as a label for an English /b/.

(3) Bilinguals might employ a *single standard* in making a voicing judgment such that their labeling functions would be similar to those of monolingual speakers of either English or Spanish, illustrated in Figure 3c by the steep dashed-line functions. A single standard could also be reflected by a steep labeling function lying somewhere between the monolingual Spanish and English functions.

Similar alternative results might be expected from a discrimination task, with bilinguals using the same set of VOT stimuli.

(1) A double standard would be reflected in double

peaks in discrimination over both Spanish and English crossover locations. Again, this might be considered the pattern of "bilingual" perception.

(2) The results might also show broad discrimination peaks covering the area of Spanish and English labeling crossovers. Again, this could be interpreted as a pattern of bilingual perception.

(3) A single standard could be shown as a sharp peak over either Spanish or English monolingual crossover locations or a peak falling somewhere between the two.

If bilingual perception follows the course of monolingual perception, it would be predicted that for each bilingual a peak in discrimination would correspond with the location of a labeling crossover.

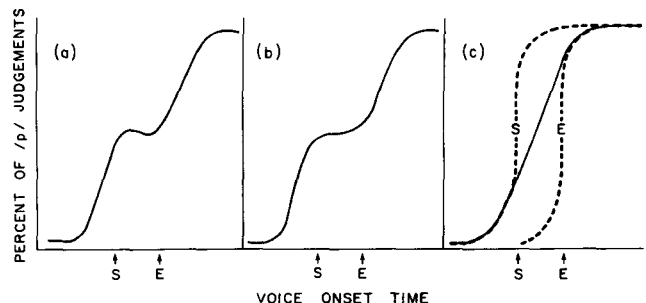


Figure 3. Three theoretical functions describing percent judgments of /p/ for bilinguals. The locations of 50% crossovers in labeling for monolingual Spanish and English groups are indicated on each graph by S and E, and idealized labeling functions for the two groups are plotted by dashed lines.

METHOD

Subjects

The subjects of this experiment were eight bilingual Spanish-English adults. English was the primary language, or language first learned, for three of the subjects (B.R., J.S., P.N.) and Spanish the primary language for five (N.L., F.M., D.D., R.M., E.G.). All but one subject (B.R.) acquired their second language either upon entering school in the primary grades or earlier. The subjects used both languages to varying degrees in their daily lives, but typically one language was spoken at home and the other in school or place of work. The bilingual subjects were between the ages of 17 and 30 and all came from well-educated backgrounds. Five were female (B.R., F.M., D.D., R.M., E.G.) and three male (J.S., P.N., N.L.).

A test of bilingualism was devised in order to provide for a measure of phonetic balance in Spanish and English as well as a criterion for entry into the subject pool. All eight subjects answered five questions in single, complete sentences and read five prepared sentences which contained sounds known to bring out an accent in unskilled speakers of that language. Twenty such utterances, a set in English and a set in Spanish, produced by each of the bilinguals, contributed to a total of 80 Spanish utterances and 80 English utterances. To the English utterances were added a set of 10 utterances each from four monolingual speakers of English, all from the Boston area, and a set of 10 utterances each from four individuals recognized as speakers of English with a Spanish accent. This made an additional 80 utterances which were added to the 80 English utterances produced by bilinguals. The grand total of 160 English utterances were spliced together in random order on a single tape with 12 sec between each utterance and 15 sec after every 10. A similar type of 160 utterances in Spanish was prepared.

The test for bilingualism consisted of the judgments of teachers of English and Spanish who listened to the tapes and judged each utterance as either being produced with an accent or not. Three teachers of English judged the English tape and three teachers of Spanish the Spanish tape. The speech material was recorded on a Sony 800 tape recorder using a Sony Model F-265 external cardioid microphone. Taped material was presented on a Tandberg 6000X tape deck with judges listening over matched Sharp Model 10 earphones. A bilingual subject met the criterion for entry into the experiment if at least 80% of his utterances were judged to be without an accent, i.e., at least 48 of the total 60 judgments of his utterances (10 utterances \times 3 judges \times 2 languages). Eight out of nine bilinguals interviewed passed the entry test. The results of this test were also used to order subjects in terms of their phonetic competence, reflected in their score in their test for each language.

