

The modification of speech perception and production in second-language learning

LEE WILLIAMS

University of Denver, Denver, Colorado 80208

Seventy-two native Spanish speaking children enrolled in programs to teach English as a second language and 24 monolingual English children were tested in speech perception and production tasks with the purpose of determining whether or not there is learning at the phonetic level during second-language acquisition. Performance with the phonetic feature voicing was studied by means of measuring changes in the perception and production of the acoustic property voice onset time. Variables of interest across Spanish-speaking subjects were degree of exposure to English and age. Age was the only group variable for the English-speaking subjects. Results with the Spanish-speaking children showed significant changes in both perception and production of voicing towards the English monolingual pattern. Age had a significant effect in production, but not in perception. It was concluded that learning at the phonetic level does occur during second-language acquisition.

Resurging interest in improved methods for teaching a second language has stimulated a concomitant growth in related research (Hakuta & Cancino, 1977; Segalowitz, 1976). However, studies directed towards an understanding of second-language learning at the phonological and phonetic levels have been remarkably scarce, despite the fact that "native-like" pronunciation of the new tongue will depend upon considerable mastery of its phonological system and phonetic inventory. In fact, several issues related to phonological and phonetic development in second-language learning remain unclear for lack of a sufficient body of data. Two perennial questions have been: the importance of the age of onset of learning and the interaction between age and degree of exposure to the second language. These questions have been scarcely touched in scientifically conducted investigations (Asher & Garcia, 1969; Oyama, 1976; Streeter & Landauer, 1976). Studies which have been carried out have focused primarily on competence in speech production, leaving out important issues related to the perception of the second language.

The purpose of this set of experiments was to

The research reported here was supported by a research fellowship awarded by the National Institutes of Health (MTLH Branch 1 FO1 MN 48740) and a fellowship awarded by the National Institutes of Health (NICHD Award 1 F22 H00 1912-01). The author would like to acknowledge with gratitude the assistance of Dr. Alvin M. Liberman, Dr. Franklin S. Cooper, Dr. Arthur Abramson, and Dr. Leigh Lister of the Haskins Laboratories, whose help was instrumental in launching this project. I would also like to thank Dr. Kenneth N. Stevens of the Massachusetts Institute of Technology, who lent support and guidance at the stage of completion of the research. The author's current address: University of Denver, Psychology Department, Denver, Colorado 80208.

address the following issues: If new phonetic distinctions are acquired, which are characteristic of the second language, what is (1) the temporal course of acquisition, (2) the relationship between changes in patterns of perception and patterns of production, and (3) the relationship between acquisition and the age of the learner?

Experimental Paradigm

The investigation described here focuses on the acquisition, in both perception and production, of the English voicing distinction by Puerto Rican Spanish children learning English as a second language. The choice of this phonetic contrast was motivated by the fact that a great deal is already known about cross-language differences in the acoustic-phonetic realization of the voicing contrast in Spanish and English (Abramson & Lisker, 1970, 1973; Lisker & Abramson, 1964; Williams, 1977a, 1977b). The voicing contrast is also known to cause difficulty for the second-language learner moving from either language to the other (Stockwell & Bowen, 1970).

In previous work it has been established that the property voice onset time (VOT), provides an acoustic variable sufficient to allow for the separation of voiced (e.g., [b], [d], [g])¹ from voiceless (e.g., [p], [t], [k]) homorganic stop consonants in word-initial position in many languages (Lisker & Abramson, 1964). Voice onset time refers to the duration between the release of consonantal articulation and the onset of glottal vibration, or voicing. By convention, voicing which begins prior to release, voicing lead, is assigned negative values in milliseconds and voicing which starts after release, voicing lag, is assigned positive values. Voice onset time can then be described in terms of discrete states of the glottis going from

voicing lead, onset prior to release, to voicing lag, onset following release.

It has been found that voiced and voiceless members of contrasting pairs (e.g., [b] vs. [p]) are produced by speakers of a given language with characteristic ranges of VOT (Williams, 1977b). These ranges and their mean values can differ across different language groups. For example, the range in VOT for a Spanish [b] in word-initial position is -180 to -30 msec (\bar{X} = -116 msec) and the range for a Spanish [p] in the same position is 0 to +60 msec (\bar{X} = +15.6 msec). English values for [b] are typically bimodal (Lisker & Abramson, 1964; Williams, 1974; Zlatin, 1974). In one body of data (Williams, 1974), two out of eight English speakers characteristically produced the [b] category with a degree of voicing lead (negative VOT values) ranging from -115 to -40 msec (\bar{X} = -85 msec). The other six speakers produced the voiced category with voicing lag (positive VOT values), ranging from 0 to +40 msec (\bar{X} = +12 msec). All speakers produced [p^h] with voicing lag, ranging from +20 to +115 msec (\bar{X} = +62.6 msec). Thus, each language keeps separate its voiced and voiceless stop categories in terms of VOT, although the distribution and means of voice onset time values can differ significantly across different languages, as in Spanish and English (Williams, 1977a).

Perceptual studies have shown, furthermore, that monolingual speakers of Spanish and English perceive members of a synthetic VOT continuum in a manner that coincides with their distribution of voiced and voiceless tokens in production. For example, within each language, division of a VOT continuum into phonemic categories /b/ and /p/, in a labeling task, typically occurs at a point lying between the distribution of VOT values for the two categories in production (Williams, 1977a, 1977b). This division occurs on the average for Spanish monolinguals at -4 msec and for English monolinguals at +25 msec VOT. The locations of maxima in discriminating pairs of stimuli drawn from a VOT continuum generally correspond with the division in labeling.

Given these facts of cross-language differences in the production and perception of voice onset time, the specific questions which this study addresses are the following:

(1) Within the approximate compulsory school age range in the United States (8-16 years), will formerly monolingual Spanish children learn to perceive voiced and voiceless word-initial stops, in terms of VOT, with a pattern similar to that of native speakers of English?

(2) If young second-language learners demonstrate acquisition of the voicing distinction in such manner, will this learning also be manifested in their ability to produce these categories like a native speaker?

