

# Early learners' discrimination of second-language vowels

Anders Højen<sup>a)</sup> and James E. Flege<sup>b)</sup>

Speech and Hearing Sciences Division, University of Alabama, Birmingham, Alabama 35294-2041

(Received 21 March 2005; revised 7 February 2006; accepted 14 February 2006)

It is uncertain from previous research to what extent the perceptual system retains plasticity after attunement to the native language (L1) sound system. This study evaluated second-language (L2) vowel discrimination by individuals who began learning the L2 as children ("early learners"). Experiment 1 identified procedures that lowered discrimination scores for foreign vowel contrasts in an AXB test (with three physically different stimuli per trial, where "X" was drawn from the same vowel category as "A" or "B"). Experiment 2 examined the AXB discrimination of English vowels by native Spanish early learners and monolingual speakers of Spanish and English (20 per group) at interstimulus intervals (ISIs) of 1000 and 0 ms. The Spanish monolinguals obtained near-chance scores for three difficult vowel contrasts, presumably because they did not perceive the vowels as distinct phonemes and because the experimental design hindered low-level encoding strategies. Like the English monolinguals, the early learners obtained high scores, indicating they had shown considerable perceptual learning. However, statistically significant differences between early learners and English monolinguals for two of three difficult contrasts at the 0-ms ISI suggested that their underlying perceptual systems were not identical. Implications for claims regarding perceptual plasticity following L1 attunement are discussed. © 2006 Acoustical Society of America. [DOI: 10.1121/1.2184289]

PACS number(s): 43.71.Ft, 43.71.Es, 43.71.Hw [ARB]

Pages: 3072–3084

## I. INTRODUCTION

It is well established that the age at which a second language (L2) is learned will exert a long-term effect on segmental phonetic perception. "Early learners" first exposed to their L2 as children resemble native speakers to a greater extent than do "late learners" whose first exposure to the L2 occurred in late adolescence or early adulthood (Flege *et al.*, 1999; Flege and MacKay, 2004; Yamada, 1995). It is uncertain, however, whether early learners will in time come to perceive L2 vowels and consonants just like native speakers of the target L2. Some studies have provided evidence of substantial differences between early learners and L2 native speakers whereas others have shown evidence of little or no difference. These divergent findings raise questions regarding the extent to which the speech perception system remains plastic following attunement to the native language (L1) sound system. The aim of this study, therefore, was to further evaluate early learners' perceptual abilities. It did so by examining the discrimination of English vowels by native Spanish speakers who had learned English as an L2 in childhood.

### A. Previous research

Infants begin to perceive speech in a language-specific way during the first year of life (e.g., Kuhl *et al.*, 1992; Werker and Tees, 1984). As they develop perceptual pro-

cesses specialized for perceiving phonemic contrasts used to contrast meaning (words) in the ambient language, infants' sensitivity to contrasts not serving such a function declines. The influence of this attunement to the L1 sound system is evident in the difficulties that adult L2 learners may have in distinguishing certain nonnative phonemic contrasts (e.g., in native Japanese speakers' difficulty in distinguishing English /ɪ/ and /I/).

It has been proposed that as people grow older, L2 learning is hindered by a loss of neuroplasticity that accompanies normal neural maturation (e.g., Long, 1990; Scovel, 2000). More recently, a loss of plasticity for speech perception has been discussed in the context of the effects of L1 perceptual attunement on subsequent perceptual learning rather than, or in addition to, a maturationally defined decrease of neuroplasticity that affects basic aspects of sensory coding, storage, and retrieval associated with speech learning. More specifically, it has been argued that L2 speech acquisition is impeded by the acquisition of the L1 phonological system (e.g., Flege, 1995; Iverson *et al.*, 2003; Kuhl, 2000; McClelland, 2001; Pallier *et al.*, 1997, 2003; Sebastián-Gallés and Soto-Faraco, 1999).

One influential account of the effect of L1 perceptual attunement on L2 perception is the native language magnet (NLM) theory (e.g., Kuhl *et al.*, 1992; Kuhl and Iverson, 1995). According to the NLM, during the first year of life L1 phonetic prototypes begin to act as perceptual "magnets" that "warp" auditory input. This would be expected, for example, to reduce the discriminability of L1 vowel tokens near the category center (the prototype) whereas equally distant tokens from the category boundary region, or from two different categories, would be expected to remain highly discriminable.

<sup>a)</sup>Present affiliations: University of Southern Denmark and Højen Research. Address for correspondence: Praestegaardsvej 2 DK-8900 Randers, Denmark; electronic mail: anders@andershojen.dk

<sup>b)</sup>Present affiliation: Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy.

The NLM predicts that as an L1-specific “perceptual map” is developed, the identification of L1 speech sounds will become more robust whereas learning to perceive instances of new, L2 categories may become difficult. This might hold true, for example, if tokens of two distinct L2 vowel categories straddle the center of a vowel category in the L1 of a learner. Tokens of both L2 vowels would tend to be “attracted” by the same L1 prototype, effectively reducing perceptual distance between the L2 vowel tokens. For a native speaker of the L2, on the other hand, the perceived distances between the two tokens would be augmented by the perceptual magnet effect given that the tokens would be differentially categorized, and thus attracted by two different prototypes (see also Flege *et al.*, 1994).

Iverson *et al.* (2003) suggested that an L1-conditioned warping of the auditory properties of incoming speech occurs at an early phonetic or late auditory level of processing and may be difficult to reverse later in life. This view was supported by the results of five studies examining the perception of Catalan vowels by young native Spanish adults who had begun learning Catalan in Barcelona by the age of 6 years. Using different testing techniques, all assessed the perception of Catalan /e/ and /ɛ/, which straddle the category center of Spanish /e/. In each study, the Spanish early learners of Catalan differed from native speakers of Catalan in perceiving Catalan /e/ and /ɛ/ (Bosch *et al.*, 2000; Pallier *et al.*, 1997, 2001; Sebastián-Gallés and Soto-Faraco, 1999). Most recently, Sebastián-Gallés *et al.* (2005) presented Catalan words and nonwords (created by changing /e/ to /ɛ/ and /ɛ/ to /e/) for lexical decision. The early learners obtained significantly lower scores than Catalan native speakers did, and also showed a wider range of scores.