Stimuli

The stimuli used in the perception tasks were produced on the parallel resonance synthesizer at the Haskins Laboratories in New Haven, Connecticut. They duplicated acoustic parameters used in the studies by Abramson and Lisker (1970, 1973). The speech stimuli were CV syllables, each made up of an initial stop consonant with formant transitions appropriate for a labial release, [b] or [p], followed by a three formant pattern appropriate for the steady state vowel [a]. Stimuli spanned a range of VOT values centered at the point of release, which was assigned the value of zero. The stimulus series varied in 10-msec steps of VOT, except for the range between -50 to +50, which varied in 5-msec steps. To represent voicing lead only the low-frequency harmonics (Formant 1) of the periodic source were used. Voicing lag was represented by the onset of a periodic source at Formants 1, 2, and 3. During the interval between release and the delayed onset of the periodic source, Formants 2 and 3 were excited by a noise source with suppression of the first formant. All stimuli were approximately 600 msec long.

The stimuli were recorded directly from the Haskins Laboratories speech synthesizer onto audio tapes for use in the labeling

and discrimination tasks. The 31 labeling test stimuli, which covered a VOT range of -100 to +100 msec, were arranged in eight random orders. There were 4 sec between each stimulus and 10 sec between every block of 10. The discrimination test was made up of 34 different stimulus pairs which covered the range of VOT from -150 to +150 msec arranged in eight random orders. All pairs differed by 20 msec of VOT. Pair members were separated by an interval of 1 sec, and each pair was separated from the following by 4 sec, with a 10-sec interval between every block of 10 pairs.

Procedure

There were four conditions for each of the eight subjects in the experiment, with a single testing session devoted to each condition: labeling and discrimination with English presentation and labeling and discrimination with Spanish presentation. The order of these was counterbalanced across subjects. The procedure was carried out in Spanish for one set of trials and in English for another set. Each session was preceded by 10 min of conversation in the language of presentation with an experimenter who was a native speaker of the language. Instructions and experimental materials were also in that language, and brief conversations occurred while the stimulus tapes were being changed.

Subjects in the labeling task listened to lists of the test syllables and identified each at time of presentation as being either a /ba/ or a /pa/ by pointing to a picture of an object whose name began with that syllable.¹ There was a Spanish picture set and an English picture set. Subjects in the discrimination task listened to pairs of syllables and were asked to judge whether the successively presented pair members sounded the same or different by pointing to a picture of two different faces for a "different" response and to a picture of two identical faces for a "same" response. The stimuli were presented to subjects by a Tandberg Model 6000X tape deck via a set of Sharp Model 10 earphones. Each bilingual made a total of 16 decisions on each stimulus or stimulus pair per language condition.

Each subject also produced 16 examples of word-initial voiced and voiceless labial stops in each language: two letters ([b] and [p]), two frames (single words and sentences), four vowel contexts in each language (see Table 1). These constituted the sample from which production data were derived.

All production words were repeated by the subjects after model utterances presented by an experimenter. A native English-speaking experimenter conducted the English presentation, and

Table 1
Words and Sentences Used in English and Spanish
for Production Data

SINGLE WORDS	WORDS IN SENTENCES
BOASTER	I DON'T LIKE THE BOASTER MUCH
POSTER	HELP ME PUT THE POSTER UP.
BEAKED	THE BIRD HAS A BEAKED BILL.
PEAKED	THE ROOF HAS A PEAKED TOP.
BALMY	IT IS A BALMY DAY.
PALMY	IT IS A PALMY BEACH.
BAYING	SEE THE DOGGIE BAYING LOUDLY.
PAYING	SEE THE MOMMIE PAYING IT.
BOCA	NO EXISTEN PROBLEMAS EN BOCA CERRADA.
POCA	DAME AGUA EN POCA CANTIDAD.
BULLA	HAY MUCHA BULLA AQUÍ.
PULLA	EL NIÑO SE RIÓ DE LA PULLA GRACIOSA.
BAÑO	VOY A DARME UN BAÑO CALIENTE.
PAÑO	EL NIÑO TIENE UN PAÑO SUCIO.
BESO	MAMÁ, DAME UN BESO GRANDE.
PESO	TENGO SOLO UN PESO PLATEADO.