(3) In the acquisition of new VOT values, characteristic of the English voicing contrast, will the age of the learner and the degree to which he has been exposed to English be systematically related to patterns of change?

METHOD

Subjects

The subjects in this experiment were 24 native English speaking children and 72 children who were native speakers of Spanish and had come from Puerto Rico to the mainland United States without prior speaking ability in or comprehension of English. All Spanish-speaking subjects were from small rural communities of Puerto Rico where little or no English is spoken and none had made return trips to the island since their arrival on the mainland. All of these children were enrolled in bilingual programs in schools of the Boston metropolitan area. However, they came from homes where only Spanish was spoken and lived in Spanish-speaking neighborhoods. English-speaking subjects were all natives of Boston, attending public school there. They were all monolingual in English, thus presenting the putative target behavior to which the Puerto Rican Spanish second-language learners were moving in their English-language performance. It was not possible to locate a similar group of monolingual Spanish-speaking Puerto Rican children.

The Spanish-speaking subjects were selected from three populations with differing amounts of exposure to English, based on their length of stay on the mainland: 0 to 6 months, 1½ to 2 years, and 3 to 3½ years. Each exposure group was further divided into two age groups: 8- to 10-year-olds and 14- to 16-year-olds. Twelve subjects were in each of the six experimental groups. Similarly, there were 12 English monolingual children in each of the same age groupings.

Materials

The experimental stimuli used in the perception task were drawn from the voice onset time (VOT) continuum of labial consonant-vowel (CV) syllables produced at Haskins Laboratories by Abramson and Lisker (1970, 1973). The acoustic characteristics of this stimulus set have been described elsewhere (Williams, 1977a). Audio tapes for labeling and discrimination tasks were recorded directly from the Haskins Laboratories speech synthesizer at 7½ rps. For labeling, four different random orders were prepared using 31 CV syllables ranging from -100 to +100 msec VOT with 4 sec between each stimulus and 10 sec after every block of 10. For discrimination, there were four random orders of all pairs, which differed by 20 msec VOT, within the -150 to +150-msec range. Each permutation of the two stimuli in the 34 pairs occurred an equal number of times. There was 1 sec between pair members, 4 sec between successive pairs, and 10 sec after every block of 10.

Procedure

Stimuli were presented by a Tandberg (6000 X) tape deck via matched earphones (Sharp 10). Because it was anticipated that some Spanish-speaking children would have difficulty comprehending English, all instructions were carried out in Spanish for the Spanish-speaking subjects by a native speaker of that language. English-speaking subjects were presented the tasks in English. Since it had been previously ascertained that the language of presentation does not form a perceptual set for language using this procedure (Williams, 1977a), it was hypothesized that presenting the tasks in two different languages for the two groups would not bias the results. A forced-choice procedure and pointing response were employed in both labeling and discrimination tasks. In labeling, /b/ or /p/ responses were indicated by the children's pointing to a picture of one of two objects or events whose

names begin with "b" or "p" ("baño" or "paño" for the Spanish and "barking" or "parking" for the English presentation). In discrimination, the subjects indicated whether or not the two members of a stimulus pair were the same or different by pointing to either a picture of two identical faces or to a picture of two different faces. Each subject participated in both labeling and discrimination sessions (counterbalanced), contributing to a total of 16 decisions per subject per stimulus or stimulus pair.

In the production task, all 72 Spanish-speaking subjects produced 16 examples of word-initial voiced and voiceless ([b], [p]) stop consonants in four different vowel environments ([ey], [a], [Ow], [Iy]). Eight examples were produced as single words and eight were embedded in a carrier sentence (see Williams, 1977a). Similar sets were obtained in both English and Spanish. All words were repeated by the subjects after a model utterance presented by either a native speaker of English or Spanish. This was done in order to overcome anticipated shyness on the part of the young Spanish subjects. It was hypothesized that an imitation task would tap the subjects' best performance, which could be used to assess changes which might be observed between groups. Production data were not initially obtained for English-speaking children. However, a partial set of such data were later collected in order to interpret a confound in the Puerto Rican Spanish results (see Results section).

Utterances were recorded at 7½ rps on a Sony (800-B) tape recorder using an exterior cardioid microphone (Sony, F26S). Words were recorded at the end of the final testing session, single words before sentences. 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 the place where 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 within formants. Measurements of VOT were first made in millimeters and converted into milliseconds, rounding off the nearest 5 msec.

RESULTS

Perception

In Figures 1 and 2, raw labeling scores are plotted by age as a function of the percent of /p/ judgments.

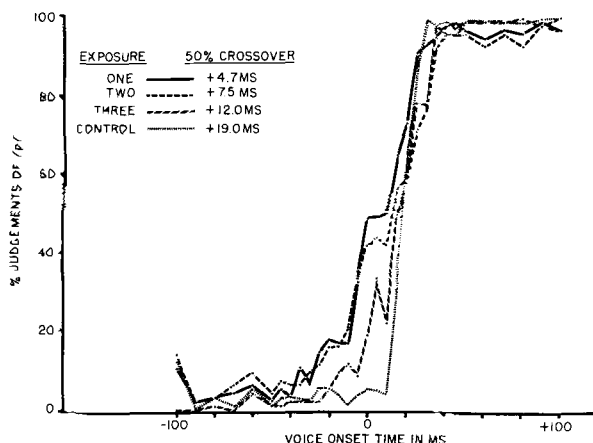


Figure 1. Pooled labeling functions for native Spanish-speaking children and English controls, 8 to 10 years old. Spanish children are represented by three exposure groups, according to time on the mainland United States: one, 0-6 months; two, 1½ to 2 years; three, 3 to 3½ years. The *n* for each data point per group is 96.