The research on early native Spanish learners of Catalan suggested a number of important conclusions. One was that the speech perception system “does not seem able to easily develop new phonemic categories” after perceptual attunement to the L1 phonological system (Bosch *et al.*, 2000, p. 193). Another was that early learners exhibit a “lack of behavioral plasticity” (Pallier *et al.*, 1997, p. B9) with respect to learning to perceive contrastive phonemes of an L2. Still another was that “there are severe limitations to the malleability of the initially acquired L1 phonemic categories even under conditions of early and extensive exposure” (Sebastián-Gallés and Soto-Faraco, 1999, p. 120).

The results of two studies examining the perception of English vowels by native Italian immigrants to Canada suggested a different conclusion regarding the perceptual ability of early learners. Both of these studies used a triadic categorical discrimination test (CDT) in which the task was to identify the serial position of an odd item out (in change trials) or to indicate that all three tokens were instances of a single category (in no-change trials). Flege *et al.* (1999) tested early native Italian learners of English who reported using Italian relatively often (“high-L1-use” group) or seldom (“low-L1-use” group). The low-L1-use group did not differ from native English (NE) monolinguals in discriminating English vowels whereas the high-L1-use group was found to differ marginally from the NE group. A similar result was obtained by Flege and MacKay (2004), who observed a significant vowel

discrimination difference between NE speakers and an early high-L1-use group but not an early low-L1-use group. The authors interpreted this finding to mean that early exposure to the L2 does not guarantee the formation of new phonetic categories for L2 vowels, but that category formation will be more likely for individuals who use their L1 infrequently.

The research with Spanish learners of Catalan suggested that early learners are unlikely to perceive L2 vowels in a nativelike way whereas the research with Italian learners of English suggested that early learners may closely resemble L2 native speakers, at least those who use the L1 infrequently. The basis for the apparent divergence in findings obtained for early L2 learners in Canada and Spain (Barcelona) is uncertain. It was not reported how often the early learners in Barcelona used their L1 (Spanish). Perhaps they differed from native speakers because they used their L1 as frequently as the “high-L1-use” participants in the Canadian research (*viz.* an approximately equal L1-L2 use). Other possible explanations for the differing results obtained in Canada and Spain include differences in the perceived relation between the L1 and L2 vowels examined, use of diverse testing techniques that might have induced different levels of processing, the age of the participants, and the amount of foreign-accented L2 input the participants might have received.

Obtaining a better understanding of early learners’ perception of L2 vowels is important because it relates to the more general issue of perceptual plasticity in human speech. Native-nonnative differences might be attributed to a loss of plasticity arising either from a reduced capacity to use one or more processes needed to develop language-specific, long-term memory representations for vowels and consonants, or from the persistent influence of prior learning on subsequent learning [e.g., the early-level “warping” of input due to the effect of L1 “magnets;” see Iverson *et al.* (2003)]. Alternatively, native-nonnative differences might be attributed to differences in attention and/or input, or to the inability to sustain an optimal representation of more than one linguistic system (Grosjean, 1989).

Even assuming that the speech perception system remains fully plastic after the L1 sound system has been established, it is probably unreasonable to suppose that early learners’ perception of L2 vowels could ever be identical to that of L2 native speakers. This is because perceptual maps are presumably based on the statistical properties of individuals’ linguistic input (e.g., Kuhl, 2000; Kluender *et al.*, 1998), which can never have been the same for individuals with different L1s. That being the case, identical perception by groups of early learners and L2 native speakers should probably not be used as a litmus test for whether the speech perception system does or does not remain plastic following L1 perceptual attunement.

## B. The present study

The aim of most previous L2 vowel perception research evaluating perceptual plasticity has been to determine if groups of early learners did or did not differ significantly from a control group of L2 native speakers. A more informa-

tive way to assess perceptual learning by early learners may be to compare their performance to that of two control groups, *viz.* L2 native speakers (as in past research) and monolingual native speakers of the early learners' L1. The present study compared the English vowel discrimination of early native Spanish learners of English to that of English monolinguals and Spanish monolinguals. Comparing the early learners' performance to that of English monolinguals provided a way to determine if their perceptual learning was complete. Comparing their performance to that of Spanish monolinguals provided a way to determine how much the early learners had learned, even in the absence of complete learning.

Including Spanish monolinguals ruled out the use of testing techniques that presupposed, even indirectly, a knowledge of English, e.g., forced choice identification or goodness rating. Although the CDT used in the research on Italian learners of English has certain advantages, it does not control response bias as effectively as another discrimination test widely used in cross-language and L2 research, the categorial AXB test (e.g., Best *et al.*, 1996, 2001; Harnsberger, 2001; Polka, 1995; Polka and Bohn, 1996).

A problem for the categorial AXB test, on the other hand, is that it has often yielded near-ceiling discrimination scores (90%–100% correct) for pairs of foreign vowels likely to have been perceived as instances of a single L1 category (Best *et al.*, 1996; Polka, 1995; Polka and Bohn, 1996). This suggests that the AXB paradigm may not constitute a “phonetically sensitive” test of L2 vowel discrimination in the sense that pairs of L2 vowels showing a 2-to-1 cross-language mapping pattern might be successfully discriminated on the basis of low-level (perhaps auditory) memory codes rather than different language-specific phonetic codes. Such a response would present a problem for the present study, given the plan to use the Spanish monolingual control group as a baseline to assess how much perceptual learning the early learners had shown. The Spanish monolingual reference group could not serve this function if it obtained near-ceiling scores due to the use of nonphonetic encoding strategies. Thus, for purposes of this study, an ideal (“phonetically sensitive”) AXB test would be one that yielded near-ceiling scores for English monolinguals but near-chance scores for Spanish monolinguals for English vowels heard as instances of a single Spanish vowel category.