a native Spanish speaking experimenter the Spanish presentation. Utterances were recorded at the end of the final two testing sessions on a Sony Model 800-B tape recorder with an exterior Sony cardioid microphone Model F-26S. Single words were recorded before words in sentences, and the order of presentation within both sets was randomized for each subject. The language in which words were produced was also counterbalanced across subjects, half of the subjects producing English and half Spanish words first. Wide-band spectrograms were made of all utterances from which measurements of VOT were taken. Two vertical lines were drawn on the spectrograms, one at the point of labial release, corresponding to that point when an abrupt change in spectral energy is visible, and the other at the point of voicing onset corresponding to the appearance of the first regularly spaced vertical striations at the levels of all three formants. Measurements of VOT were first made in millimeters and converted into milliseconds, rounding off to the nearest 5 msec.

RESULTS

Perception

Table 2 shows the location of 50% labeling crossovers and discrimination peaks for each subject according to language of presentation. It can be seen from the table that presenting the task in either language does not consistently increase the probability that subjects will have crossover locations closer to the monolingual value for that language. In fact, five of the eight subjects (F.M., N.L., B.R., D.D., P.N.) had crossovers closer to the English location under conditions of Spanish presentation than they did under conditions of English presentation, indicating that this factor was not a good predictor of a Spanish- or English-like performance in perception. In a similar manner, language of presentation did not systematically alter the location of peaks in discrimination, nor was there a systematic relationship between primary language and peak location. The generally small differences according to language of presentation which existed for each subject in the location of crossovers and

in the peaks of discrimination were judged to be the result of error variance, and the data for labeling and for discrimination were pooled for each subject across language of presentation (see Figure 4).

With the exception of two subjects (F.M. and N.L.), the labeling functions are sharp and monotonic, suggesting that most bilinguals were using a single criterion for judging the identity of the stimuli. The raw data from the labeling task were converted by probit transformation (Finney, 1952) and crossover locations were taken from best fitting straight line functions drawn to each bilingual's transformed data. Two analyses of variance were then performed in order to compare labeling performance of bilinguals with that of Spanish and English monolinguals from a prior study (Williams, 1974). The mean bilingual crossover location was significantly different from mean locations of both the Spanish monolinguals [$F(1,14) = 8.91, p = .01$] and English monolinguals [$F(1,14) = 8.71, p = .01$]. However, unlike the monolingual data, the bilingual group data did not appear to be homogeneous but rather broke into two sub-groups: three individuals with Spanish-like crossover locations, N.L. = -1, F.M. = -1, B.R. = +3, and five individuals with English-like locations, J.S. = +14, D.D. = +16, R.M. = +17, P.N. = +18.5, and E.G. = +22 (after probit transformation). A test comparing the crossover means of the two subgroups of bilinguals revealed a significant difference, [$t(6) = 3.7, p < .01$].

A further analysis of Spanish and English subgroups was performed testing the hypothesis that the observed subgroups are bona fide Spanish and English groups representing a sample with characteristics of the Spanish and English monolingual populations. This analysis just missed significance [$F(7,8) = 3.3, p > .05$].

Discrimination data for the bilinguals also divided into the same two subgroups. Subjects with Spanish-like crossovers (N.L., F.M., B.R.) had discrimination functions which were broad, covering the entire range of VOT from the monolingual Spanish location of -4 msec to the English location of +25 msec. This was a pattern predicted for listeners with a double standard in discrimination, or "bilingual" perception. The subjects with English-like crossovers (J.S., D.D., R.M., P.N., E.G.) had sharper discrimination functions, which fell closer to the English crossover location of +25 msec. Again we see in a perceptual task a split in the performance of bilinguals, with the same individuals falling into two different groups: five with a pattern of perceptual discrimination more like that of English monolinguals and three with a pattern which might be described as more bilingual.