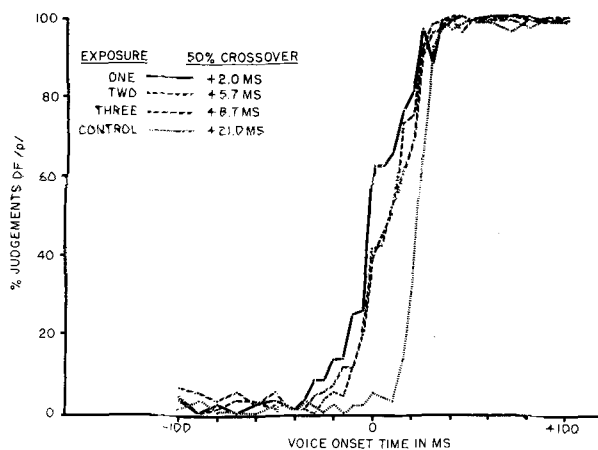


Figure 2. Pooled labeling functions for native Spanish-speaking children and English controls, 14 to 16 years old. Spanish children are represented by three exposure groups, according to time on the mainland United States: one, 0-6 months; two, 1½ to 2 years; three, 3 to 3½ years. The *n* for each data point per group is 96.

For the Spanish groups, this is further broken down by degree of exposure. Also shown are the mean crossover locations for each group after data conversion (see below). The point at which a function crosses the 50% location is conventionally used to define a labeling crossover between two phonemic (perceived) categories, in this case between the voiced and voiceless labial stops, /b/ and /p/.

Several observations can be made about the results. For Spanish groups of both ages, with increasing exposure to English, the locations of 50% crossovers move towards the average English-speaking children's crossover locations, at +19 and +21 msec VOT and evidence for change is found at the earliest period of exposure. There also appears to be an age effect, i.e., the movement seems greatest for youngest subjects. Although performance with monolingual Puerto Rican Spanish children was not obtainable, the crossover location of such a group would presumably be close to the average adult monolingual Puerto Rican Spanish crossover of -4 msec VOT (Williams, 1977b). Thus, movement towards the monolingual English crossover for children (which lies well into the positive VOT region) represents movement away from the likely location of a monolingual Spanish crossover location for children, i.e., in the negative region.

In order to test for the significance of age and exposure effects in the Spanish results, percentage scores from all groups were first converted by probit transformation (Finney, 1952) and crossover locations were taken from the best fitting straight line functions drawn to each subject's transformed data. Crossover values from individual subject's line functions were then submitted to a two-way analysis of variance,

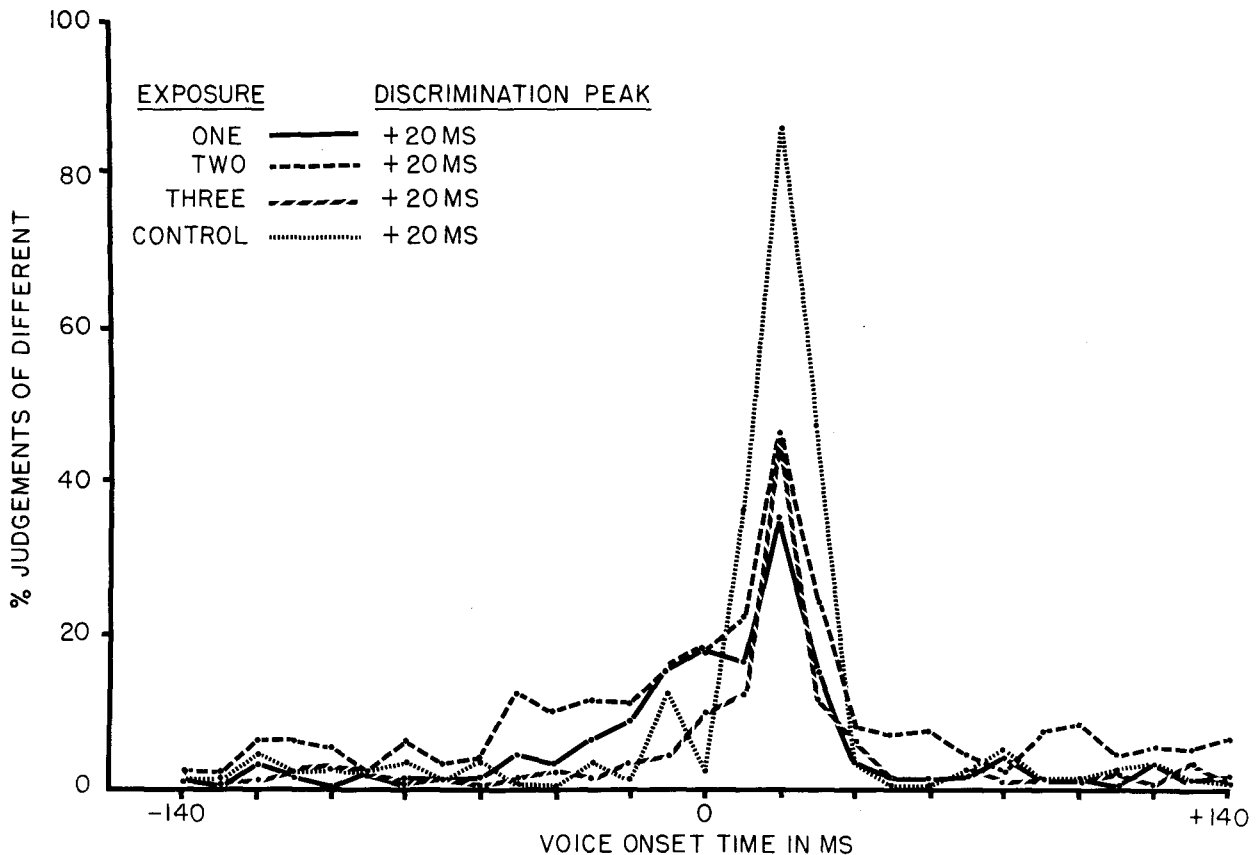


Figure 3. Pooled discrimination functions for native Spanish-speaking children and English controls, 8 to 10 years old. Spanish children are represented by three exposure groups, according to time on the mainland United States: one, 0 to 6 months; two, 1½ to 2 years; three, 3 to 3½ years. The *n* for each data point per group is 96.

age (2) by exposure (4), treating the English controls as a fourth exposure group. The results of this analysis showed significance for the effect of exposure [$F(3,88) = 18.62, p < .001$] but not for age or an Age by Exposure interaction. Specific post hoc comparisons revealed significant differences between Spanish and English groups [$F(1,88) = 48.27, p < .001$] and between Spanish exposure levels 1 vs. 2 and 3 [$F(1,88) = 5.53, p = .02$].