The decision was made to use the categorial AXB paradigm here, but to carry out preliminary research (experiment 1) to find a variant that would reduce scores to near-chance levels for contrasting foreign vowels heard as instances of a single L1 category. This ambitious goal was not fully met in experiment 1, but this experiment did nonetheless reveal a way to make the AXB test more phonetically sensitive. The procedures developed in experiment 1 were therefore used in experiment 2 to assess the discrimination of English vowels by native Spanish early learners, English monolinguals, and Spanish monolinguals.

## II. EXPERIMENT 1

The aim of this experiment was to identify a phonetically sensitive variant of the categorial AXB discrimination test, which would yield near-ceiling scores for listeners who perceive contrasting vowels as distinct phonemes, but near-chance scores for listeners who perceive contrasting foreign vowels as instances of a single L1 category. Such a pattern of scores for Spanish and English monolinguals in experiment 2 would make it possible to assess the extent of perceptual learning by native Spanish early learners. It was not possible to recruit a sufficient number of Spanish monolinguals for this preliminary experiment as well as experiment 2. Therefore, the decision was made to test English adults on vowels drawn from a language they had never heard before, *viz.* Danish.

Previous cross-language research using the AXB discrimination test has often yielded near-ceiling scores for vowels even when tokens drawn from contrasting foreign vowel categories were classified as instances of a single L1 category (Best *et al.*, 1996; Polka, 1995; Polka and Bohn, 1996). This evidence of successful within-category vowel discrimination indicates that listeners are able to base their judgments in an AXB test on more than just language-specific phonetic codes.

Hessen and Schouten (1992) discussed the nature of processes that might be applied in the discrimination of speech sounds that are perceived as either within-category or between-category contrasts. Combining elements from dual processes theory (Fujisaki and Kawashima, 1970) and trace-context theory applied to speech (Macmillan *et al.*, 1988), Hessen and Schouten's (1992) phoneme perception theory (PPT) posits three processes for the encoding of stimuli drawn from a speech sound continuum: trace coding, context (or position) coding, and labeling. In *labeling*, the trace is mapped onto the permanent context (the language-specific sound system) and is given a phonetic label (Hessen and Schouten term it a “phoneme label”). In *context coding*, the trace is mapped onto the experimental context. The listener positions each stimulus in relation to the full range of stimuli presented and gives it a temporary sublabel, possibly relating it to “anchors” among the stimuli in the form of extreme or best category members. The sensory *trace* is stored with all acoustic information intact (except for sensory noise) but decays rapidly after a few hundred milliseconds. Seen from the perspective of the PPT, foreign vowel tokens given different phonetic labels (between-category contrasts) would be discriminated based on the labels, whereas tokens given the same label (within-category contrasts) must be discriminated using context codes or the sensory trace.

Above-chance categorial AXB discrimination of foreign vowels given the same phonetic label might be based on context codes. (Discrimination is less likely to be based on the sensory traces when all stimuli in a trial are physically different.) For example, native English speakers might label (covertly or overtly) the German vowels in an [y]-[y]-[u] trial as “two foreign-sounding u-vowels” followed by “one overly clear English u-vowel” (or some other *ad hoc* label). The use of such a strategy would be unfortunate in experi-

ment 2, where we wished to determine if Spanish early learners manage to discriminate English vowels like native English speakers by generating contrasting phonetic labels (which, of course, implies the establishment of new perceptual English vowel categories).

Accordingly, procedures that might hinder trace or context-coding strategies were evaluated. According to the PPT (Hessen and Schouten, 1992), increased “context variance”—obtained by increasing the range of stimuli presented—limits listeners’ ability to hold context codes in memory. This motivated our decision to evaluate the influence of block structure. In previous applications of the AXB test, the trials testing various contrasts of interest have been presented in separate blocks. This experiment compared the discrimination of four Danish vowel contrasts when presented in separate blocks (small stimulus range) and when presented in one large randomized block (large stimulus range).

A second modification of the standard AXB format evaluated in this experiment was suggested by the results obtained in a study that focused on discrimination of stress placement (Dupoux *et al.*, 2001). The vowel stimuli were produced with either a relatively high or low fundamental frequency ( $F_0$ ). The  $F_0$  variations that occurred within trials were clearly audible, but were independent of category identity. We reasoned that noncritical  $F_0$  variation would make context codes less useful for discrimination by diverting attention away from the dimensions that signaled vowel category identity. In addition, even if categorically relevant acoustic information became part of the context code for a stimulus, uncorrelated or misleading  $F_0$  information might place demands on memory that could negatively affect performance. Finally,  $F_0$  variation would likely also hamper the use of sensory traces.

The questions raised by this methodological experiment were as follows. First, would English monolinguals obtain lower scores in a nonblocked than blocked condition? Second, would the introduction of  $F_0$  variation serve to reduce scores? Third, would the two manipulations just mentioned affect performance on vowels perceived as within-category contrasts but not between-category contrasts (for which discrimination could be based on phonetic labels)?

## A. Methods

### 1. Participants

Most of the 45 English monolinguals (“E-monolinguals”) recruited for this experiment came from the southeastern portion of the U.S. (28 were from Alabama). None of the E-monolinguals reported being able to speak a language other than English, and none had previous exposure to Danish. Fifteen participants each (12 females) were randomly assigned to one of three conditions: “blocked” (the traditional approach in AXB tests), “mixed” (in which trials testing four contrasts were randomly presented in one large block), or “mixed/ $F_0$ ” (one large block,  $F_0$  variation introduced within trials).