It is possible that differences in factors independent of the perceptual test itself might help interpret the split in bilingual performance. In order to in-

Table 2
Location of 50% Labeling Crossovers and Major Discrimination Peaks for Bilinguals According to Language of Presentation

SUBJECTS	SPANISH PRESENTATION		ENGLISH PRESENTATION		SPANISH PRESENTATION		ENGLISH PRESENTATION		
FM <u>S</u>	0	>	-2		+20	<	+10	+25	
NL <u>S</u>	-2	>	-4		+20	=	-20		
BR <u>E</u>	+5	>	-1		+20	>	-12		
DD <u>S</u>	+20	>	+10		+20 / +30	<	-30	+40	
JS <u>E</u>	+14	<	+19		+20	=	-20		
RM <u>S</u>	+16	<	+20		+20	=	-20		
PN <u>E</u>	+21	>	+20		+20	+30	>	-20	
EG <u>S</u>	+22	=	+22		+20	+30	<	-20	+40
SPANISH MONO VALUE	-4		ENGLISH MONO VALUE	+25	SPANISH MONO. VALUE	-10	ENGLISH MONO VALUE	-20	
LABELING					DISCRIMINATION				

S or an E following a bilingual's initials indicates whether Spanish or English is his primary language.

investigate this possibility, certain facts were examined which were concerned with the bilingual's language performance, such as: a rank order of bilinguals according to their daily use of Spanish, a percentage score on the test of bilingualism for Spanish and English, a mean negative voice onset time value in isolated words with initial voiced stops in Spanish and English. Bilinguals were rank-ordered according to 50% crossovers from Spanish-like to English-like values, and Spearman rank order correlations between this measure and rank orders of the other five measures were done. None of the correlations were significant. The scattered distribution of primary language along the dimension of rank order of 50% labeling crossovers and discrimination peaks (see Table 2) indicates that neither does language first acquired help explain the bilingual split in perceptual performance.

Production

When production data from Spanish initial stops produced by bilinguals and Spanish monolinguals were compared in a one-way analysis of variance, combining isolated and sentence context utterances, there was no significant difference in the values of VOT in Spanish words produced by the two groups. When the data from bilinguals speaking Spanish were analyzed separately in a two-way analysis of variance, phoneme category ([b] vs. [p]) by frame (isolated words vs. words in sentences), two significant results emerged. The effect of phoneme category was highly significant [$F(1,7) = 206.43, p < .001$], reflecting the fact that distributions of VOT for initial voiced and voiceless stops were essentially nonoverlapping in Spanish produced by both bilingual and Spanish monolingual speakers. This statement must be qualified, since there was also a