The shape of labeling functions for different groups is also informative. For the older Spanish groups (Figure 2), labeling functions are generally steep and monotonic, which is characteristic of the two English child groups as well as adult monolingual groups, tested with the same set of stimuli (Williams, 1977a). The functions for the younger subjects, on the other hand (Figure 1), show either plateaus or reversals in slope direction in the area shared by the monolingual Spanish and English crossovers. The question arises as to whether the three younger groups' labeling slopes take these forms simply as a result of averaging. For example, within each group, a portion of the subjects may have crossover locations closer to the Spanish monolingual location and another por-

tion of the subjects may have locations closer to that of English monolinguals such that the distribution of responses in each group would be bimodal. A cumulative function of averaged data of this kind could produce flattening or dipping of the slope in the area between the two true crossovers. On the other hand, individual subjects might be using a double, or bilingual, standard in making judgments, the outcome being that the distribution of their individual responses might be bimodal in the matter described above. Examination of the raw data indicates that both factors may be operating to some degree. However, the data are not extensive enough to tease out a clear bimodal distribution for either individuals or groups.

In Figures 3 and 4, discrimination data are plotted by age, and exposure for Spanish groups, as a function of the percent of "different" responses to pairs of stimuli. (In fact, all pair members differed by 10 msec VOT). The outstanding characteristic of the discrimination data is the degree to which the functions of Spanish groups coincide with those of English groups, peaking in all cases at +20 msec VOT, which is the exact location of the average English child

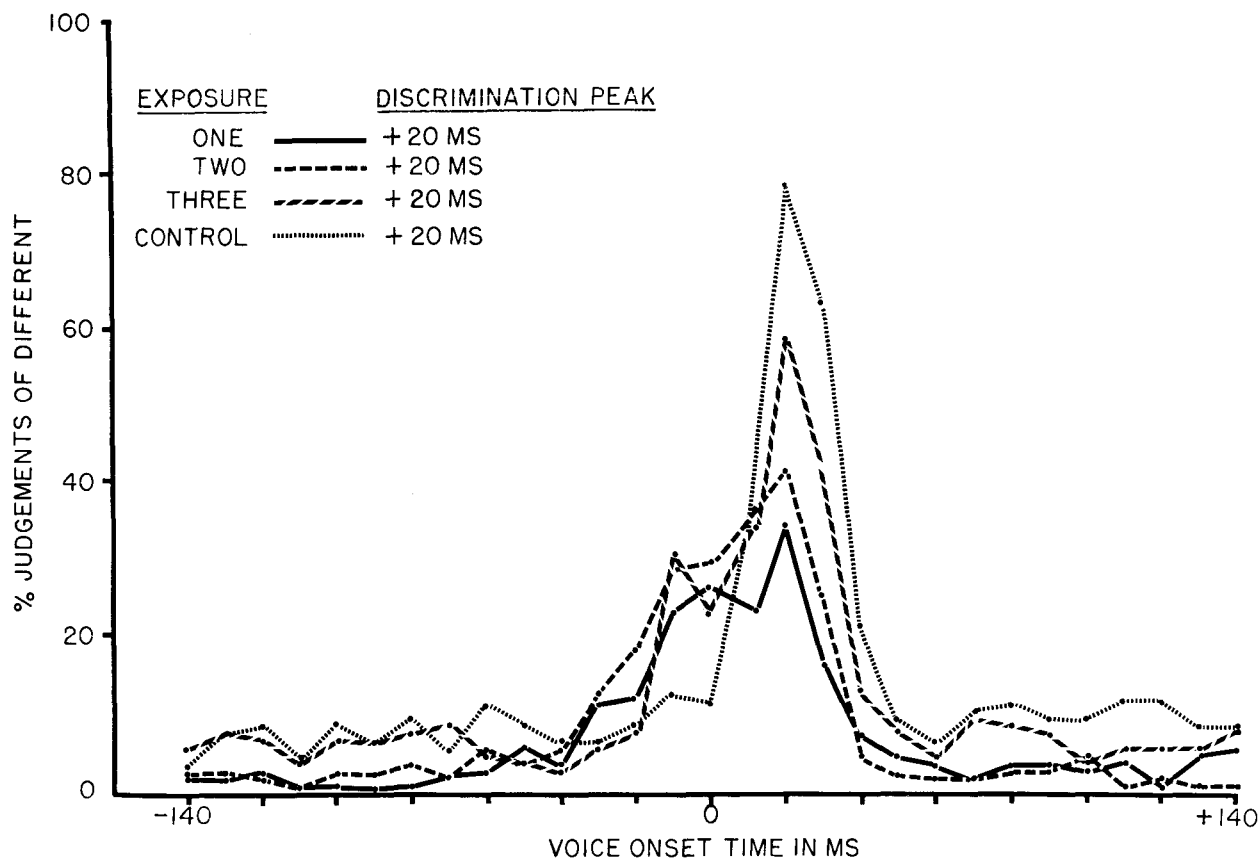


Figure 4. Pooled discrimination functions for native Spanish-speaking children and English controls, 14 to 16 years old. Spanish children are represented by three exposure groups, according to time on the mainland United States: one, 0 to 6 months; two, 1½ to 2 years; three, 3 to 3½ years. The *n* for each data point per group is 96.

labeling crossover. The functions for the three older groups also have secondary peaks falling near the average adult monolingual Spanish crossover location at -4 msec. For both age groups, there is the suggestion of an exposure effect, i.e., the greater the amount of exposure to English, the higher the peak, English-speaking children having the greatest amount of exposure.