Participants assigned to the three conditions differed little in mean age (blocked=27 years; mixed=28 years,

mixed/ $F_0$ =27 years). Three native speakers of Danish (mean age=34 years, 2 females) served as a reference group and took the test in the mixed/ $F_0$  condition. One Dane had been living in the U.S. for two months when tested, the other two for less than 2 weeks. All 48 participants were paid, and all passed a pure tone hearing screening at octave frequencies from 250 to 4000 Hz (25 dB).

### 2. Stimuli

Nonwords were used as stimuli because, in L1 development at least, correct discrimination rates may differ between real words and nonwords (e.g., Stager and Werker, 1997). Had real Danish words been used in the present experiment, the Danes might have enjoyed a lexical advantage not shared by the E-monolinguals.

The first author, a native Danish speaker then living in the United States, selected the four vowel contrasts to be examined based on published descriptions of Danish and English vowels (Grønnum, 1998; Steinlen, 2005) and his own intuitions about which Danish vowels might be difficult for E-monolinguals to discriminate. The inference that E-monolinguals will tend to hear Danish [u:<sup>?</sup>] and [ɔ:<sup>?</sup>] as two distinct English vowels (/u/ and /o/, respectively) was confirmed by a preliminary cross-language classification test in which ten Native English speakers classified the tokens of the eight Danish target vowels used to form the four contrasts. This Danish contrast ([u:<sup>?</sup>]-[ɔ:<sup>?</sup>]) was therefore designated the “easy” contrast. Three other Danish contrasts were designated the “difficult contrasts” (D1, D2, D3) because, as expected, the two vowels in each contrast tended to be classified as a single English vowel: [e:<sub>1</sub>] and [e] 85% of the time, usually as English /u/; [i:<sup>?</sup>] and [e:<sub>1</sub>:<sup>?</sup>] 99% of the time, usually as English /i/; [y] and [ø] 88% of the time, usually as English /u/.

Different flanking consonants were used for the four Danish contrasts: /gl\_s/ for the easy [u:<sup>?</sup>]-[ɔ:<sup>?</sup>] contrast, /fl\_s/ for the [e:<sub>1</sub>]-[e] contrast (D1), /kl\_s/ for the [i:<sup>?</sup>]-[e:<sub>1</sub>:<sup>?</sup>] contrast (D2), and /tj\_s/ for the [y]-[ø] contrast (D3). These consonant sequences are legal in English. One might object that the initial /tj/ context for D3 is “foreign” for English speakers; however, its typical phonetic realization ([tʃ]) closely resembles English [tʃ] (Grønnum, 1998).

The stimuli were produced by a native Danish male and recorded (Shure SM81 microphone, Tascam DA-302 DAT) in a sound booth. All stimuli used in the blocked and mixed conditions, and half of those used in the mixed/ $F_0$  condition, were produced in a declarative carrier sentence (*Det hed virkelig \_\_\_\_\_ engang*, approximately “It really was \_\_\_\_\_ once”). The remaining half of the stimuli used in the mixed/ $F_0$  condition were produced in an interrogative sentence (*Hed det virkelig \_\_\_\_\_ engang?*, “Was it really \_\_\_\_\_ once?”), which yielded a relatively high  $F_0$  in the variable target words.

After the digitization (22.05 kHz, 16 bit), the CCVC stimuli were edited out and normalized for rms amplitude. Eight tokens for each of the eight Danish target vowels were selected. The intent was to select tokens that minimized acoustic phonetic variation within the eight tokens of each vowel category. The exception, of course, was  $F_0$  for tokens

selected for the mixed/ $F_0$  condition. For this condition, the four tokens per vowel category produced in the declarative frame had a mean  $F_0$  of 108 Hz at the vowel midpoint (range=102–117), and the four tokens per vowel produced in the interrogative frame had a mean  $F_0$  of 139 Hz (range =134–146 Hz).

### 3. Procedure

All 48 participants were tested on 256 AXB trials. Each of the four contrasts was tested by 64 trials with all possible AB combinations of the eight tokens of each vowel presented with an equal number of each trial type (AAB, ABB, BAA, and BBA). There was a complete and balanced rotation of the stimuli in the three positions in the triad; the three stimuli in each trial were always physically different. Participants in the blocked condition heard trials testing the four contrasts in separate, counterbalanced blocks (Latin square). For participants in the mixed condition, trials testing the four contrasts were presented in a single randomized block (with appropriate breaks). The same held true for the mixed/ $F_0$  condition which, in addition, introduced variation in  $F_0$ .<sup>1</sup>

The stimuli were presented via headphones at a self-selected comfortable level to the participants, who were tested individually in a sound booth. The participants were told to push a button labeled “1-2” if they heard a match between the first and the second stimuli (i.e., A and X), and the “2-3” button if the second and third stimuli (X and B) matched. The interstimulus interval (ISI) for all trials was 1000 ms. A 1200-ms interval occurred between each response and the following trial.

To ensure that participants understood the task, the test was preceded by a practice block with 20 easy trials consisting of stimuli not used in the experiment. Feedback was provided during this block, but not during the experiment. Also unlike the experiment, trials could be replayed. All participants were required to respond correctly to at least 18/20 practice trials before proceeding to the experiment. All did so within four attempts.

The same 256 trials were presented in the blocked and the mixed conditions. In the mixed/ $F_0$  condition, half of tokens had relatively high (H)  $F_0$  and half had a relatively low (L)  $F_0$ . Each trial in this condition manifested one of the following types of  $F_0$  pattern. In “no  $F_0$  change” trials, all three tokens had either H or L  $F_0$ . In “neutral” trials,  $F_0$  could not influence responses because the  $F_0$  of the X stimulus differed from both A and B (HLH or LHL). In “congruent”  $F_0$  trials,  $F_0$  similarities “suggested” the correct response because the  $F_0$  level of the X stimulus was the same as that of the categorically same vowel. Conversely, in “incongruent” trials, the  $F_0$  suggested the wrong response, i.e., participants would always go wrong if they based their responses solely on  $F_0$  similarities among the three stimuli in a trial (e.g., HHL  $F_0$  in an [y]-[ø]-[ø] trial).

