Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation

James Emil Flege
Department of Biocommunication, University of Alabama at Birmingham, University Station, Birmingham, Alabama 35294

Wieke Eefting
Institute of Phonetics, University of Utrecht, Utrecht, The Netherlands

(Received 11 September 1986; accepted for publication 14 October 1987)

This study examined imitation of a voice onset time (VOT) continuum ranging from /da/ to /ta/ by subjects differing in age and/or linguistic experience. The subjects did not reproduce the incremental increases in VOT linearly, but instead showed abrupt shifts in VOT between two or three VOT response "modes." The location of the response shifts occurred at the same location as phoneme boundaries obtained in a previous identification experiment. This supports the view that the stimuli were categorized before being imitated. Children and adults who spoke just Spanish generally produced only lead and short-lag VOT responses. English monolinguals tended to produce stops with only short-lag and long-lag VOT values. The native Spanish adults and children who spoke English, on the other hand, produced stops with VOT values falling into all three modal VOT ranges. This was interpreted to mean that they had established a phonetic category {t^h} with which to implement the voiceless aspirated realizations of /t/ in English. Their inability to produce English /p,t,k/ with the same values as native speakers of English must therefore be attributed to the information specified in their new English phonetic categories (which might be incorrect as the result of exposure to Spanish-accented English), to partially formed phonetic realization rules, or both.

PACS numbers: 43.70.Fq, 43.70.Hs, 43.70.Kv

INTRODUCTION

One aim of speech production research is to understand how talkers physically realize in time and space the phonological units (phonemes) that distinguish words. It is generally agreed that the information specified at a phonemic level of representation is less detailed than the motoric codes used to guide articulatory movements. Developmental studies indicate that it takes children several years to establish the motor skills needed to realize phones in a mature fashion (Flege and Eefting, 1986; Flege *et al.*, 1987). The present study was concerned with the speech learning that takes place in second-language (L2) acquisition.

Flege and Eefting (1987a) found that two groups of native Spanish adults and a group of 9- to 10-year-old children realized English /p,t,k/ with voice onset time (VOT) values that were intermediate to those observed for age-matched Spanish and English monolinguals (see also Caramazza *et al.*, 1973; Major, 1987). The seeming inability of these subjects to produce English stops authentically was probably not due to their having passed a "critical period" for L2 learning (see Flege, 1987b), for the subjects in all three groups had begun learning English by 6 years of age.

Two broad hypotheses exist concerning the cause of the difference between the native and non-native speakers. Flege (1987a,c) hypothesized that L2 learners whose native language (L1) realizes /p,t,k/ with short-lag VOT values show "compromise" VOT values in producing English /p,t,k/ because they do not establish phonetic categories for English stops. It was hypothesized that L2 learners judge tokens of [p,t,k] in their L1 and tokens of [p^h, t^h, k^h] in their L2 to be realizations of the same phonetic categories even though they can detect auditorily the acoustic differences between corresponding L1 and L2 stops. Differences in how they realize /p,t,k/ in L1 vs L2 might be attributed to the establishment of realization rules for producing stops in L2. Based on differences in how L2 learning influences the production of stops in L1, Flege (1987d) later hypothesized that individuals who begin learning English L2 as young children *do* establish separate phonetic categories for English /p,t,k/. The present study used an imitation task to test this hypothesis for the native Spanish subjects whose production and perception of English stops were examined previously (Flege and Eefting, 1987a).

The phonological and phonetic differences between Spanish and English are illustrated in Fig. 1 in terms of the model of speech production underlying the two alternative hypotheses just presented. (The terminology and symbols are derived from Keating, 1984). The model specifies three levels of organization: phonemic, phonetic, and motoric. Three universal phonetic categories having distinct patterns of laryngeal timing are available for implementing stop phonemes: voiced {d}, voiceless unaspirated {t}, and voiceless aspirated {t^h}.

Spanish and English use different phonetic categories to implement the contrast between /t/ and /d/. In Spanish, {d} is used to implement /d/ and {t} implements /t/. The realization rules used in Spanish to output the phonetic cate-

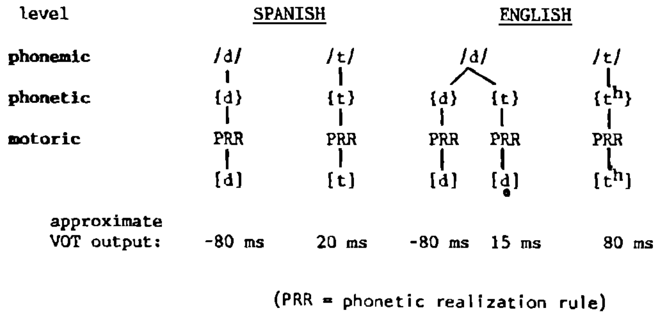

FIG. 1. Illustration of a model of speech production that relates phonological representations to phonetic output.

gories {d} and {t} yield stops with VOT values of approximately − 80 ms and 20 ms, respectively, in word-initial prestressed position.[1] In English, /d/ is implemented by either {d} or {t}, and /t/ is implemented by {tʰ}. The realization rules used in English to output {d} and {t} result in VOT values of about − 80 ms and 20 ms. The rule used to implement {tʰ} yields VOT values of approximately 80 ms (Williams, 1977a; Lisker and Abramson, 1964; Flege and Eefting, 1986).

Phonetic realization rules are needed in a model of speech production to account for talker's ability to modify articulation (e.g., when speaking louder, more rapidly, more clearly, or with greater emphasis). They are also needed to account for cross-language differences in how universal phonetic categories are realized. For example, small but systematic differences in absolute VOT values distinguish the voiceless aspirated stops found in Saudi Arabian Arabic, English, and Danish (Flege and Port, 1981; Christensen, 1984). An intermediate phonetic level of representation between phonemes, which are specified by distinctive features, and phonetic realization rules, which are specified in terms of motor codes, is motivated by differences in how English /d/ may be realized. Most native speakers of English realize /b,d,g/ with both lead and short-lag VOT values (Lisker and Abramson, 1964, 1967; Flege and Massey, 1980), suggesting that English word-initial stops may be produced with vocal fold adduction near the beginning of the stop closure interval, or at about the time of stop release.[2]

Imitation is generally regarded as consisting of three distinct processes: perception of structural properties in the stimuli being imitated, coding and storage in memory, and regeneration in the form of a motoric code suitable for skilled movement. There is evidence that subjects' categorization of stop consonants prevents them from imitating variations in VOT accurately.[3] Yeni-Komshian *et al.* (1977) examined imitation by monolingual adult native speakers of Lebanese Arabic, a language in which /t/ is realized with short-lag VOT values of about 20 ms and /d/ with lead VOT values of about − 60 ms. None of the eight subjects showed a linear increase in VOT corresponding to the incremental increases in VOT in a continuum ranging from /da/ to /ta/. Seven subjects showed a discontinuity in their responses at a location which seemed to correspond to the phoneme boundary between Arabic /d/ and /t/. The discontinuities consisted of a large increase in VOT, a change from stops to fricatives, a change in place of articulation (/da/ to /ka/), or a change in syllable structure (from CV to V).

The remaining adult subject examined by Yeni-Komshian *et al.* (1977) produced syllables beginning with a fricative in response to stimuli with − 120- to 0-ms VOT, syllables initiated by an apical or velar stop for the 10- to 50-ms stimuli, and /ha/ in response to the 60- to 140-ms stimuli. This suggested that although he was unable to realize stops reliably with VOT values in the three "modal" ranges characteristic of the three universal phonetic categories, he recognized the phonetic distinction between voiced and voiceless unaspirated stops, as well as that between voiceless unaspirated and aspirated stops (a distinction not found in Lebanese Arabic). Children aged 3 to 6 years of age may be more likely than adults to recognize the existence of three phonetic categories along a VOT continuum. Yeni-Komshian *et al.* (1968) found that, like adults, most native English children showed a shift in response mode at a single location along the continuum (from short-lag to long-lag VOT, or from short-lag stops to isolated vowels). However, four of ten Arabic children showed response shifts at two locations along the VOT continuum.[4]

The results obtained by Yeni-Komshian and her colleagues (Yeni-Komshian *et al.*, 1968, 1977) suggest that, after categorizing the stimulus to be imitated, subjects tend to reproduce the stimulus with syllables found in their phonetic repertoire. Native speakers of Spanish may need to establish a long-lag phonetic category {tʰ} before beginning to establish realization rules suitable for producing English /t/ with appropriate VOT values. If the Spanish subjects examined by Flege and Eefting (1987a) do not produce long-lag VOT values when imitating long-lag stimuli in the present experiment, it would suggest they produced English /p,t,k/ with VOT values that were too short because they had not: (a) established a phonetic category {tʰ} for implementing English /t/, or (b) formed a realization rule for outputting {tʰ}. If they do produce responses with long-lag VOT values in response to the long-lag stimuli, it would suggest that individuals who learn English as an L2 by the age of 5–6 years are capable of forming phonetic categories for L2 phones and establishing phonetic realization rules for outputting their new L2 categories. This, in turn, would suggest that the Spanish subjects who produced English /p,t,k/ with VOT values that were too short did so because their phonetic categories were inaccurate, perhaps due to their hearing Spanish-accented English (in which /p,t,k/ are produced with VOT values that are shorter than in nonaccented English; see Flege and Hammond, 1982).