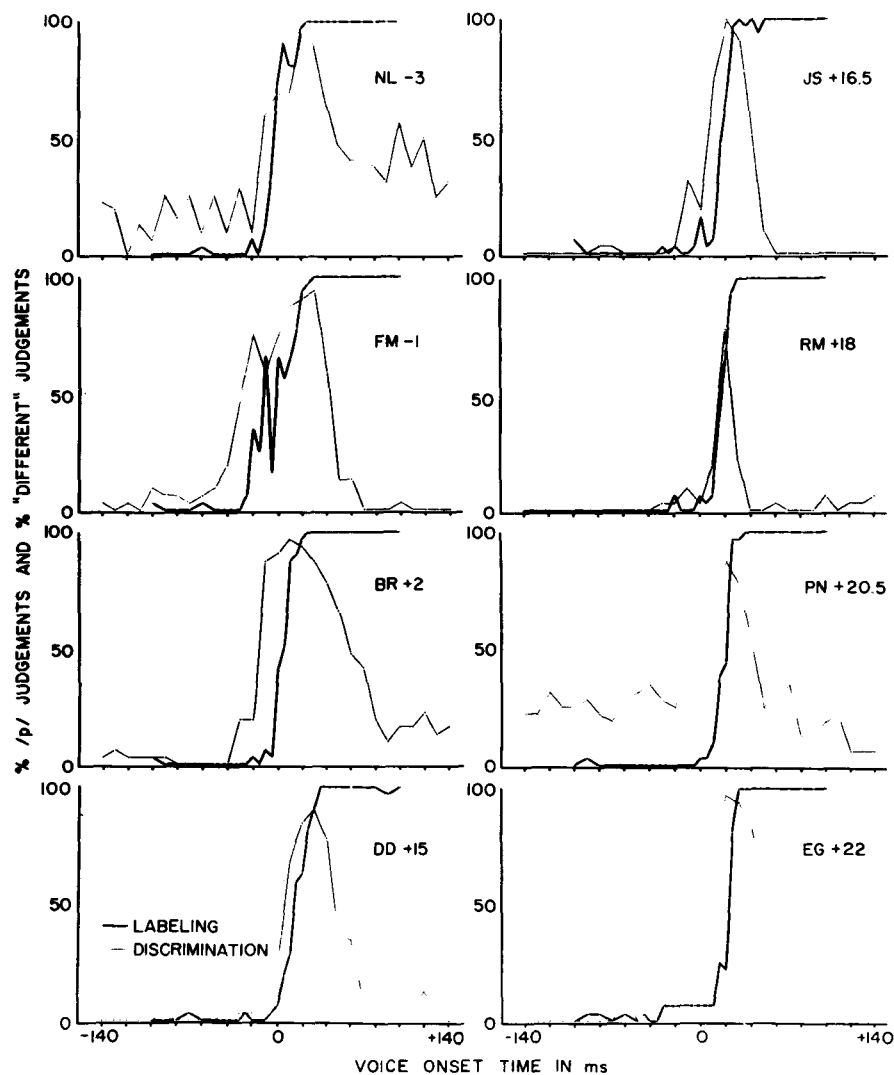


Figure 4. Labeling and discrimination functions for eight bilinguals. Crossover locations, prior to probit transformation, are indicated on the graph for each bilingual.

significant Phoneme Category by Frame interaction [$F(1,7) = 9.70$, $p = .02$], such that putting words into sentences had the effect of reducing both negative voicing lead in [b] and positive voicing lag in [p]. Similar findings of reduction of voice onset time values in sentence context have been reported by other investigators (Klatt, 1975; Lisker & Abramson, 1967). When the data on VOT values in English initial stops produced by bilinguals and monolingual English subjects of the previous study were combined into a three-way analysis of variance, a significant difference in language emerged, [$F(1,4) = 8.14$, $p = .01$]. A significant Language by Phoneme Category interaction, [$F(1,14) = 16.4$, $p = .01$] revealed that this was primarily attributable to the fact that bilinguals, on the average, tend to produce initial voiced stops in English words with more prevoicing than English monolinguals. In fact, bilinguals had a value of VOT for [b] in initial words which lay roughly between the mean values found for monolingual Spanish and English initial stops. When an analysis was done of the bilingual group alone speaking English, the only significant effect was phoneme category [$F(1,7) = 190.57$, $p = 0.1$].

DISCUSSION

With regard to the purpose of determining the nature of perceptual behavior for Spanish-English bilinguals, the labeling and discrimination data indicate that subjects do not have a double standard for judging the identity of speech stimuli employed in this experiment. The individual labeling functions did not conform to any of those predicted for perception by a bilingual using a double standard (see Figure 3). Instead, the functions were typically steep and monotonic, closely resembling the shape of pooled functions for monolinguals (see Figures 1 and 2). This suggests that bilinguals were dividing the voicing continuum into voiced and voiceless domains by utilizing changes in the acoustic variables that lay in a fairly restricted region of the speech-sound continuum.