### B. Results

As expected, the three Danes obtained high scores for the “easy” contrast ( $M=100\%$ ) and also the three difficult contrasts (D1  $M=99\%$ , D2  $M=96\%$ , D3  $M=94\%$ ). The

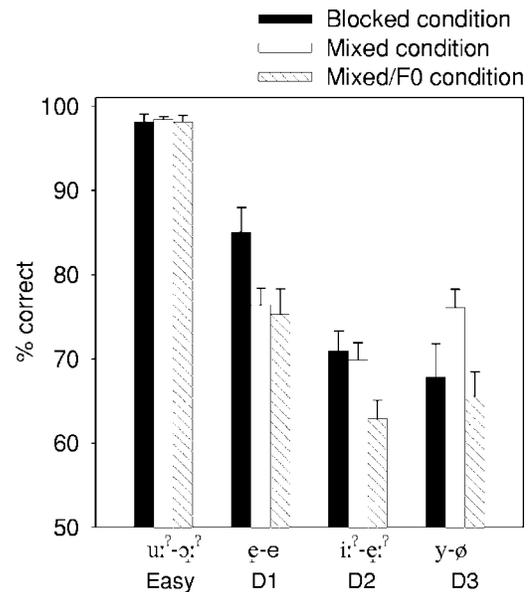


FIG. 1. Mean percent correct AXB discrimination of one “easy” and three “difficult” (D1, D2, D3) Danish vowel contrasts by three groups of English monolinguals assigned to one condition each. Error bars bracket 1 standard error.

E-monolinguals obtained high scores for the easy contrast ( $M=98\%$ ) but, as shown in Fig. 1, substantially lower scores for the three difficult contrasts. Their scores for the difficult contrasts ranged from 63% correct for [i:~e:], in the mixed/ $F_0$  condition, to 85% correct for [e~e], in the blocked condition. Single sample  $t$  tests showed that the E-monolinguals’ scores for all three difficult contrasts significantly exceeded the chance level of 50% correct in all three testing conditions [ $t(14)=4.4$  to  $13.5$ ,  $p<0.001$ ].

The E-monolinguals’ scores were submitted to an ANOVA in which condition (blocked, mixed, mixed/ $F_0$ ) served as a between-subjects factor and vowel contrast (four levels) was a repeated measure. The ANOVA yielded a significant interaction [ $F(6, 126)=4.2$ ,  $p<0.001$ ], which was explored through tests of simple effects. The simple effect of condition was nonsignificant for the easy contrast [ $F(2, 42)=0.1$ ,  $p=0.95$ ] but significant or marginally significant for the three difficult contrasts [D1  $F(2, 42)=3.9$ ,  $p=0.027$ ; D2  $F(2, 42)=3.9$ ,  $p=0.029$ ; D3  $F(2, 42)=3.2$ ,  $p=0.051$ ]. Tukey tests revealed that lower scores were obtained for D1 and D2 in the mixed/ $F_0$  condition than in the blocked condition ( $p<0.03$ ). No other differences between conditions reached significance.<sup>2</sup> The lack of a significant difference between the blocked and mixed conditions indicated that simply combining the trials testing the four contrasts in a single randomized block was not sufficient to lower discrimination scores. This suggested that the difference between the blocked and mixed/ $F_0$  conditions was largely due to the manipulation of  $F_0$ . The effect of  $F_0$  differences in the mixed/ $F_0$  condition was therefore explored further.

The E-monolinguals’ scores for the easy contrast were at ceiling (98% to 99%) regardless of the  $F_0$  type of trial. However, for all three difficult contrasts, scores were higher in trials manifesting a “congruent”  $F_0$  type (D1 89%, D2 80%, D3 81%) than an “incongruent”  $F_0$  type (D1 63%, D2 50%,

D2 55%). Put another way, higher scores were obtained when reliance on *F0* similarities would yield a correct response than when reliance on *F0* would yield an incorrect response. Intermediate scores were obtained for trials with no *F0* change (D1 80%, D2 67%, D3 68%) and trials with a “neutral” *F0* type, which suggested neither the correct nor the incorrect response (D1 80%, D2 66%, D3 69%).

### C. Discussion

This experiment aimed to develop AXB testing procedures that would yield near-ceiling scores for vowels heard as distinct phonemes (and thus given different labels) but near-chance discrimination scores for foreign vowels heard as instances of a single L1 category (and thus given the same label). This goal was only partially met.

As intended, the E-monolinguals assigned to all three conditions (testing formats) obtained near-ceiling scores for the easy Danish contrast (as did the native Danish reference group for the easy contrast and all three difficult contrasts). It is likely that the E-monolinguals obtained near-ceiling scores for the easy contrast by generating two different phonetic label codes when hearing the Danish vowels making up this contrast. Also as intended, the E-monolinguals obtained much lower scores for three difficult Danish contrasts.

Contrary to our intent, however, the E-monolinguals' scores for all three difficult contrasts were significantly above chance in all conditions. Importantly, though, significantly lower scores were obtained for two contrasts when the effect of both manipulations examined here (i.e., mixed block structure and *F0* variation) were combined (mixed/*F0* condition) than when standard AXB testing procedures were used (blocked condition). Assuming that the E-monolinguals based discrimination of the three difficult contrasts on context codes (Hessen and Schouten, 1992), the results suggested that the modifications hindered but did not completely prevent the use of such codes.