There is some controversy concerning the extent to which imitation responses bypass the process of phonetic categorization. Mynah birds can reproduce human speech sounds without processing them phonetically (Marler and Mundinger, 1971). Chistovich *et al.* (1966) hypothesized that imitation partially bypasses categorization. This is consistent with the finding that Broca's aphasics may imitate sounds more correctly than produce them spontaneously (Trost and Canter, 1974). Yeni-Komshian *et al.* (1968), however, found that native English children tended to show a response shift between the 30- and 40-ms stimuli in a /da/ and /ta/ continuum, while the native Arabic children showed response shifts at shorter VOT values. The apparent cross-language difference in the location of the response discontinuity suggested that the child subjects had identified the stimuli covertly before reproducing them. Since children's imitations might be expected to show the influence of categorization to a lesser extent than adults' (Garnica and Edwards, 1977; Leonard *et al.*, 1978; Barton, 1978; Klein, 1979), one might expect the same to be true for adults' imitations.

The location of response discontinuities in each subject's imitation responses was determined in the present study. If imitation is influenced covertly by categorization, the locations of the discontinuities should not differ from the category boundaries obtained in the forced-choice identification experiment reported by Flege and Eefting (1987b). If children's imitation is shaped less by categorization than adults', this might be true for the adult but not child subjects.

I. METHODS

A. Subjects

Seven groups of subjects differing in age and/or linguistic experience were formed, each consisting of five male and five female subjects with normal hearing and speech (according to self-report or that of their teacher). Two groups consisted of monolingual native speakers of English. Group EA ("English adults") consisted of adults affiliated with the University of Alabama at Birmingham whose mean age was 26 years. Group EC ("English children") consisted of 9-to10-year-old fourth graders from a parochial school in Birmingham.

The subjects in three groups were native speakers of Spanish who had begun learning English as an L2 before the age of 6 years. The subjects in all three groups spoke English with a detectable Spanish accent in the authors' opinion, although degree of accent was not quantified. The subjects in group BC ("bilingual children") were 8- to 9-year-old Puerto Rican children who had been attending a private English-speaking school in Mayaguez, Puerto Rico, for an average of 3.6 years at the time of the study. The parents of these children were native speakers of Spanish from Puerto Rico who spoke to their children in Spanish at home.

The young adult subjects in group LCB ("later childhood bilinguals") were born and raised in Puerto Rico and had never lived in a place where English was the dominant language. They had enrolled in a school where English was the language of instruction (like the subjects in BC) at the age of 5–6 years, and remained there for an average of 7.1 years. The subjects in group ECB ("earlier childhood bilinguals") had the same mean age (19 years) as those in group LCB. Like the subjects in group LCB, they had attended a Spanish-speaking public high school in Puerto Rico (with one exception in both groups), and were attending a Spanish-speaking university at the time of the study. Unlike the subjects in LCB, they had been born on the mainland U.S. (or had been taken there shortly after birth), where they stayed for an average of 9.7 years and attended public school for 6.4 years. Although the subjects in ECB indicated that Spanish was the language usually spoken in their home, one parent of two subjects was a native speaker of English. Perhaps because of their somewhat earlier and probably more massive exposure to English, three subjects in ECB regarded English as their first language.

Two groups of subjects consisted of native speakers of Spanish who had never lived outside Puerto Rico and could not speak English. It appears that these subjects had received little exposure to English, for TV and radio programs are broadcast exclusively in Spanish in Mayaguez. Subjects in group SA ("Spanish adults") were engineering students at the University of Puerto Rico in Mayaguez with a mean age of 20 years. They will be referred to here as "monolingual" for the sake of convenience, even though they had studied English as a foreign language in school for 12 years. Their exposure to English seems to have been quite limited since most of their formal instruction in English focused on grammar and reading, and came from native speakers of Spanish. The adults in SA were not able to engage in a simple English conversation prior to the experiment with one of us who does not speak Spanish (JEF). The subjects in group SC ("Spanish children") were 9-year-old fourth graders from a public school who had received 1 year of English-language instruction (nominally 1 h per day) from a non-native speaker of English. These children could not respond in English to questions such as "What is your name?".

B. Stimuli

A 16-member continuum ranging from /da/ to /ta/, which has been described elsewhere in detail (Flege and Eefting, 1986), was synthesized using a version of the Klatt (1980) software synthesizer. Briefly, VOT was incremented in 10-ms steps from -60 to $+90$ ms to include stops in the three modal VOT ranges used to realize stops in Spanish and English. The synthesis parameters were based on careful isolated tokens of "da" and "ta" produced by a native speaker of English (JEF). The same low-intensity 10-ms release burst occurred in all 16 stimuli. Creaky voice was simulated at the end of the vowel by decreasing intensity and F_0. The relatively long (approximately 100 ms) formant transitions caused category boundaries to occur at somewhat longer values than in previous studies employing stimuli with shorter transitions (see Flege and Eefting, 1986).

C. Procedures

The adults and child subjects examined by Yeni-Komshian *et al.* (1968, 1977) produced syllables not initiated by a stop when imitating a continuum ranging from /da/ to /ta/, perhaps because they were simply told to "repeat" the stimuli and were given no practice. To avoid nonstop responses, the subjects in the present study identified the stimuli before imitating them. This ensured thorough familiarity with the stimuli. The stimuli were presented twice on each trial, with an interstimulus interval set at 0.7 s, to help the subjects focus attention on the parameters which varied across stimuli. The subjects were told to imitate the second presentation of each stimulus "as closely as possible" during the fixed 2.5-

s intertrial interval. Instructions were given in Spanish to the subjects in SA and SC, in English to the subjects in groups EA and EC, and in Spanish or English to the subjects in BC, LCB, and ECB.

Six separate randomizations of the 16-member continuum were recorded (Sony model TCD5M) for later presentation via headphones (TDH-49) at 76-dB SPL (A). Imitation responses were recorded (Technics model RS-M235) using an omnidirectional microphone (Shure model 578) positioned about 30 cm from the mouth.

D. Data analyses

Spectrograms were made (Kay model 7800) of the subjects' imitations of the final five randomizations of the 16 stimuli using a 300-Hz analyzing filter and expanded time function to improve temporal resolution. Of the 5600 stops (70 subjects × 16 stimuli × 5 repetitions), only 1.8% were unmeasurable. (One subject in group EA produced the entire vowel portion with creaky voice, which made VOT difficult to measure accurately.) Voice onset time was measured using a ruler to the nearest 0.5 mm (2.0 ms) from the beginning of the release burst to the onset of periodicity in the region of the second formant in the following vowel (in voiceless unaspirated and aspirated stops), or from the beginning of low-frequency striations to the beginning of the release burst (in prevoiced stops). Measurement reliability was estimated by remeasuring from duplicate spectrograms three stops produced by five randomly selected adult and child subjects. The mean difference between the two sets of measurements averaged 0.27 mm (1.1 ms), with a range of 0–0.5 mm (0–2.0 ms).

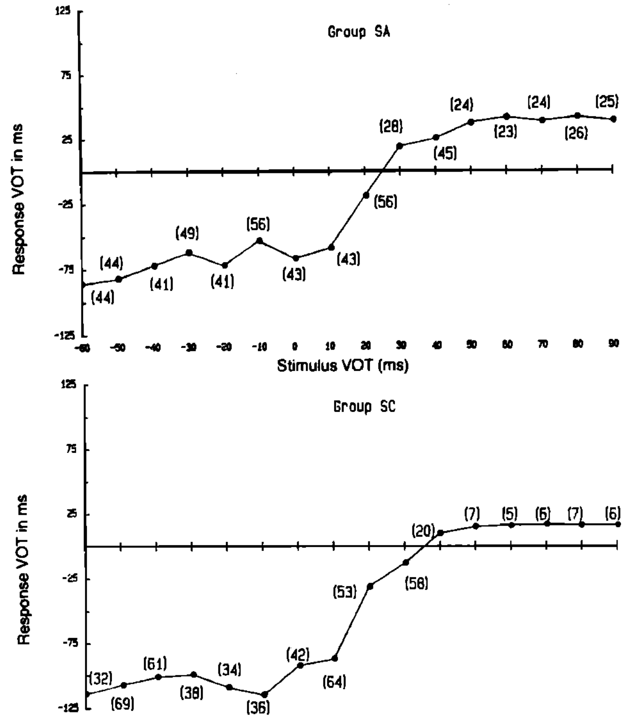

FIG. 2. The mean VOT response values produced by monolingual adult (SA) and child (SC) native speakers of Spanish who imitated a /da/ to /ta/ continuum in which VOT ranged from − 60 to 90 in ms. Each mean in this and the following figures is based on ten median values unless otherwise noted. Standard deviations are in parentheses in this and the following two figures.