A similar conclusion can be made based on the results of discrimination for those individuals with English-like performance. Only the three bilinguals with broad peaks in discrimination showed any indication of sensitivity to acoustic-phonetic information present at both the English and the Spanish crossover locations. Results of discrimination suggest the general principle that for all of the bilinguals differences between stimuli which were in the English crossover region of the VOT continuum could be discriminated. Only those bilinguals who demonstrated a Spanish-like pattern in dividing the continuum into voiced and voiceless domains in labeling could discriminate differences in stimuli in the Spanish crossover location.

With regard to the secondary purpose of this experiment, the results give no evidence that a language-specific perceptual set influences the responses of subjects in labeling and discrimination tasks using the present set of stimuli. Similar results have been reported by Caramazza, Yeni-Komshian, Zurif, and Carbone (1973) with English-French bilinguals. This finding does not constitute proof that a language-specific set does not influence the perception of speech. It only indicates that the conditions of this experiment do not elicit such an effect.

The voice onset time measurements within each language made from words spoken by Spanish-English bilinguals generally conform to those results reported previously for Spanish and English monolinguals (Lisker & Abramson, 1964; Williams, 1977). The only significant difference in the distribution of VOT values between bilinguals and monolinguals is in the production of the English word-initial voiced stop. Bilinguals, on the average, carry over into English the acoustic-phonetic feature prevoicing. This carry-over from Spanish to English in terms of voice onset time would presumably not interfere in any way with the perceptual acceptability of the word-initial voiced stops spoken in English by bilinguals, since prevoicing of initial voiced stops has repeatedly appeared in studies of the English voicing contrast, presumably an allophonic variant typical of the speech of some English speakers (Lisker & Abramson, 1964; Smith & Westbury, 1975; Zlatin, 1974). Producing an English initial voiced stop in an isolated word with prevoicing will, if anything, increase the acoustic-phonetic distance between the voiced and voiceless phonemes which would presumably accentuate their perceived contrast.

The results of this study appear at first glance to present somewhat contradictory findings. In production, Spanish-English bilinguals preserve the characteristics of each language in which they are fluent. On the other hand, results in a perceptual labeling task show that most bilinguals divide a VOT continuum into voiced and voiceless domains at a single compromise point lying close to either the Spanish or English monolingual dividing point. This would suggest that the bilingual cannot separate perceptually both the Spanish and English voicing contrasts. The results of the discrimination task compliment this by demonstrating that most bilinguals cannot discriminate stimuli in the two VOT ranges where maximum English and Spanish monolingual discrimination occurs. However, it is premature to conclude that most bilinguals studied were unable to distinguish a major phonemic class in both of their languages. A reasonable explanation for the perceptual results of this study can be found by examination of the acoustic characteristics of the synthetic speech stimuli used as well as the characteristics of their natural-speech counterparts.

In the synthetic VOT series, three acoustic properties, in addition to differences in VOT, varied parametrically in the region bracketing the English perceptual crossover location at +25-msec VOT. While these variables are probably a direct acoustic result of articulatory gestures which underlie differences in the timing of voicing onset, they may serve as perceptual cues independent of and in addition to the VOT cue, where VOT defines timing between release and voicing onset.

(1) The presence, absence, or varying duration of aspiration or aperiodic energy in the interval between articulatory release and the onset of voicing. Aspiration is visible on a spectrogram of a naturally produced English initial voiceless stop at all formants above Formant 1. The presence of aspirated formants has been demonstrated to provide a positive cue for initial voicelessness to English listeners (Winitz, LaRiviere, & Herriman, 1975).

(2) The absence of periodic acoustic energy at the level of Formant 1, during aperiodic excitation of the vocal tract referred to as "first formant cutback" (Liberman, Delattre, & Cooper, 1958). There is also evidence that the presence or absence of low-frequency periodic energy in the region of the first formant provides a perceptual cue for the initial voicing contrast for English listeners (Delattre, Liberman, & Cooper, 1955; Libermann et al., 1958; Lisker, 1975).

(3) Differences in the degree and temporal extent of formant transitions under conditions of periodic excitation of the vocal tract. There is some evidence that this acoustic variable may provide cues for the initial voicing contrast in English (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Stevens & Klatt, 1974; Summerfield & Haggard, 1974).