The effect of increasing the range of stimuli by presenting four contrasts randomly in a single block (mixed condition) rather than blocked by contrast (blocked condition) was nonsignificant. This suggested that the participants in this experiment did not generate context codes by positioning the stimuli in relation to other stimuli in the experimental context, but rather in relation to phonetic categories stored in long-term memory, a possibility also envisaged by Hessen and Schouten (1992). Context variance has previously been shown to be range dependent in experiments using synthetic stimuli (e.g., Macmillan *et al.*, 1988). The absence of a range effect in the present study, using natural rather than synthetic stimuli, raises the possibility that neither between-category nor within-category discrimination is range dependent for natural stimuli.

The significant differences between the blocked and mixed/*F0* condition, but the lack of differences between the blocked and mixed conditions, suggested that differences between blocked and mixed/*F0* were due largely to the *F0* manipulation. Additional consideration of scores in the four *F0* types in the mixed/*F0* condition suggested that *F0* variation might have been more effective in lowering scores had

only the incongruent *F0* type been used. The scores for the incongruent *F0* type averaged just slightly above chance (56% correct) for the three difficult contrasts. The decision was made, therefore, to apply the procedures of the mixed/*F0* condition in experiment 2 but only to use the incongruent and the neutral *F0* types.

### III. EXPERIMENT 2

This experiment compared the discrimination of English vowels by native Spanish adults who had begun learning English in the U.S. as children (“early learners”) to that of monolingual native speakers of English (“E-monolinguals”) and Spanish (“S-monolinguals”). One of the English vowel contrasts examined here was expected to be easy for the S-monolinguals to discriminate because the two vowels in it were likely to be heard as distinct Spanish vowels. However, the other three contrasts (designated D1, D2, D3) were expected to be difficult because the contrasting vowels were likely to be heard as instances of a single Spanish vowel category. Pilot tests with late learners (native Spanish speakers who had begun learning English as adults) on a range of English vowel contrasts suggested that the difficult contrasts examined here were the three most difficult English vowel contrasts for native Spanish speakers, at least for vowels spoken by the native English speaker who provided the stimuli.

In experiment 1, E-monolinguals with no prior exposure to Danish obtained significantly above-chance discrimination scores for Danish vowels that were heard as instances of a single English vowel category. This suggested that even the most difficult condition (i.e., the mixed/*F0* condition, in which trials testing four contrasts were mixed together and *F0* variation was added to the stimuli) did not prevent completely the use of low-level or context codes. Two additional steps were therefore taken in this experiment to further reduce percent correct scores. This was considered important to do given the nature of the question addressed by this experiment: Would the early learners' discrimination of English vowels more nearly resemble that of S-monolinguals or E-monolinguals?

One step involved the *F0* manipulation. In experiment 1, the *F0* manipulation lowered E-monolinguals' scores for difficult contrasts mainly when the *F0* difference among the three stimuli in a trial was of the “incongruent” *F0* type. Therefore, only the incongruent *F0* type was included here, along with the “neutral” *F0* type, to maintain some variability.

The second step was motivated by the evidence (Cowan and Morse, 1986) that vowel discrimination is poorer at an ISI of 0 ms than 250, 500, 1000, or 2000 ms (both for within- and between-category trials).<sup>3</sup> We therefore included a 0-ms ISI as well as the 1000-ms ISI used in experiment 1. As described earlier, the PPT of Hessen and Schouten (1992) posits that label codes are generated rapidly, whereas context codes—essentially refinements of the phonetic code—are generated more slowly. If the ISI between items to be discriminated is too short, context codes might be stored in an incomplete and inaccurate form. This should render them less reliable for correct discrimination, thereby lowering

TABLE I. Mean characteristics of the three groups of participants in experiment 2. Ranges are in parentheses.

	Early learners	E-mono-linguals	S-mono-linguals
Gender	13 f, 7 m	11 f, 9 m	8 f, 12 m
Chronological ages	31 (20–53)	31 (21–45)	29 (19–46)
Years of education	16 (13–22)	17 (13–21)	10 (3–20)
Age of exposure to E <sup>a</sup>	6 (2–10)	...	...
Years of exposure to E <sup>a</sup>	25 (9–45)	...	...
Self-rated S ability <sup>b</sup>	6.0 (3–7)	...	...
Self-rated E ability <sup>b</sup>	6.6 (4–7)	...	...

<sup>a</sup>English.

<sup>b</sup>Self-rated Spanish (S) or English (E) ability, 1=poor, 7=good.

scores in cases where a label code is insufficient. Such interruption of the generation of context memory codes can probably be understood as informational backward recognition masking (see Watson, 2005). For ISI=0 ms, the second and third stimuli in the AXB triads will serve as masks for preceding stimuli. Subsequent masks have been shown to interrupt auditory processing of a stimulus. For example, Massaro (1970) found that identification of a test tone was impeded by an immediately following masking tone but improved with increasing ISI.

## A. Methods

### 1. Participants

The 60 participants were recruited in Birmingham, AL. All were required to be between the ages of 19 and 55 years, to report normal hearing, and to pass a pure-tone hearing screening. Twenty participants each were assigned to one of three groups: “S-monolinguals” (Spanish monolinguals), “E-monolinguals” (English monolinguals), and “early-learners” (native Spanish adults who had begun learning English as an L2 in childhood). As shown in Table I, the three groups did not differ significantly in chronological age [ $F(2,57)=0.51$ ,  $p=0.6$ ], but did differ in years of formal education [ $F(2,57)=30.5$ ,  $p<0.001$ ]. A Tukey test indicated that the S-monolinguals had received less education than the E-monolinguals and early learners ( $p<0.001$ ), who did not differ from one another.