II. RESULTS

A. Mean imitation responses

Unlike the native Arabic subjects examined by Yeni-Komshian *et al.* (1968, 1977), the imitation responses of the subjects in this study always began with a stop consonant. There was a great deal of intra- and inter-subject variability in the imitation responses, so the median of the five imitations of the 16 stimuli by each subject was used in calculating the mean values for the seven groups shown in Figs. 2–4.

These figures reveal that the subjects in all seven groups showed a transition region between two fairly constant response modes in imitating the VOT continuum. The monolingual Spanish subjects (groups SA and SC) differed markedly from the monolingual English subjects (groups EA and EC) in imitating the stimuli with lead and long-lag VOT. The Spanish adults (SA) generally imitated the − 60- to 10-ms stimuli with lead VOT. They imitated the 20- to 40-ms stimuli with lead, short-lag, or long-lag VOT, and the 50- to 90-ms stimuli with short-lag or long-lag VOT.

The Spanish children (SC) generally imitated the − 60- to 10-ms stimuli with lead VOT. They imitated the 20- to 40-ms stimuli with lead or short-lag VOT, and the 50- to 90-ms stimuli with short-lag VOT.

The English adults (EA) generally imitated the − 60- to 30-ms stimuli with short-lag VOT, the 40-ms stimulus with short- or long-lag VOT, and the 50- to 90-ms stimuli

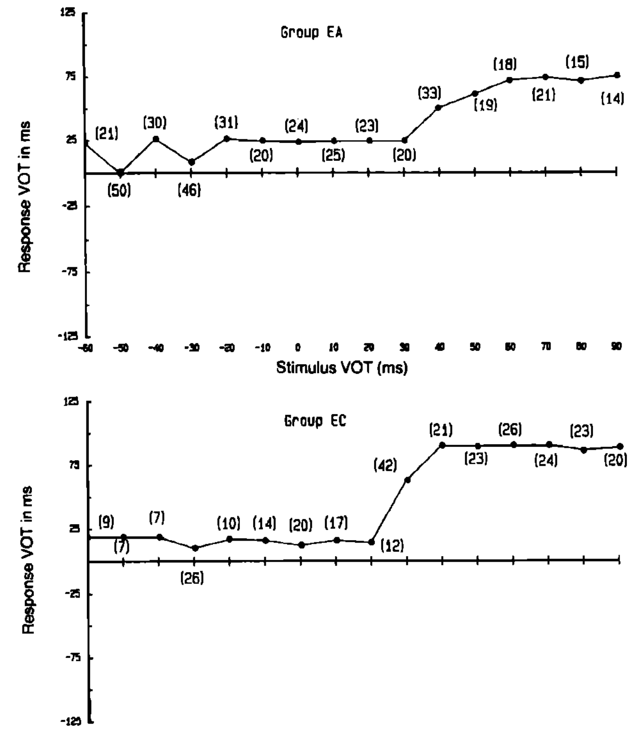

FIG. 3. The mean values produced by monolingual English adult (EA) and child (EC) speakers of English who imitated a VOT continuum, in ms. The mean for EA was based on just nine median values.

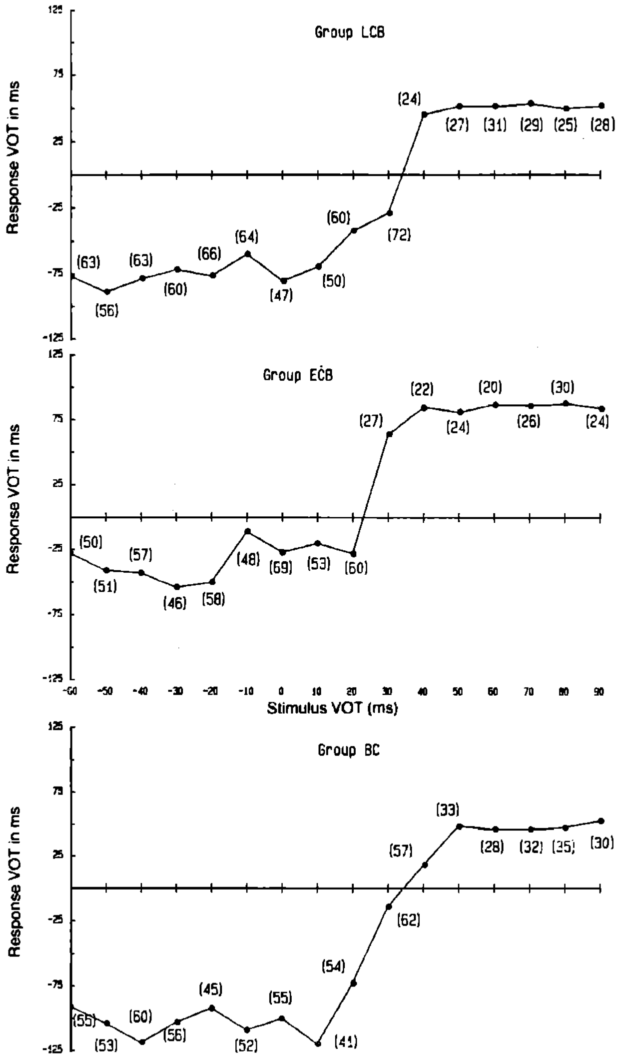

FIG. 4. The mean values produced by adult native Spanish speakers of English (ECB, LCB) and child native Spanish speakers of English (BC) imitating a VOT continuum, in ms.

with long-lag VOT. The English children (EC) generally imitated the − 60- to 20-ms stimuli with short-lag VOT, the 30-ms stimuli with short- or long-lag VOT, and the 40- to 90-ms VOT with long-lag VOT.

The native Spanish speakers of English differed from both the Spanish and English monolinguals. The adults in LCB imitated the − 60- to 20-ms stimuli with lead or short-lag VOT, the 30-ms stimulus with lead, short-lag, or long-lag VOT, and the 40- to 90-ms stimuli with short-lag or long-lag VOT. The adults in ECB also imitated the − 60- to 20-ms stimuli with lead or short-lag VOT, but imitated the 30-ms stimulus with short-lag or long-lag VOT, and the 40- to 90-ms stimuli with long-lag VOT values. The children in BC differed from the adults in LCB and ECB in imitating the − 60- to 20-ms stimuli almost exclusively with lead VOT. They imitated the 30- and 40-ms stimuli with lead, short-lag, or long-lag VOT values, and the 50- to 90-ms stimuli with short- or long-lag VOT.

The seven groups differed in how they imitated the stimuli with the greatest "lead" VOT values (i.e., the − 60- to − 30-ms stimuli). The mean VOT value produced in response to the − 60- to − 30-ms stimuli was calculated for each group. The adult English monolinguals in group EA had longer mean VOT values than the adult Spanish monolinguals in group SA (1 ms vs − 75 ms). The means obtained for the adult Spanish speakers of English in ECB and LCB (− 42 and − 79 ms, respectively) were intermediate to those obtained for the adult monolingual groups. The monolingual English children in group EC had longer mean VOT values than the monolingual Spanish children in SC (17 ms vs − 106 ms). The mean value obtained from Spanish children who spoke English (BC) (− 104 ms) was not intermediate to the Spanish and English monolingual children, but closely resembled that of the monolingual Spanish children.

The effect of the subject group on the mean VOT values produced in response to the − 60- to − 30-ms stimuli was significant [$F(6,62) = 14.4$, $p < 0.01$]. *Post-hoc* tests (Newman–Keuls, $\alpha = 0.05$) revealed that both groups of English monolinguals (EA, EC) had longer VOT values than any native Spanish group (SA, SC, LCB, ECB, and BC). Voiced stops such as /d/ may be realized with either lead or short-lag VOT values in English. This finding is, therefore, consistent with the view that short-lag stops are physiologically easier to produce than prevoiced stops (Kewley-Port and Preston, 1974). The adult Spanish speakers of English in ECB had longer VOT values than the monolingual Spanish subjects (SA, SC) because some of them resembled the native English subjects in imitating the − 60- to − 30-ms stimuli with short-lag VOT values.