By contrast, in that part of the synthetic VOT continuum in which the average monolingual Spanish labeling crossover and discrimination peak are found, at -4 and -10 msec, respectively, only differences in negative values of voice onset time, i.e., in the presence vs. absence or duration of prevoicing, provide acoustic information upon which a perceptual distinction may be made between initial voiced and voiceless stops. In the synthetic series used in this study, as in natural speech, prevoicing is a low-frequency, low-intensity form of periodic energy. An interpretation of the perceptual results reported for Spanish *monolinguals* is that prevoicing of sufficient duration can provide a positive cue for initial voicedness, while the absence of this acoustic property provides a positive cue for initial voicelessness. However, given its relatively low intensity compared to other potential voicing cues, such as those described above for the English contrast, there is reason to suspect that there may be other acoustic properties which supply perceptual

cues for initial voicing even to the Spanish monolingual. Potential cues might be provided by:

(1) Differences between initial voiced and voiceless stops in the intensity, duration, and frequency distribution of the brief noise burst which accompanies articulatory release (Halle, Hughes, & Radley, 1957; Klatt, 1975; Zue, 1976).

(2) Differences in voice-pitch-change immediately following release across initial voiced and voiceless stops (Lea, 1973) which have been demonstrated to provide perceptual cues for initial voicing (Fujimura, 1971; Haggard, Ambler, & Callow, 1970).

Acoustic cues such as the above may be more perceptually salient than the relatively low-energy cue provided by differences in prevoicing. These potential cues for the initial Spanish voicing contrast are not present in the synthetic VOT continuum used in this study, yet the Spanish-English *bilingual* may rely on them to a greater extent than the Spanish monolingual due to his exposure to English, which includes acoustic properties other than VOT which vary across initial voiced and voiceless stops. If this were true, using a synthetic continuum which did not include these potential voicing cues for the Spanish initial voicing contrast but did include multiple cues for the English contrast, as was the case in the series of stimuli used in this study, the obtained results would be predicted: bilingual performance would show evidence for a perceptual distinction between initial voiced and voiceless stops in English but not in Spanish.

The fact that three out of eight bilinguals in the study showed some perceptual sensitivity to differences in stimuli lying in the negative or prevoiced region of the VOT continuum while five did not suggests that individual differences probably exist in the manner in which bilinguals use acoustic-phonetic cues. A more exact account of such differences and of the possible role of multiple acoustic cues for initial voicing for the Spanish-English bilingual would come first from a careful study of acoustic properties which vary systematically in natural-speech versions of the initial voicing contrast in English and Spanish in addition to VOT, defined as timing between release and voicing onset. To establish the perceptual role of any such property would require further testing with synthetic speech sound continua in which these properties were varied systematically. One set of results predicted by such an investigation would be that the bilingual has the ability to distinguish perceptually the phonemic contrasts in both of his languages, but that he does not use, as perceptual cues for a given contrast, the same acoustic properties used by monolingual speakers. Becoming bilingual may entail, among other things, a modification in the use of acoustic information present in the speech signal.