E-monolinguals were required to be native speakers of American English, to have NE-speaking parents, to have never lived outside the U.S., and to be unable to speak a language other than English proficiently. The S-monolinguals were required to have been born and raised in a predominantly Spanish-speaking country (Mexico—16, Guatemala—1, Venezuela—1, Argentina—1, Colombia—1). Despite the fact that the S-monolinguals were living in Birmingham when tested (mean residence in the U.S.=3.8 years, s.d.=4.6), they had either very limited or no profi-

ciency in English according to self-report. To verify this, the experimenter vocally asked candidates for inclusion in the S-monolingual group eight simple questions in English (e.g., *What day was it yesterday?* or *What kind of food do you like to eat?*). The S-monolinguals were able to provide a response—often just a single word—to an average of just 2.7 of the 8 questions (range=0–6).

The early learners were required to have learned Spanish as an L1 from native Spanish-speaking parents and to have learned English as an L2 in childhood. Sixteen early learners were born in a predominantly Spanish-speaking country and had immigrated to the U.S. in childhood. One was born in the U.S. but moved soon thereafter to Mexico before returning to the U.S. later in childhood. The remaining three were born and raised in a predominantly Spanish-speaking community in the U.S.; they reported being unable to understand English upon entering preschool or elementary school.

Participant characteristics are shown in Table I. The age at which the early learners had begun learning English was indexed by self-reported age of first exposure (“AOE”) to English. Their English-language experience was indexed by self-reported years of exposure (“YOE”) to English. All participants were paid.

### 2. Stimuli

A total of 64 English vowel stimuli were obtained in the following way. A list of nonwords was created by inserting eight English vowels ([i u ɪ e<sup>ɪ</sup> a ʌ ʊ o<sup>ʊ</sup>]) into a CC\_C frame. A native English male produced the nonwords in declarative and interrogative sentence frames (*I say\_\_\_\_\_now again; Can you say\_\_\_\_\_now again?*). As in experiment 1, different consonants flanked the vowel pairs of interest: /gl\_s/ for vowels forming the [i]-[u] contrast, /kl\_s/ for the [ɪ]-[e<sup>ɪ</sup>] contrast, /tʃ\_s/ for the [a]-[ʌ], and /fl\_s/ for [ʊ]-[o<sup>ʊ</sup>] contrast. All of the consonant sequences are legal in both Spanish and English.

The sentences were recorded and the CCVC nonword stimuli edited as in experiment 1. For each vowel category, four tokens were taken from the declarative sentence frame and four from the interrogative sentence frame.  $F_0$  values were measured using an autocorrelation technique by placing a 40-ms window at the vowel midpoint. As shown in Table II, the  $F_0$  values were higher for stimuli taken from the interrogative than the declarative frame, and so the two sets of stimuli were designated the “high- $F_0$ ” (or H) and “low- $F_0$ ” (L) stimuli.

The frequencies of the first two vowel formants ( $F_1, F_2$ ) of the English stimuli at the vowel midpoint have been plotted in Fig. 2 along with mean  $F_1$  and  $F_2$  values for the five vowels of Spanish [as spoken in a /b\_t/ context by adult males, from Bradlow (1995)]. The perceived phonetic distance of vowels drawn from two languages cannot be accurately determined by means of distances in an  $F_1$ - $F_2$  space. However, the distances evident in Fig. 2 suggest that S-monolinguals might tend to classify the English [ɪ] and [e<sup>ɪ</sup>] tokens as instances of Spanish /e/, the [a] and [ʌ] tokens as Spanish /a/, and the [ʊ] and [o<sup>ʊ</sup>] tokens as Spanish /o/. The [i] and [u] tokens might be classified as instances of Spanish

TABLE II. Mean vowel duration and fundamental frequency ( $F_0$ ) values at vowel midpoint for stimuli produced in interrogative and declarative sentence frames. Ranges are in parentheses.

	$F_0$ (Hz)		Duration (ms)	
	Interrogative	Declarative	Interrogative	Declarative
[i]	170 (164–178)	116 (110–120)	149 (143–158)	113 (109–117)
[u]	176 (170–182)	110 (107–111)	128 (122–133)	107 (97–121)
[ɪ]	168 (160–177)	117 (113–123)	110 (101–116)	94 (93–95)
[eʰ]	169 (163–174)	114 (113–114)	157 (153–161)	132 (121–140)
[ɑ]	163 (153–170)	121 (119–122)	173 (160–188)	151 (132–163)
[ʌ]	167 (160–170)	121 (113–132)	118 (110–124)	114 (110–121)
[ʊ]	172 (166–176)	136 (118–148)	133 (129–134)	113 (91–126)
[oʊ]	163 (158–172)	129 (124–139)	178 (170–185)	162 (148–171)

/i/ and /u/, respectively, at least if the fronting of the English [u] did not cause Spanish speakers to perceive tokens of this category as being instances of Spanish /i/.<sup>4</sup>

Inferences based on formant values were formally evaluated following the discrimination experiment described below. The 64 English stimuli from the discrimination test were aurally presented a single time in quasi-random order to the S-monolinguals. Their task was to classify each token as being most like one of the five vowels of Spanish, “i”, “e”, “a”, “o”, or “u”. (Data for one S-monolingual were lost due to equipment malfunction.) In agreement with the acoustic data, the English [ɪ] and [eʰ] tokens were usually classified as Spanish /e/ (128 and 124 judgments, respectively, out of a maximum 152); the [ɑ] and [ʌ] tokens were usually classi-

fied as Spanish /a/ (in 152 and 118 judgments); and the [ʊ] and [oʊ] tokens were usually classified as Spanish /o/ (116 and 134 judgments). The [i] and [u] tokens were classified as different vowels, Spanish /i/ (151 judgments) and /u/ (134 judgments).