The mean value produced in imitating the stimuli with VOT values ranging from 60 to 90 ms was also calculated for each subject. The monolingual English adults (EA) imitated these stimuli with longer VOT values than the monolingual Spanish adults (SA) (73 ms vs 40 ms). The mean for LCB was intermediate (51 ms) to those obtained for the English and Spanish monolinguals, but the mean for ECB (86 ms) was greater than the monolingual English adults'. The monolingual English children in group EC had longer VOT than the monolingual Spanish children in group SC (89 ms vs 17 ms). The mean for group BC was intermediate (48 ms) to that of the monolingual English and Spanish children.

Analysis of these data suggested that both age and amount of English-language experience shaped the subjects' imitation of the 60- to 90-ms stimuli. The effect of the subject group on mean VOT was significant [$F(6,62) = 13.3$, $p < 0.01$]. The monolingual Spanish children (SC) had shorter VOT values than the subjects in any other group ($p < 0.05$). The subjects in groups SA, BC, and LCB had shorter VOT than the subjects in EA, EC, and ECB ($p < 0.05$). The subjects in LCB, but not ECB, may have differed from the monolingual English adults because the subjects in ECB were exposed to English at an earlier age than those in LCB. However, it should be noted that the adults in LCB began learning English at the same age as the children in BC, who had shorter VOT values than the monolingual English children in EC.

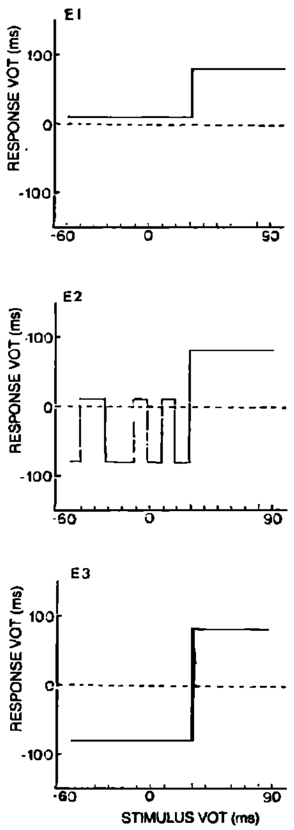

FIG. 5. Stylized response patterns representing the median VOT values produced by subjects imitating a /da/ to /ta/ continuum in which VOT ranged from − 60 to 90 ms.

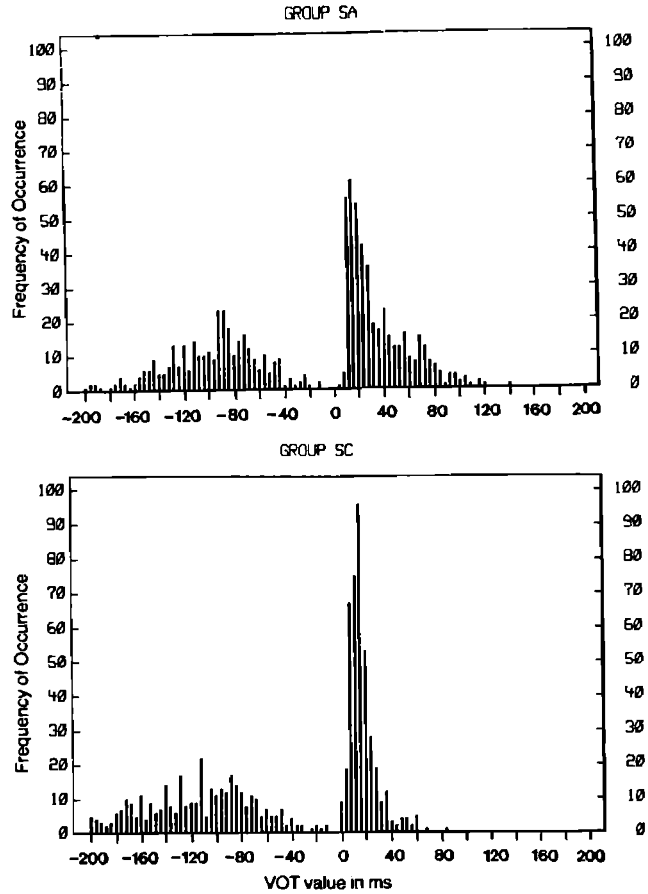

FIG. 6. The frequency of VOT values in stops produced by the native Spanish subjects in groups SA and SC as they imitated the members of a VOT continuum. The histograms are based on 784 observations for SA, and 798 observations for SC. A 4-ms bin size is used in this and the following two figures.

B. Individual response patterns

The median values for each subject were assigned to one of the five response patterns illustrated in Fig. 5. (Individual subject data conforming to these patterns are presented in Table III.) Table I indicates which of the five response patterns best described the 16 median imitation response values for each subject. (For this and subsequent analyses, long-lag stops were defined as stops with VOT values exceeding 35 ms.)

Nine of the ten monolingual English children (EC) showed the "English" pattern E1 (short-lag and long-lag responses), while all ten monolingual Spanish children (SC) showed the "Spanish" pattern, S (lead and short-lag responses). Significantly fewer bilingual children (BC) than monolingual Spanish children showed pattern S $[X(1) = 15.4, p < 0.01]$, and more children in BC than SC showed one of the other four (non-Spanish) patterns (i.e., E1, E2, E3, S/E) $[X(1) = 7.0, p < 0.01]$.

TABLE I. The imitation response patterns observed for monolingual English and Spanish children (EC, SC), bilingual Spanish/English children (BC), monolingual English and Spanish adults (EA, SA), and early and late childhood bilinguals (ECB, LCB). The "Spanish" (S), "English" (E1–E3), and "Spanish–English" (S/E) patterns are illustrated in Fig. 5 (see text).

	Subject									
Group	1	2	3	4	5	6	7	8	9	10
EC	E1	E1	E1	E1	E1	E2	E1	E1	E1	E1
SC	S	S	S	S	S	S	S	S	S	S
BC	S/E	S	S	S/E	S	E2	E3	S/E	S/E	E3
EA	E1	···	E1	E1	E1	E1	E2	E1	E2	E1
SA	E1	E3	S	S/E	S	E3	S	S	S	S
ECB	E2	E2	E2	E2	E2	E3	E3	E2	E1	E1
LCB	E3	E3	S	S	E2	E2	S	E3	S	E1

An effect of learning English on the response patterns was less clearly evident for the adult Spanish speakers of English (ECB, LCB) because of how the "monolingual" Spanish adults (SA) imitated the stimuli. Only six subjects in SA showed the "Spanish" pattern, S. Fewer of the adult native Spanish speakers of English showed pattern S (four in LCB, none in ECB). There was not a significant difference in the proportion of subjects in ECB and LCB compared to those in SA who showed one of the four non-Spanish patterns (E1, E2, E3, S/E).

C. Histogram analysis

Frequency histograms plotting the 900 (16 stimuli × 5 repetitions × 10 subjects) imitation responses for the subjects in each group were prepared to provide a fine-grained method for pinpointing concentrations of responses ("peaks") that might be indicative of stable motor control patterns. If the native Spanish subjects developed a $\{t^h\}$ category for producing /t/ in English, they might be expected to produce a cluster of responses with VOT values falling in the long-lag VOT range. The grouped histograms are presented in Fig. 6 (groups SA and SC), Fig. 7 (groups EA and EC), and Fig. 8 (groups ECB, LCB, and BC).

The Spanish monolinguals in groups SA and SC had peaks in the lead VOT range (modes at − 88 and − 110 ms, respectively) and in the short-lag VOT range (modes at 18 ms for both groups). The monolingual English subjects in groups EA and EC did not have peaks in the lead range. They realized stops with lead VOT values only sporadically. They had peaks in the short-lag range (22- and 18-ms modes) and long-lag range (70- and 82-ms modes). The native Spanish speakers of English in groups LCB, ECB, and BC represented a composite of the patterns observed for the Spanish and English monolinguals, showing peaks in the lead range (modes of − 82, − 82, − 110 ms), short-lag range (modes of 26, 22, and 14 ms), and long-lag range (modes of 70, 86, and 62 ms).