In order to test for significance of some of these effects, all individual percent difference scores, across the peak areas, between -30 and $+40$ msec VOT, were submitted to a three-way multivariate analysis of variance (MANOVA), age (2) by exposure (4) by discrimination pair, as a repeated measure (8). Results showed significant effects of age [$F(8,80) = 2.323$, $p < .03$] and exposure [$F(8,80) = 2.86$, $p < .001$]. Separate post hoc comparisons revealed a significant difference only between the individual exposure levels English vs. all Spanish groups [$F(8,80) = 7.36$, $p < .001$]. Comparisons between English and Puerto Rican groups at several specific VOT discrimination pair levels were also significant [$F(1,87) =$ for 0 msec VOT, 10.09, $p < .002$; for +20 msec VOT, 39.8, $p < .001$; for +30 msec VOT, 29.92, $p < .001$]. The latter results

indicate that in the range bounding the monolingual English labeling crossover, from 0 to +30 msec VOT, discrimination of English-speaking children was at a significantly higher level than that of Spanish-speaking children. Discrimination of the range across the monolingual Spanish crossover from 0 to -20 msec VOT, on the other hand, was superior for the Spanish- vs. English-speaking children, reflected in the secondary peaks of discrimination for the older Spanish children. Those differences were not significant, however.

Production

Voice onset time values in milliseconds, derived from spectrograms of the production data in English and Spanish, were subjected to a five-way repeated measures analysis of variance, with age (2) and exposure (3) as between-subject variables and phonetic category ([p] or [b]), language (English or Spanish), and frame (produced as single words or in sentences) as within-subject variables. The significant results of this analysis are presented in Table 1. As can be seen in the table, all of the main effects were significant at the .01 level or better. The pattern of results can be summarized as follows: Phonetic category—In both

Table 1
Analysis of Variance Table of Significant Results

Source	df	MS	F
Between			
Age	1	21,693.29	15.57***
Exposure	2	7,375.25	5.29**
Error	56	1,393.27	
Within			
Frame	1	15,871.85	33.75***
Error	56	470.25	
Language	1	31,242.68	47.05***
Error	56	663.97	
Phonetic Category	1	786,532.44	853.79***
Error	56	921.27	
Phonetic Category by Frame	1	46,462.27	96.40***
Error	56	482.00	
Age by Exposure by Frame	2	1,526.85	3.25*
Error	56	470.25	
Phonetic Category by Language by Frame	1	3,534.53	8.06**
Error	54	438.60	
Age by Exposure by Phonetic Category by Frame	2	5,128.86	10.64***
Error	56	482.00	

* $p < .05$; ** $p < .01$; *** $p < .001$.

languages tokens of [b] are produced with negative VOT values (voicing lead) and tokens of [p] ([p], Spanish or [p^h], English) with positive VOT values (voicing lag). Language—The amount of voicing lead of [b] is greater in Spanish and the amount of voicing lag of [p] is greater in English. Frame—Placing a word in a sentence frame reduces the amount of VOT for both lead (negative) and lag (positive) values. Exposure—There is generally a change in VOT for [b] and [p] tokens towards values more appropriate for English as a function of exposure to English; this occurs for both English and Spanish productions. Age—The exposure effect is greatest for the younger group.

Statements about all of the main effects must be interpreted in terms of several significant interactions. A significant Phonetic Category by Frame interaction reflects the fact that the reduction in VOT from single word to sentence frame operates to a greater extent with [b] than with [p]. This effect interacts again with language as the reduction is greater in English than in Spanish. When the effects of age, phonetic category, and frame are examined across exposure, a somewhat complicated picture emerges. At exposure level 2, a characteristic of the data can be identified which contributes heavily to the interactions found, i.e., Age by Exposure by Frame and Age by Exposure by Phonetic Category by Frame. When the group means for phonetic category are plotted by frame, age, and language across exposure, curvilinear functions typically emerge. This contrasts with either a linear function or curvilinear function of reverse direction at another level. This can be found in the data for both age groups at both levels of exposure and is attributable to what appears to be a high degree of variability in the data for the second exposure

group. When exposure level 2 is disregarded and exposure levels 1 and 3 are compared, the picture is somewhat simplified. This is illustrated in Figures 5 and 6. Clear differences across phonetic category, frame, and age exist at both exposure levels with a consistent trend in changing VOT, i.e., decreasing lead and increasing lag.

A comparison of the results of this study and those from a study of Spanish and English adult monolinguals in which the same set of words were used produces the following generalizations:

(1) Both of the young second-language-learner age groups use less voicing lead in producing Spanish [b]s and a greater amount of voicing lag in producing Spanish [p]s than adult Spanish monolinguals, and these differences are greater for the younger groups.

(2) Both second-language-learner age groups use a greater amount of voicing lead when producing English [b]s and less voicing lag when producing English [p]s than do monolingual English speakers and the differences are greater for the older group.

These two findings may be restated as an increase over time in English-like VOT characteristics when second-language learners are producing Spanish words and a decrease in Spanish-like VOT characteristics when they are producing English words. Both changes interact with age.

The age effect was particularly interesting in this set of results since, in conjunction with the direction of change in the production of English words across exposure, it suggests that younger children are more likely to “learn” a second language faster than older children, at the phonetic level. However, age and exposure were confounded in this sample, i.e., it could be that age in conjunction with exposure to English