REFERENCES

- ABRAMSON, A., & LISKER, L. Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the Sixth International Congress of Phonetic Sciences, 1967, Prague*. Prague: Academia, 1970. Pp. 569-573.
- ABRAMSON, A., & LISKER, L. Voice timing perception in Spanish word-initial stops. *Journal of Phonetics*, 1973, 1, 1-8.
- CARAMAZZA, A., YENI-KOMSHIAN, G., ZURIF, E., & CARBONE, E. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America*, 1973, 54, 421-428.
- COOPER, F. S., DELATTRE, P. C., LIBERMAN, A. M., BORST, J. M., & GERSTMAN, L. G. Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, 1952, 24, 597-606.
- COOPER, W., BLUMSTEIN, S., & NIGRO, G. Articulatory effects on speech perception: A preliminary report. *Journal of Phonetics*, 1975, 3, 87-98.
- DELATTRE, P. C., LIBERMAN, A. M., & COOPER, F. S. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 1955, 27, 769-773.
- DENT, L. Voice onset time of spontaneously spoken Spanish voiceless stops. *Journal of the Acoustical Society of America*, 1976, 59, Supplement 1, S41.
- FINNEY, D. *Probit analysis; a statistical treatment of the sigmoid response curve*. Cambridge, England: The University Press, 1952.
- FUJIMURA, O. Remarks on stop consonants: Synthesis experiments and acoustic cues. In L. L. Hammerich, R. Jakobson, & E. Zwirner (Eds.), *Form and substance: Phonetic and linguistic papers presented to Eli Fischer-Jorgensen*. Denmark: Akademisk Forlag, 1971. Pp. 221-232.
- HAGGARD, M. S., AMBLER, S., & CALLOW, M. Pitch as voicing cue. *Journal of the Acoustical Society of America*, 1970, 47, 613-617.
- HALLE, M., HUGHES, G. W., & RADLEY, J. P. A. Acoustic properties of stop consonants. *Journal of the Acoustical Society of America*, 1957, 29, 107-116.
- KLATT, D. Voice onset time, frication and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 1975, 18, 686-706.
- KOLERS, P. A. Reading and talking bilingually. *American Journal of Psychology*, 1966, 79, 357-376.
- LEA, W. A. Influences of phonetic sequences and stress on fundamental frequency contours of isolated words. *Journal of the Acoustical Society of America*, 1973, 53, 346. (Abstract)
- LIBERMAN, A. M., DELATTRE, P. C., & COOPER, F. S. Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, 1958, 1, 153-167.
- LISKER, L. Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America*, 1975, 57, 1547-1551.
- LISKER, L., & ABRAMSON, A. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, 20, 384-422.
- LISKER, L., & ABRAMSON, A. Some effects of context on voice onset time in English stops. *Language and Speech*, 1967, 10, 1-28.
- LISKER, L., & ABRAMSON, A. The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the Sixth International Congress of Phonetic Sciences, 1967, Prague*. Prague: Academia, 1970. Pp. 563-567.
- MACNAMARA, J. The linguistic independence of bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 729-736.
- MACNAMARA, J., KRAUTHAMER, M., & BOLGAR, M. Language switching in bilinguals as a function of stimulus and response uncertainty. *Journal of Experimental Psychology*, 1968, 78, 208-215.
- SMITH, B. L., & WESTBURY, J. R. Temporal control of voicing during occlusion in plosives. *Journal of the Acoustical Society of America*, 1975, 57, Supplement 1, S71.
- STEVENS, K. N., & KLATT, D. H. The role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, 1974, 55, 653-659.
- STOCKWELL, R. P., & BOWEN, J. D. *The sounds of English and Spanish*. Chicago: The University of Chicago Press, 1970.
- SUMMERFIELD, A. Q., & HAGGARD, M. P. Perceptual processing of multiple cues and contexts: effects of following vowel upon stop consonant voicing. *Journal of Phonetics*, 1974, 2, 279-295.
- WILLIAMS, L. *Speech perception and production as a function of exposure to a second language*. Unpublished PhD dissertation, Harvard University, 1974.
- WILLIAMS, L. The voicing contrast in Spanish. *Journal of Phonetics*, in press.
- WINITZ, H., LARIVIERE, C., & HERRIMAN, E. Variations in VOT for English initial stops. *Journal of Phonetics*, 1975, 3, 41-52.
- ZLATIN, M. Voicing contrast: perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America*, 1974, 56, 981-994.
- ZUE, V. *Acoustic characteristics of stop consonants: A controlled study*. Unpublished ScD dissertation, Massachusetts Institute of Technology, 1976.

NOTES

1. A pointing response was used in this study in order to avoid the possible contamination of a perceptual decision by an orthographic influence resulting from a written response or by acoustic feedback resulting from the subjects speaking their responses. The kind of contamination anticipated by autofeedback would result from perceptuomotor adaptation, an effect reported by Cooper, Blumstein, and Nigro (1975).

(Received for publication August 31, 1976;
revision accepted December 7, 1976.)