The group histograms in Figs. 6–8 suggested that the subjects in LCB, ECB, and BC showed three clear response modes in imitating the VOT continuum. However, since the peaks might have been due to the imitations of a small number of subjects in each group, histograms plotting the 90 imitation responses of each subject were also prepared and visually inspected by one of us (JEF). A peak was considered to have occurred if four or more tokens occurred in a maximum of two adjacent 4-ms bins in the lead, short-lag, and long-lag VOT values.[5] If a peak centered at a value of more than 90 ms (the longest VOT in the stimuli) occurred, it was designated an "extra-long-lag" peak. The same was true for the second of two peaks with values of 35 ms or more.[6]

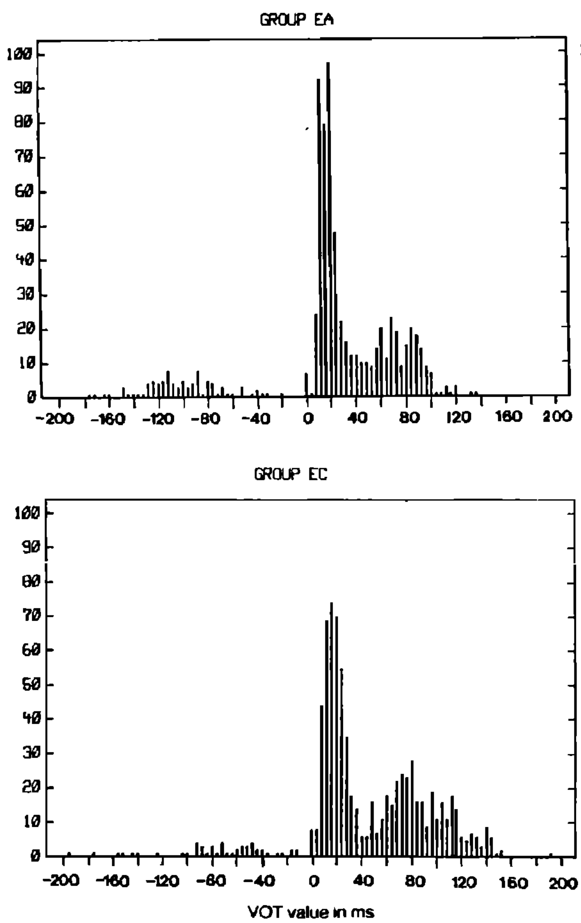

FIG. 7. The frequency of VOT values in stops produced by the native English subjects in groups EA and EC as they imitated the members of a VOT continuum. The histograms are based on 720 observations for EA, and 799 observations for EC.

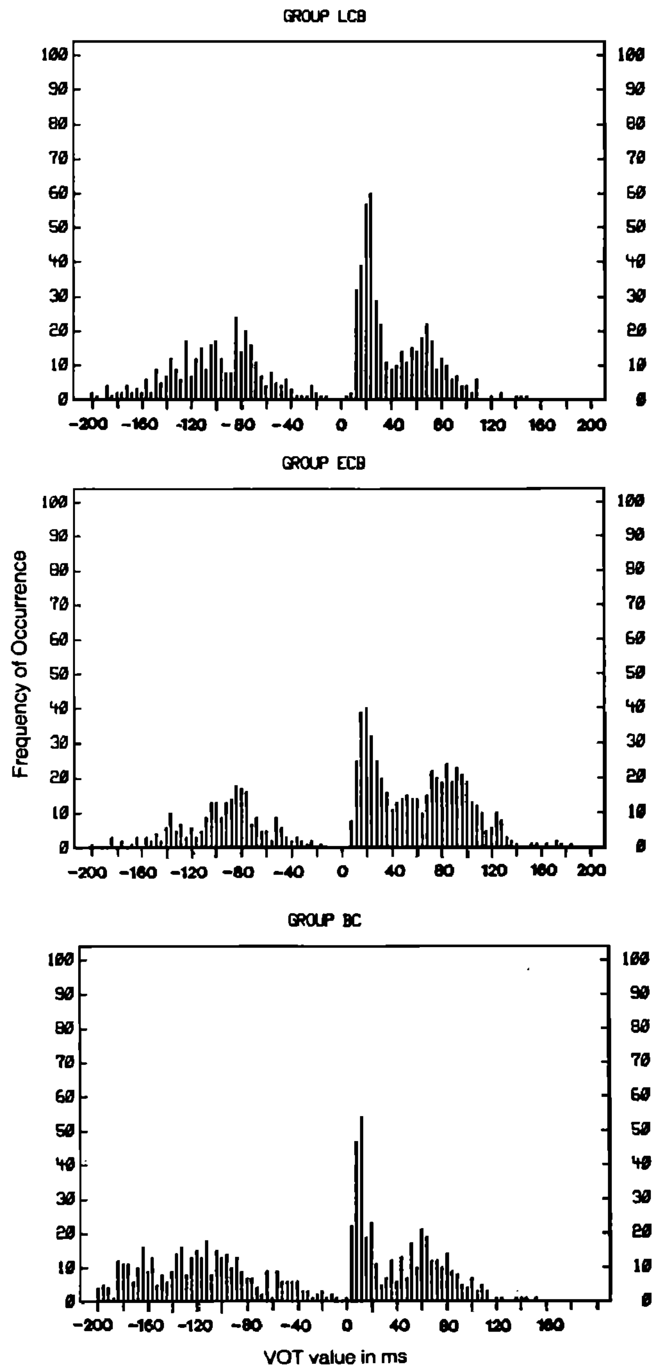

FIG. 8. The frequency of VOT values in stops produced by the native Spanish speakers of English in groups LCB, ECB, and BC as they imitated the members of a VOT continuum. The histograms are based on 800 observations for all three groups.

Table II lists the peaks evident for individual subjects in each group. Consistent with the grouped data, most of the monolingual Spanish children (SC) had lead and short-lag peaks, most of the monolingual English children (EC) had short-lag and long-lag peaks, and most of the bilingual children (BC) showed a composite pattern. All ten children in BC had lead peaks, seven had short-lag peaks, and eight had long-lag peaks. Significantly more children in BC than SC had a long-lag peak $[X^2(1) = 8.0, p < 0.01]$.

More of the monolingual Spanish adults (SA) than monolingual English adults (EA) had lead peaks (9/10 vs 3/9). Fewer subjects in SA than EA had long-lag peaks (4/10 vs 9/9). Most (17/20) of the adult native Spanish speakers of English (LCB, ECB), like the monolingual Spanish adults, had lead peaks. Like the monolingual English adults, most (16/20) of them also had long-lag peaks. The proportion of subjects in LCB/ECB and SA who had long-lag peaks did not differ significantly because four subjects in SA had long-lag peaks. However, the proportion of all 30 native Spanish speakers of English (groups BC, LCB, ECB) who had long-lag peaks was significantly greater than the proportion of monolingual Spanish subjects (SC and SA) with long-lag peaks $[X(1) = 7.71, p < 0.01]$.

The number of subjects in each group who showed peaks in all three modal VOT ranges (i.e., lead, short-lag, and long-lag) was tabulated. There was just one such subject in EC, none in SC, and six in BC. There were three in EA, one in SA, four in LCB, and six in ECB. Significantly more Spanish speakers of English (ECB, LCB, BC) than Spanish monolinguals (SA, SC) had three peaks $[X^2(1) = 8.25, p < 0.01]$.

TABLE II. The location of peaks in individual subjects' histograms. "Lead" indicates the lead VOT range, "s-lag" the short-lag range, "l-lag" the long-lag range, and "xl-lag," a range with even longer VOT than long-lag (see text).