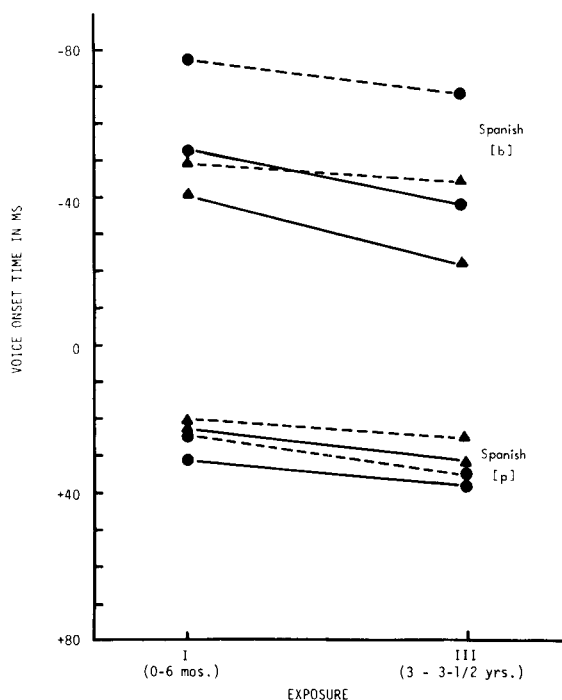


Figure 5. Average VOT values for Spanish utterances with initial [b] and [p] for exposure groups I and III. Age and frame variables are represented as: 8 to 10 years, single words, ●—● and sentences, ▲—▲; 14 to 16 years, single words, ●---● and sentences ▲---▲.

accounts for the more English-like VOT values of the younger Spanish-speaking children or it could be that age alone is critical. Younger children may produce VOT values closer to the English values for physiological reasons, independent of exposure to English. The absence of a significant Age by Exposure interaction makes it difficult to determine which was operating. Therefore, additional production data were collected from monolingual English children of the Boston area, using only words in isolation. The words were recorded and measured in the same manner as done with the Spanish-speaking children, using the same two age groupings, with 12 children in each group.

It is of interest that Spanish words spoken by English-speaking children were also sought, but the data were not collectible because these children were too reluctant to imitate Spanish. It was hypothesized that finding age to be nonsignificant for native English-speaking children, producing English words, but for age to have a significant effect for Spanish-speaking children, producing English, would support the interpretation that age is important only in the context of learning a second language.

Measured VOT values from the English single words, produced by both English- and Spanish-speaking children were entered into a two-way analysis of variance, age (2) by language group (2). The

results showed significance for the overall age factor [$F(1,65) = 5.26, p = .025$] and language group factor [$F(1,65) = 69.74, p < .001$], with no significant Age by Language Group interaction. However, post hoc contrasts showed that the age comparison within the native Spanish-speaking groups was significant [$F(1,65) = 7.24, p < .009$] but not the comparison within the native English-speaking groups.

DISCUSSION

The results of the experiments reported here provide concrete evidence for perceptual learning at the phonemic and phonetic levels during second-language acquisition. This learning may be manifested directly in the form of perceptual performance or indirectly as changes in speech sound production, which presumably ensue from changes in the perception of acoustic-phonetic material. The fact that changes in the direction of monolingual English performance was observed for native Spanish-speaking children in both labeling and discrimination tasks, with minimal exposure, suggests that very little experience may be required to activate this form of perceptual learning.

There was only slight evidence from individual or group data, in either labeling or discrimination results, to support the view that the children tested were retaining a Spanish-like standard while acquiring one

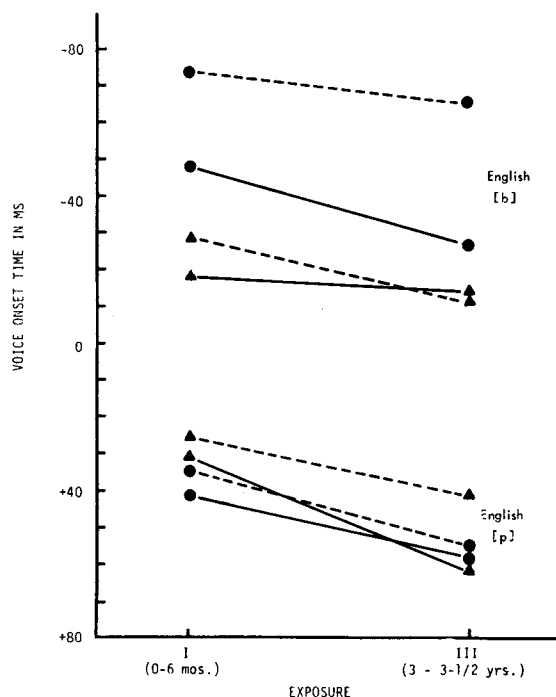


Figure 6. Average VOT values for English utterances with initial [b] and [p] for exposure groups I and III. Age and frame variables are represented as: 8 to 10 years, single words, ●—● and sentences, ▲—▲; 14 to 16 years, single words, ▲---▲ and sentences ●---●.

more English-like. The only support for this position was the occurrence of plateaus and slope reversals in the labeling data for younger groups and the secondary peaks in discrimination over the Spanish voicing boundary for the older groups. Several possible explanations for these findings can be considered. The absence of strong evidence for the development of a double standard in perceptual performance suggests that a learner may choose some compromise value along an acoustic-phonetic dimension when his native and second language divide this dimension at different locations. If this were true, however, it would seem difficult for the learner to keep separate, in perception, the two sets of contrasting acoustic-phonetic tokens which underlie the same phonemic distinction in his two languages. For example, if an individual learner's crossover in labeling this series occurred at +10 msec VOT, then voicing-lead stops (i.e., negative VOT) as well as short lag stops (e.g., +5 msec VOT) would all fall into the *perceived* voiced (/b/) category, even though Spanish voiceless ([p]) stops are frequently *produced* with from 0 to +5 msec VOT. Thus, Spanish voiced and voiceless stops might be confused by this listener, if VOT constituted the only cue to initial stop voicing.

It is possible that the task itself, requiring a binary decision, forced the listeners into a two-category division of the continuum, eliminating evidence for a double division. However, the discrimination task did not present such a constraint. The discrimination results still show an absence of discrimination maxima at VOT values near the monolingual Spanish boundary, even though adult Spanish monolinguals do show peak discrimination in the prevoiced region (Williams, 1977b).

Perhaps, during the process of acquiring a second language, the learner becomes highly sensitized to acoustic-phonetic information relevant to his second language and temporarily less sensitive to phonetic information relevant to his primary language. Such a process could account for the lack of evidence for both Spanish- and English-like patterns of perception.