	English adults (EA)				English children (EC)			
S#	Lead	s-lag	l-lag	xl-lag	Lead	s-lag	l-lag	xl-lag
1	*	22	62	*	*	18	64	*
2	...	...	...	...	*	14	78	*
3	*	14	58	*	*	30	50	106
4	*	18	76	*	*	26	*	118
5	*	22	62	94	*	22	*	92
6	−118	22	70	*	−50	14	84	*
7	−82	22	*	94	*	22	74	102
8	*	18	74	*	*	14	82	*
9	−122	14	86	*	*	22	*	114
10	*	14	46	*	*	10	62	*
	Spanish adults (SA)				Spanish children (SC)			
S#	Lead	s-lag	l-lag	xl-lag	Lead	s-lag	l-lag	xl-lag
1	*	22	42	*	−138	18	*	*
2	−102	38	64	*	−82	18	*	*
3	−90	22	*	*	−126	18	*	*
4	−118	*	54	74	−140	18	*	*
5	−90	34	*	*	−168	12	*	*
6	−112	*	50	*	−110	10	*	*
7	−106	18	*	*	−86	16	*	*
8	−98	30	*	*	−74	16	*	*
9	−66	14	*	*	−90	14	*	*
10	−70	18	*	*	−166	22	*	*
	Group LCB				Group ECB			
S#	Lead	s-lag	l-lag	xl-lag	Lead	s-lag	l-lag	xl-lag
1	−154	26	58	*	−102	16	74	94
2	−146	34	78	*	−146	38	*	126
3	−74	22	*	*	−82	18	74	*
4	−122	26	*	*	−74	22	70	*
5	−82	26	*	94	−66	18	78	*
6	−66	22	70	*	−78	*	86	104
7	−82	22	*	*	−126	*	54	102
8	−94	*	70	*	*	18	62	*
9	−114	18	*	*	*	26	82	100
10	*	26	60	*	−90	18	90	*

Group BC			
Lead	s-lag	l-lag	xl-lag
−162	*	66	*
−126	6	46	*
−98	10	*	*
−86	22	64	*
−158	14	*	*
−62	14	62	*
−70	*	62	102
−162	22	66	*
−94	*	62	*
−110	12	82	102

D. Effect of categorization

The location of discontinuities in the subjects' imitation responses seemed to match closely the phoneme boundaries obtained from an identification experiment in most instances. To quantify this relationship, a procedure was developed to determine the location of the discontinuities in each subject's data. The procedure is illustrated in Table III for five subjects. Asterisks mark the location of the discontinuities. In pattern S (observed for most monolingual Spanish subjects), about half of the stimuli were imitated with lead VOT and the remaining half with short-lag VOT. The location of the discontinuity was considered to be 15 ms since the shift occurred between the 10- and 20-ms stimuli. In patterns E1–E3 (seen for native speakers of English as

TABLE III. Illustration of how the location of discontinuities in five subjects' imitation responses were specified in ms. "S" designates a typical "Spanish" pattern, "E1–E3" three possible "English" patterns, and "S/E" a composite English/Spanish pattern. The discontinuity was considered to have fallen between the two responses marked with asterisks.

Stimulus #	VOT	S	E1	E2	E3	S/E
1	−60	−79	20	24	−162	−166
2	−50	−69	16	−81	−186	−113
3	−40	−53	24	16	−166	−194
4	−30	−89	18	−73	−174	−162
5	−20	−81	18	−77	−101	−150
6	−10	−71	16	22	−141	−170
7	0	−71	18	16	−145	−154
8	10	−83*	16	−77	−117	−162
9	20	16*	18	−85*	−109	12
10	30	18	28*	49*	−133*	20*
11	40	16	63*	87	44*	65*
12	50	20	61	49	57	57
13	60	16	59	77	49	49
14	70	18	77	81	44	49
15	80	20	55	77	44	38
16	90	16	79	69	49	65

TABLE IV. The mean location of response discontinuities for subjects in seven groups, along with the mean phoneme boundaries obtained from the same subjects in a previous study (Flege and Eefting, 1986), in ms.

Response discontinuities Subject group						
EA	EC	SA	SC	ECB	LCB	BC
41.7	28.0	21.0	17.0	26.0	28.0	32.0
(8.7)	(4.8)	(8.4)	(10.3)	(3.2)	(8.2)	(9.5)
Phoneme boundaries Subject group						
EA	EC	SA	SC	ECB	LCB	BC
42.4	29.9	23.2	16.7	27.4	28.9	32.6
(5.0)	(4.3)	(10.8)	(7.2)	(3.0)	(6.2)	(7.0)

well as some of the native Spanish subjects), the shift is from short-lag (or from lead and short-lag) to long-lag responses. In pattern S/E, a composite of Spanish and English patterns, there is a shift from lead to short-lag responses and a shift from short-lag to long-lag responses. The location of the discontinuity was considered to have occurred between the short-lag and long-lag responses in these instances.

The discontinuities occurred at about the same location as the phoneme boundaries. The mean location of the response discontinuities obtained for the seven subject groups is presented in Table IV along with the mean phoneme boundaries obtained by Flege and Eefting (1986). The mean (signed) difference for the grouped data was only 1.2 ms. The mean (unsigned) difference between the discontinuities and phoneme boundaries was 6.1 ms (s.d. = 5.4) for the 69 subjects for whom two sets of data were available. The maximum divergence for any subject was 15 ms.

The two sets of data were submitted to a mixed-design ANOVA, which yielded a significant effect of group [$F(6,62) = 9.76, p < 0.01$], but not a significant condition or group by condition interaction. *Post-hoc* tests (Newman–Keuls, $\alpha = 0.05$) showed that discontinuities occurred at a greater VOT value for the monolingual English adults (EA) than for any other subject group. The discontinuities for the monolingual Spanish children (SC) occurred at shorter values than for subjects in any other group except the monolingual Spanish adults (SA). These same differences were obtained in an analysis of the phoneme boundary data (Flege and Eefting, 1986). There was only one between-group difference not seen in the phoneme boundary data: The mean location of the response discontinuities was shorter for SA than BC.

III. DISCUSSION

A. Category formation

An imitation task was used here to help determine whether native speakers of Spanish who had been observed to produce English /p,t,k/ with significantly shorter voice onset time (VOT) values than native English speakers (Flege and Eefting, 1987a) did so because they lacked phonetic categories ($\{p^h\}$, $\{t^h\}$, $\{k^h\}$) for implementing /p,t,k/ in English. We hypothesized that a native speaker of Spanish would need to establish a phonetic category $\{t^h\}$ and a realization rule for physically outputting $\{t^h\}$ in order to produce English /t/ with appropriate VOT values (see the Introduction).

The most important finding was that the Spanish speakers of English produced stops with VOT values falling into the three "modal" VOT ranges (i.e., lead, short-lag, long-lag) when imitating stimuli with VOT values ranging from − 60 to 90 ms. Histograms plotting the frequency of occurrence of VOT values in the responses of individual subjects and entire groups were prepared. The group histograms revealed concentrations of VOT values ("peaks") in three locations for the Spanish speakers of English (ECB, LCB, EC). Peaks were evident in just the lead and short-lag ranges for the Spanish monolinguals (SA, SC), and in just the short-lag and long-lag ranges for the English monolinguals (EA, EC). Histogram for individual subjects revealed much the same pattern. A significantly larger proportion of the native Spanish speakers of English (ECB, LCB, BC) showed long-lag peaks (80%) than monolingual Spanish subjects (SA, SC) (20%). Significantly more subjects in groups ECB, LCB, and EC (53%) than in groups SC and SA (5%) showed peaks in all three modal VOT ranges.

Evidence that the bilingual subjects had a long-lag "mode" of production indicates that they were able to distinguish short-lag and long-lag stops perceptually, even though both kinds of stops would be identified phonologically as /t/. It also demonstrated that they were able to physically realize stops with short- and long-lag values rapidly and reliably in the short time available between imitation trials. In terms of the model presented in the Introduction, this suggested the Spanish speakers of English had formed a phonetic category $\{t^h\}$ for implementing English /t/, and had established a realization rule for outputting $\{t^h\}$.

Evidence for the establishment of a $\{t^h\}$ category was more straightforward for children than adults owing to differences between children and adults in the Spanish "monolingual" groups. None of the monolingual Spanish children (group SC) showed long-lag peaks, but four monolingual adults (group SA) had long-lag peaks. Although the adults in SA had studied English for 12 years in school, we have referred to them here as "monolingual" because they were unable to carry on a simple conversation in English and had never lived in an English-speaking environment.[7]

Three subjects in SA indicated on a language background questionnaire that they could speak English "a little." Two of these subjects were among the four producing long-lag stops. Limited exposure to English /t/ may have enabled the four subjects in SA to imitate stops with long-lag VOT. The children in group SC, who did not produce long-lag stops, were truly monolingual. Results obtained by Yeni-Komshian *et al.* (1968, 1977) for native Arabic children and adults with little or no previous exposure to long-lag stops suggested that adults may be somewhat better able than children to imitate long-lag stops. Further research will be needed to determine whether the apparent adult–child difference seen here was due to an age-related difference in ability to reproduce novel phones, previous phonetic experience, or both.

In one instance, the imitation data revealed a between-group difference not evident in spontaneous production. In spontaneous production, the adults in ECB and LCB realized /p,t,k/ with shorter VOT values than native speakers of English, but did not differ from one another (Flege and Eefting, 1987a). In the present study, the subjects in ECB imitated the 60- to 90-ms VOT stimuli with significantly longer VOT values than the subjects in LCB, but did not differ significantly from the native English adults.

The subjects in ECB began learning English as young children in the U.S., while those in LCB began learning English when they enrolled in a bilingual school in Puerto Rico at the age of 5–6 years. The between-group difference in imitation might be attributed to earlier, greater, or more authentic English-language input for the subjects in ECB than LCB. Whatever the cause, the imitation task revealed tacit knowledge concerning English /t/ among the subjects in ECB that was not evident in spontaneous production.

The conclusion that the native Spanish speakers of English developed a phonetic category $\{t^h\}$ for implementing English /t/ is somewhat surprising in view of the finding (Flege and Eefting, 1987a) that they produced English /p,t,k/ with significantly shorter VOT values than native English speakers. The difference in VOT between native and non-native speakers may have arisen because the $\{t^h\}$ categories established by the Spanish speakers of English differed from native English speakers'. The Spanish speakers of English undoubtedly often heard English spoken by other native speakers of Spanish. In Spanish-accented English, /t/ is often realized with VOT values intermediate to the values typical for English and Spanish. It is also possible that the native Spanish speakers of English had not yet finished developing a realization rule which would result in acoustic output that adequately reflected the (correct) information encoded in their $\{t^h\}$ categories.

Our conclusion that the native Spanish subjects developed $\{t^h\}$ categories for implementing English /t/ is consistent with data pertaining to the effect of L2 learning on the production of stops in the native language (L1). Flege and Eefting (1987a) reported that the subjects in groups ECB, LCB, and BC realized /p,t,k/ with significantly shorter VOT values in Spanish words than age-matched Spanish monolinguals (SA, SC). The L2 effect on L1 production was hypothesized to have arisen from a universal phonetic principle termed "polarization." Keating (1984) hypothesized that the somewhat longer VOT values in short-lag realizations of Spanish /p,t,k/ than in short-lag realizations of English /b,d,g/ arise because Spanish monolinguals attempt to enhance the contrast between lead and short-lag stops (i.e., realizations of /b,d,g/ vs /p,t,k/), while English monolinguals attempt to enhance the contrast between short-lag and long-lag stops.

If the native Spanish subjects formed a phonetic category $\{t^h\}$, they might attempt to enhance the contrast between their short-lag and long-lag realizations of /t/ (in Spanish and English, respectively). This assumes that the phonetic categories used in producing two languages are part of a single system. It also assumes that enhancing the VOT contrast between short-lag and long-lag stops is more important phonetically than enhancing the VOT contrast between lead and short-lag stops.

Our conclusion that the native Spanish subjects established a $\{t^h\}$ category for stops in English differs from the conclusion drawn from a study (Flege, 1987a) which examined the speech of native French subjects who learned English as adults. These subjects realized the short-lag /t/ of French with significantly longer—rather than shorter—VOT values than French monolinguals.[8] It may be that individuals who learn an L2 in adulthood do not establish phonetic categories for phones in L2 that can be readily identified in terms of a category present in L1.

The hypothesis that "early L2 learners" but not "late L2 learners" form phonetic categories for certain phones in the L2 is consistent with the results of several studies of L2 production. Mack (1983) observed little difference between monolinguals and adults who learned both French and English as young children in the VOT measured in realizations of English /t/. Williams (1977b) found no VOT difference for /p/ between English monolinguals and adult subjects who learned Spanish and English as young children. A recent study (Fokes *et al.*, 1985) suggested that Arabic children may not differ from native English children in producing English stops. This stands in contrast to two earlier studies which showed that adult native speakers of Arabic produced English /p,t,k/ with significantly shorter (and therefore Arabic-like) VOT values than adult native English speakers (Flege and Port, 1981; Port and Mitleb, 1983).

Other studies, on the other hand, provide counterevidence to the "early L2 learning" hypothesis. Caramazza *et al.* (1973) found that French Canadians who began learning English "no later" than 7 years of age realized /p,t,k/ in English with VOT values that were significantly shorter (and, therefore, French-like) than native English subjects. Williams' (1979) data indicate that two groups of native Spanish (Puerto Rican) children realized English /p/ with Spanish-like mean VOT values (ca. 20–40 ms) that were shorter than values typically observed for native English children.

The studies just cited differ greatly in terms of how data were elicited. It is also likely that the subjects differed in terms of where and from whom they learned English. A great deal of further research will be needed to help resolve the issue of why, and under what conditions, non-native speakers will differ from English native speakers in realizing English /p,t,k/. To reach valid conclusions based on between-group comparisons, it will be necessary to carefully control not only the age at which L2 learning began, but also the kind and amount of L2 input.

B. Covert categorization

Unlike subjects examined in previous studies (Yeni-Komshian *et al.*, 1968, 1977), the subjects in the present study did not produce isolated vowels or syllables initiated by consonants other than /t/ or /d/ when imitating the members of a voice onset time (VOT) continuum which ranged from /da/ to /ta/. This was probably because they were thoroughly familiar with the stimuli, because the synthetic stimuli were more natural sounding, or both. Like subjects in previous VOT imitation studies, the adults and children in this study failed to reproduce accurately the incremental increases in VOT in the stimuli (Yeni-Komshian *et al.*, 1968, 1977).

Vocal imitation in humans has been described as a special capacity for discovering links between "perceived movements and their corresponding motor controls" mediated by central representations which are "closely related to the dynamics of articulation" (Studdert-Kennedy, 1986a, p. 53ff.; cf. Liberman and Mattingly, 1986). This view, which suggests that talkers will reproduce the members of a VOT continuum with the nearest equivalent in their phonetic repertoire, is consistent with results obtained in the VOT imitation studies of Yeni-Komshian *et al.* (1968, 1977). Chistovich *et al.* (1966) hypothesized, on the other hand, that imitation partially bypasses categorization, thereby resembling to some extent the reproduction of human speech by mynah birds (Marler and Mundinger, 1971).

This study supported the view that categorization precedes imitation. Each subject showed rather marked shifts in their imitation responses at some point along the VOT continuum. The location of the shifts did not differ significantly from the phoneme boundaries obtained in an earlier identification experiment (Flege and Eefting, 1987a). The location of the response shifts occurred at significantly longer values for the monolingual English than Spanish subjects. This agrees with the finding by Yeni-Komshian *et al.* (1968) that native English children tended to show response shifts at longer values along a VOT continuum than native Arabic children.

Covert categorization may have prevented subjects from making use of a more direct link between auditory input and motoric representations. The strong categorical effect on imitation seen here raises the question of whether the subjects might have been able to reproduce the VOT values in the stimuli more accurately in some other task or condition. Flege and Hammond (1982) found that monolingual English adults were able to decrease the VOT in /t/ when mimicking Spanish-accented English. Weismer and Cariski (1985) found that talkers could modify VOT when instructed to do so. Perhaps identifying the member of the /da/ to /ta/ continuum just prior to imitating them (which was intended to help avoid nonstop imitation responses) increased the subjects' tendency to categorize the stimuli.

In summary, Spanish speakers of English-produced stops with VOT values falling into three modal VOT categories when imitating the members of a VOT continuum ranging from /da/ to /ta/, while the imitation responses of Spanish and English monolinguals fell into only two modal VOT categories. This was interpreted to mean that the native Spanish subjects had formed a phonetic category, $\{t^h\}$, with which to implement the voiceless aspirated /t/ of English.

ACKNOWLEDGMENTS

This research was supported by NIH Grant No. NS20963. The authors thank E. Blau, C. Casablanca, E. Flores, and J. Serota for help locating subjects, and E. James for help preparing the manuscript.

[1]This is true, obviously, only when Spanish /d/ is realized as a stop consonant rather than a homorganic fricative.

[2]At least some of the variation between lead and short-lag acoustic realizations of English /b,d,g/ may be due to differences in vocal fold tensioning and/or medial compression rather than to laryngeal timing differences. Flege (1982) used electroglottography to infer the timing of vocal fold adduction in utterance-initial tokens of /b/. Nine of the ten subjects examined adducted the vocal folds before or during stop closure. These subjects produced either lead or short-lag stops. The remaining subject did not adduct the vocal folds until stop release, and produced only short-lag stops.

[3]Categorization may influence the imitation of fricative duration (Karno and Porter, 1980) and vowel formant frequency (Kent, 1973, 1979; Repp and Williams, 1985, 1987) to a lesser extent than it affects the imitation of stops. The influence of categorization on the imitation of vowel duration is unclear at present (see Bastian and Abramson, 1962; Neaser, 1970; Lehiste and Shockey, 1980).

[4]Two other native Arabic children showed a single response mode in imitating all of the stimuli, suggesting inability to perform the phonetic task.

[5]Preliminary analyses indicated that the optimal bin size was 4 ms, twice the temporal resolution of the technique used for measuring VOT.

[6]The presence of an "extra-long-lag" peak was treated the same as a "long-lag" peak in the ensuing analyses and discussions.