A final, and more plausible, explanation than those presented above is provided when one considers what is known about the acoustic characteristics of naturally spoken voiced and voiceless stops in Spanish and English. In both languages, several acoustic properties seem to differ systematically across initial stop voicing contrasts in addition to VOT and its correlate properties. In English, for example, there are differences in the spectral characteristics of the brief burst of aperiodic noise associated with the articulatory release of the stop (Halle, Hughes, & Radley, 1957; Klatt, 1975; Zue, Note 1), and differences in patterns of voice fundamental frequency change following release (Ewan & Kronen, 1974; House & Fairbanks, 1953; Lea, 1973). Differences in

postrelease fundamental frequency change have been shown to provide perceptual cues for the voicing with English-speaking listeners (Fujimura, 1971; Haggard, Ambler, & Callow, 1970). It has been demonstrated that when voicing lead is removed from voiced stops which initiate words spoken in Spanish, a relatively high-amplitude burst and the presence of low-frequency periodic energy following release favor "voiced" judgments by Spanish listeners (Williams, 1977b). A relatively high-amplitude burst and silence for 5-10 msec following release, on the other hand, favor "voiceless" judgments. These results suggest that information supplied by acoustic properties at the point of release, as well as voicing lead, may provide voicing cues for Spanish listeners.

Acoustic cues such as the above *may be* more perceptually salient than the relatively low-energy cue provided by differences in voicing lead. None of the above properties are available as voicing cues in the synthetic continuum used in these studies. Perhaps the second-language learner comes to rely more heavily on additional cues, such as these, as he changes from being a Spanish monolingual to a Spanish-English bilingual. If this were the case, the results obtained could be predicted, i.e., an absence of clear evidence for a developing double standard in perception (Spanish-like and English-like) using test materials lacking these potential cues.

The evidence that acoustic-phonetic properties which discriminate the English voicing contrast are available to the young second-language learner with little exposure to English suggests that this contrast may be a naturally distinctive one, perhaps due to psychoacoustic or auditory factors. A similar conclusion was expressed by Streeter and Landauer (1976) in the interpretation of their finding that Kikuyu children were able to discriminate VOT contrasts lying in the region of the English contrast (i.e., +10 vs. +40 msec VOT) without experience with that acoustic-phonetic contrast. It is noteworthy that the strongest evidence for infant discrimination of synthetic stop consonants that contrast in VOT also occurs for contrasts which bracket the English phoneme boundary (Eilers, Gavin, & Wilson, 1979; Eimas, 1975; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Trehub & Rabinovitch, 1972). Infant discrimination of degrees of voicing lead have been difficult to demonstrate (Eilers et al., 1979; Lasky, Syrdal-Lasky, & Klein, 1975; Streeter, 1976), except for infants from a language in which voicing lead is phonemic. Evidence for discrimination of contrasts in the VOT vicinity of maximum English discrimination has also been obtained with the chinchilla (Kuhl, Note 2), suggesting a nonlinguistic basis for this pattern of selective discrimination. Nonspeech stimuli, modeling the synthetic-speech VOT continua, have produced labeling and discrimination data with English listeners

which look similar to the results obtained using the synthetic-speech VOT series (Stevens & Klatt, 1974; Miller, Weir, Pastore, Kelley, & Dooling, 1976; Pisoni, 1977). These results suggest that the basis for the perception of VOT with speech-like stimuli may be, at least in part, psychoacoustic in nature.

It may not even be necessary to search beyond an analysis of the acoustic characteristics of the voicing continuum used in this study to find an explanation for the apparent superiority of discrimination of VOT contrasts drawn from the +25 msec VOT region. In the synthetic series used in the studies of voicing perception cited above, a single property, voice onset time, varies in the voicing lead region (negative VOT). By contrast, a number of properties undergo parametric change in the VOT region bracketing approximately +25 msec VOT. These are also present in the naturally produced versions of these stimuli: (1) The presence, absence, or varying duration of aspiration or aperiodic energy in the interval between articulatory release and the onset of voicing. (2) The absence of periodic acoustic energy at the level of formant one during aperiodic excitation of the vocal tract. (3) Differences in the duration and amount of frequency change of formant transitions under conditions of periodic excitation of the vocal tract.

In a discrimination task, the young second-language learner may have immediate access to this complex of contrasting properties, providing for maximal discrimination of the relevant portion of the VOT continuum. The results of discrimination in this study support this conclusion. However, to use these potential voicing cues in order to organize phonetic distinctions into two separate systems, appropriate for each language, may require further experience with the second language. This notion is supported by the labeling results of this study which show gradual change with exposure among second-language learners.

The modification from Spanish- to English-like performance observed in the perception results corresponds to a similar modification in production. This modification affects not only the production of the newly acquired language, but also the production of the native language, in terms of VOT characteristics. Thus, the young second-language learner gradually loses his Spanish-like manner of producing English words while (s)he develops interference from the second language on the primary language. The fact that these results were based on utterances which were repetitions of a model presented by a native speaker strengthens this finding. In imitation one would expect speakers to attempt correct pronunciation, avoiding the encroachment of the phonetic effects of one language upon the other. Of course, it remains to be determined whether or not the observed degree of deviation from monolingual values of VOT

constitute differences sufficient to be detected as a loss of Spanish accent in producing English, on the one hand, and an acquisition of an English accent in producing Spanish, on the other.

Of further interest would be information about the patterns of an individual speaker's progress in modifying voicing characteristics in second-language learning. For example, is the pattern of change linear or are there periods of acquisition interspersed with plateaus in progress? The variability in performance observed in the second exposure period, from 1½ to 2 years, may in fact reflect a period of individual deviation in performance from an otherwise linear process of change. A definitive answer to the question relating to individual patterns of change in production would, of course, be obtained best from a longitudinal investigation of second-language learning. It would also be of interest to determine the exact relationship between changes in perception and production, again an issue best investigated by a longitudinal study of individual learners.

To summarize, the results of the present investigation show that the young second-language learners modify, over time, their perception and production of phonetic material relevant to both languages. There is a tendency for this process to occur more rapidly for younger children, although the effect of age was not reliable in the results of perceptual tests reported here.