[7]No information is available concerning what specific instruction, if any, the subjects in SA received in school concerning the phonetic difference between /p,t,k/ in English and Spanish.

[8]A corresponding, but reverse, pattern was noted for native English speakers of French.

Abramson, A., and Lisker, L. (**1973**). "Voice-timing perception in Spanish word-initial stops," J. Phon. **1**, 1–8.

Barton, D. (**1978**). *The Role of Perception in the Acquisition of Phonology* (Indiana University Linguistics Club, Bloomington, IN).

Bastian, J., and Abramson, A. (**1962**). "Identification and discrimination of phonemic vowel duration," paper presented at the 63rd Meeting of the Acoustical Society of America, New York, 24–26 May 1962.

Caramazza, A., Yeni-Komshian, G., Zurif, E., and Carbone, E. (**1973**). "The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals," J. Acoust. Soc. Am. **54**, 421–428.

Chistovich, L., Fant, G., deSerpa-Leitão, A., and Tjernlund, P. (**1966**). "Mimicking and perception of synthetic vowels," Quart. Prog. Stat. Rep., Royal Technical University, Speech Transmission Laboratory, Stockholm **2**, 1–18.

Christensen, J. (**1984**). "The perception of voice onset time: A cross-language study of American English and Danish," Ann. Rep. Inst. Phonetics, University of Copenhagen **18**, 163–185.

Flege, J. (**1981**). "The phonological basis of foreign accent: A hypothesis," TESOL Q. **15**, 443–455.

Flege, J. (**1982**). "Laryngeal timing and phonation onset in utterance-initial English stops," J. Phon. **10**, 177–192.

Flege, J. (**1987a**). "The production of 'new' and 'similar' phones in a foreign language: evidence for the effect of equivalence classification," J. Phon. **15**, 47–65.

Flege, J. (**1987b**). "A critical period for learning to pronounce foreign languages?" Appl. Linguist. **8**, 162–177.

Flege, J. (**1987c**). "The production and perception of foreign language speech sounds," in *Human Communication and Its Disorders, Vol. III*, edited by H. Winitz (Ablex, Norwood, NJ) (in press).

Flege, J. (**1987d**). "The relevance of phonetic input to phonological development in L2 learning," paper presented at the Conference of Second-Language Acquisition, Stanford University, 17–19 July 1987.

Flege, J., and Eefting, W. (**1986**). "Linguistic and developmental effects on the production and perception of stop consonants," Phonetica **43**, 155–171.

Flege, J., and Eefting, W. (**1987a**). "Production and perception of English stops by native Spanish speakers," J. Phon. **15**, 67–83.

Flege, J., and Eefting, W. (**1987b**). "Cross-language switching in stop consonant production and perception by Dutch speakers of English," Speech Commun. **6**, 185–202.

Flege, J., and Hammond, R. (**1982**). "Mimicry of non-distinctive phonetic differences between language varieties," Stud. Second Lang. Acquis. **5**, 1–17.

Flege, J., and Massey, K. (**1980**). "English prevoicing: Random or controlled?" paper presented at the Linguistic Society of America, Albuquerque, NM, August 1980.
Flege, J., and Port, R. (**1981**). "Cross-language phonetic interference: Arabic to English," Lang. Speech **24**, 125–146.
Flege, J., McCutcheon, M., and Smith, S. (**1987**). "The development of skill in producing word-final English stops," J. Acoust. Soc. Am. **82**, 433–447.
Fokes, J., Bond, Z., and Steinberg, M. (**1985**). "Acquisition of the English voicing contrast by Arab children," Lang. Speech **28**, 81–91.
Garnica, O., and Edwards, M. (**1977**). "Phonological variation in child speech: The trade-off phenomenon," Working Pap. in Linguist. **22**, 81–87 (Dept. of Linguistics, Ohio State University, Columbus, OH).
Kalikow, D., and Swets, J. (**1972**). "Experiments with computer-controlled displays in second-language learning," IEEE Trans. Audio Electroacoust. **20**, 23–28.
Karno, E., and Porter, R. (**1980**). "Imitation of /s/ duration in VCV's," J. Acoust. Soc. Am. Suppl. 1 **67**, S38.
Keating, P. (**1984**). "Phonetic and phonological representation of stop consonant voicing," Language **60**, 286–319.
Kent, R. (**1973**). "The imitation of synthetic vowels and some implications for speech memory," Phonetica **28**, 45–52.
Kent, R. (**1979**). "Imitation of synthesized English and non-English vowels by children and adults," J. Psycholinguist Res. **8**, 43–60.
Kewley-Port, D., and Preston, M. (**1974**). "Early apical stop production," J. Phon. **2**, 195–210.
Klatt, D. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.
Klein, H. (**1979**). "Imitative or spontaneous word productions: What's the difference?" paper presented at the American-Speech-Language-Hearing Association annual meeting, Atlanta, 16 November 1986.
Lehiste, I., and Shockey, L. (**1980**). "Labeling, discrimination and repetition of stimuli with level and changing fundamental frequency," J. Phon. **8**, 469–474.
Leonard, L., Schwartz, R., Folger, K., and Wilcox, M. (**1978**). "Some aspects of child phonology in imitative and spontaneous speech," J. Child Lang. **5**, 403–416.
Liberman, A., and Mattingly, I. G. (**1986**). "The motor theory of speech perception revised," Haskins Lab. Stat. Rep. Speech Res. **82/83**, 63–93.
Lisker, L., and Abramson, A. (**1964**). "Cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.
Lisker, L., and Abramson, A. (**1967**). "Some effects of context of voice onset time," Lang. Speech **10**, 1–28.
Mack, M. (**1983**). "Psycholinguistic consequences of early bilingualism: A comparative study of performance of English monolinguals and French-English bilinguals in phonetic, syntactic, and semantic experiments," Doctoral dissertation, Brown University, Providence, RI.
Major, R. (**1987**). "English voiceless stop production by speakers of Brazilian Portuguese," J. Phon. **15**, 197–202.
Marler, P., and Mundinger, P. (**1971**). "Vocal learning in birds," in *Ontogeny of Vertebrate Behavior*, edited by H. Moltz (Academic, New York).
McLaughlin, B. (**1978**). *Second Language Acquisition in Childhood* (Erlbaum, Hillsdale, NJ).
Naeser, M. (**1970**). *The American Child's Acquisition of Differential Vowel Duration*, the Wisconsin Research and Development Center for Cognitive Learning, Technical Report 144. University of Wisconsin, Madison, WI.
Port, R., and Mitleb, F. (**1983**). "Segmental features and implementation in acquisition of English by Arabic speakers," J. Phon. **11**, 219–231.
Repp, B., and Williams, D. (**1985**). "Categorical trends in vowel imitation: Preliminary observations from a replication experiment," Speech Commun. **4**, 105–120.
Repp, B., and Williams, D. (**1987**). "Categorical tendencies in imitating self-produced isolated vowels," Speech Commun. **6**, 1–14.
Stevens, K. (**1983**). "Design features of speech sound systems," in *The Production of Speech*, edited by P. MacNeilage (Springer, New York), pp. 247–263.
Studdert-Kennedy, M. (**1986a**). "On learning to speak," Haskins Lab. Stat. Rep. Speech Res. **82/83**, 53–62.
Studdert-Kennedy, M. (**1986b**). "Development of the speech perceptuomotor system," Haskins Lab. Stat. Rep. Speech Res. **82/83**, 131–140.
Trost, J., and Canter, G. (**1974**). "Apraxia of speech in patients with Broca's aphasia: A study of phoneme production accuracy and error patterns," Brain Lang. **1**, 63–79.
Weismer, G., and Cariski, D. (**1985**). "On speakers' abilities to control speech mechanism output: Theoretical and clinical implications," in *Speech and Language: Advances in a Basic Research and Practice*, edited by N. Lass (Academic, New York), Vol. 10.
Williams, L. (**1977a**). "The voicing contrast in Spanish," J. Phon. **5**, 169–184.
Williams, L. (**1977b**). "The perception of stop consonant voicing by Spanish-English bilinguals," Percept. Psychophys. **21**, 289–297.
Williams, L. (**1979**). "The modification of speech perception and production in second-language learning," Percept. Psychophys. **26**, 95–104.
Yeni-Komshian, G., Caramazza, A., and Preston, M. (**1977**). "A study of voicing in Lebanese Arabic," J. Phon. **5**, 35–48.
Yeni-Komshian, G., Preston, M., and Port, D. (**1968**). "A cross-language study of the perception of stop consonants in American and Lebanese children," Neurocommunication Lab. Prog. Rep., Johns Hopkins University, Vol. 3, pp. 1–15.