The apparent development of compromise VOT targets in both perception and production for second-language learners may reflect a true convergence over time of the acoustic-phonetic features of the two languages instead of the development of two separate phonetic systems. However, what I would like to hypothesize is that what occurs in the process of second-language acquisition is rather a restructuring of the acoustic-phonetic space that encompasses both languages. This restructuring may involve a developing sensitivity to supplemental acoustic cues which will enable bilinguals to keep separate, in perception, the phonetic systems of their two languages. It may also entail, in production, a realignment of the total phonetic space of any single acoustic-phonetic dimension, such as VOT. Presumably, however, if a speaker is to become a fluent bilingual, without the phonetic encroachment of either language upon the other, such a change would not eliminate either the acoustic-phonetic distinction of contrasts within each language or the distinctions of the phonetic systems across the two languages.

REFERENCE NOTES

1. Zue, V. W. Acoustic characteristics of stop consonants: A controlled study. *Lincoln Laboratory Technical Report 523*, Lincoln, Massachusetts, May 1976.
2. Kuhl, P. *Predispositions for the perception of speech-sound*

categories: *A species-specific phenomenon*. Paper presented at the Conference on Early Behavioral Assessment of the Communicative Cognitive Abilities of the Developmentally Disabled, Orcas Island, Washington, May 2-6, 1976.

REFERENCES

- ABRAMSON, A., & LISKER, L. Discrimination along the voicing continuum: Cross-language tests. *Proceedings of the Sixth International Congress of Phonetic Sciences, 1967*. Prague: Academia, 1970.
- ABRAMSON, A., & LISKER, L. Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, 1973, **1**, 1-8.
- ASHER, J., & GARCIA, L. The optimal age to learn a foreign language. *The Modern Language Journal*, 1969, **53**, 336-341.
- EILERS, R., GAVIN, W., & WILSON, W. Linguistic experience and phonemic perception in infancy: A cross-linguistic study. *Child Development*, 1979, **49**, 14-18.
- EIMAS, P. D. Speech perception in early infancy. In L. Cohen & P. Salapatek (Eds.), *Infant perception from sensation to cognition II*. New York: Academic Press, 1975.
- EIMAS, P. D., STQUELAND, E. R., JUSCZYK, P., & VIGORITO, J. Speech perception in infants. *Science*, 1971, **171**, 303-306.
- EWAN, W. G., & KRONES, R. Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics*, 1974, **2**, 327-335.
- FINNEY, D. *Probit analysis: A statistical treatment of the sigmoid response curve*. Cambridge, England: University Press, 1952.
- FUJIMURA, O. Remarks on stop consonants: Synthesis experiments and acoustic cues. In L. L. Hammerich, R. Jacobson, & E. Zwirner (Eds.), *Form and substance: Phonetic and linguistic papers presented to Eli Fischer-Jorgensen*. Denmark: Akademisk Forlag, 1971.
- HAGGARD, M. S., AMBLER, S., & CALLOW, M. Pitch as a voicing cue. *Journal of the Acoustical Society of America*, 1970, **47**, 613-617.
- HAKUTA, K., & CANCINO, H. Trends in second-language-acquisition research. *Harvard Educational Review*, 1977, **47**, 294-316.
- HALLE, M., HUGHES, C. W., & RADLEY, J. P. A. Acoustic properties of stop consonants. *Journal of the Acoustical Society of America*, 1957, **29**, 107-116.
- HOUSE, A. S., & FAIRBANKS, G. The influence of consonantal environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, 1953, **25**, 105-113.
- KLATT, D. Voice onset time, frication and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 1975, **18**, 686-706.
- LASKY, R., SYRDAL-LASKY, A., & KLEIN, R. VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, 1975, **20**, 215-225.
- 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)
- LISKER, L., & ABRAMSON, A. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, **20**, 384-422.
- MILLER, J., WIER, C., PASTORE, R., KELLEY, W., & DOOLING, R. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, **60**, 410-417.
- OYAMA, S. A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, 1976, **5**, 261-284.
- PISONI, D. Identification and discrimination of the relative onset times of two component tones: Implications for voicing perceptions in stops. *Journal of the Acoustical Society of America*, 1977, **61**, 1352-1361.
- SEGALOWITZ, N. Psychological perspectives on bilingual education. In B. Spolsky & R. L. Cooper (Eds.), *Frontiers in bilingual education*. Rowley, Mass: Newbury House, 1976.
- 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: University of Chicago Press, 1970.
- STREETER, L. Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature*, 1976, **259**, 39-41.
- STREETER, L. A., & LANDAUER, T. K. Effects of learning English as a second language on the acquisition of a new phoneme contrast. *Journal of the Acoustical Society of America*, 1976, **59**, 448-451.
- TREHUB, S. E., & RABINOVITCH, M. S. Auditory-linguistic sensitivity in early infancy. *Developmental Psychology*, 1972, **6**, 74-77.
- WILLIAMS, L. *Speech perception and production as a function of exposure to a second language*. Unpublished PhD dissertation, Harvard University, Cambridge, Massachusetts, 1974.
- WILLIAMS, L. The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, 1977, **21**, 289-297. (a)
- WILLIAMS, L. The voicing contrast in Spanish. *Journal of Phonetics*, 1977, **5**, 169-184. (b)
- ZLATIN, M. Voicing contrast: Perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America*, 1974, **56**, 981-994.

NOTE

1. Square brackets are used to illustrate phonetic distinctions in production and slashes to illustrate the perceived difference between two phonetic distinctions. Thus, the phonetic distinction [b] vs. [p^h] in English is perceived as the phonemic distinction /b/ vs. /p/. The phonetic distinction underlying the same phonemic contrast in Spanish is [b] vs. [p] (unaspirated p). For simplicity, the word phonetic is used in discussing a general class of related events when both production or perception are involved. The term "acoustic-phonetic" relates directly to the acoustic properties underlying a phonetic distinction.

(Received for publication November 3, 1977;
revision accepted May 12, 1979.)