### 3. Procedure

All 60 participants were tested individually in a sound booth where stimuli were presented via headphones at a self-selected level. The 64 trials testing each of the four contrasts were presented in one block of 256 trials. Half of the trials had the neutral  $F_0$  type (which suggested neither the correct nor the incorrect response), and the remaining trials had the incongruent  $F_0$  type, which suggested an incorrect response (e.g., LLH  $F_0$  in an [ɑ]-[ʌ]-[ʌ] trial). The block containing all 256 trials was presented two times in counterbalanced order with an ISI of either 0 or 1000 ms. In both ISI conditions, a 1200-ms interval occurred between each response and the following trial. A separate randomization was used for each participant, but was maintained across the two ISI conditions.

The 256 trials presented in both ISI conditions were preceded by 32 extra trials. Instructions, practice, and admission criterion were similar to those described in experiment 1. All test materials and instructions were translated and given in Spanish to the S-monolinguals. During the interval between blocks, participants responded to a language background questionnaire and produced sentences used for another study.

## B. Results and discussion

### 1. Preliminary analysis

Lower scores were again expected for trials with the incongruent than the neutral  $F_0$  type. This expectation was tested by a (3) group  $\times$  (4) vowel contrast  $\times$  (2) ISI  $\times$  (2)  $F_0$  type ANOVA examining the 16 mean scores obtained for each participant. Scores were indeed significantly lower

- English vowels (this study)
- Spanish reference vowels (Bradlow, 1995)

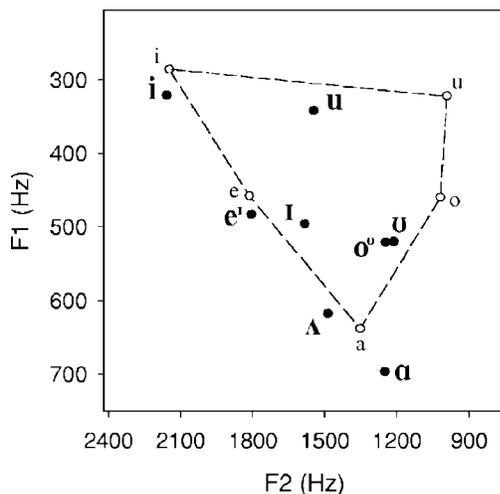


FIG. 2. Vowel chart showing first formant ( $F_1$ ) and second formant ( $F_2$ ) mean values at vowel midpoint of each of the eight English vowels used in experiment 2 (in bold). The mean values for Spanish are from Bradlow (1995).

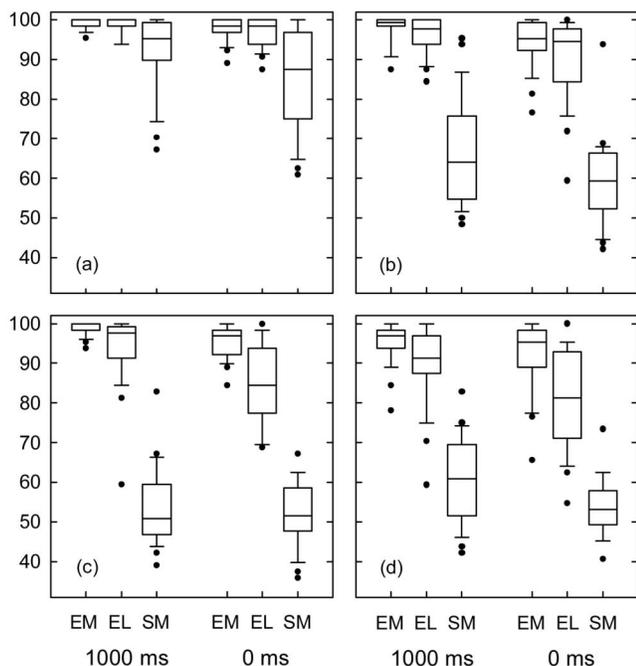


FIG. 3. Boxplots showing percent correct AXB discrimination obtained by English monolinguals (EM), early learners (EL), and Spanish monolinguals (SM). The four panels each display the results for one of the four vowel contrasts, i-u (a), a-ʌ (b), i-eʰ (c), and ʊ-oʊ (d), obtained by the three groups at a 1000 and a 0-ms ISI. Filled circles indicate individual outliers.

overall for trials with an incongruent than neutral  $F_0$  type [ $F(1, 57) = 34.2, p < 0.001$ ]. However, the lack of a group  $\times$  vowel contrast  $\times$   $F_0$  type interaction [ $F(6, 171) = 0.9, p = 0.53$ ] suggested that the relative effect of  $F_0$  type on the four contrasts was similar across the three groups. In all subsequent analyses, therefore, scores were averaged over the two  $F_0$  types.

## 2. Between-group differences

As shown in Fig. 3, the mean percent correct scores were higher overall for the easy contrast ([i]-[u],  $M = 95\%$  correct) than for the three difficult contrasts ([i]-[eʰ] 80%, [a]-[ʌ] 84%, [ʊ]-[oʊ] 79%). The early learners' scores were slightly lower ( $M = 91\%$ ) than the E-monolinguals' ( $M = 96\%$ ), but substantially higher than the S-monolinguals' ( $M = 65\%$ ). Lower scores were obtained, on average, for ISI=0 ms than ISI=1000 ms ( $M = 82$  vs. 87%). Of greatest interest, the S-monolinguals' scores for the three difficult contrasts at ISI=0 ms were all near chance (<60% correct). A series of single sample  $t$  tests indicated that S-monolinguals' scores for one difficult contrast, [i]-[eʰ], did not significantly exceed the chance level of 50% at ISI=0 ms [ $t(19) = 1.2, p = 0.24$ ] or ISI=1000 ms [ $t(19) = 1.6, p = 0.12$ ].

The eight scores obtained for each participant (4 contrasts  $\times$  2 ISIs) were examined in a mixed-design (3) group  $\times$  (2) ISI  $\times$  (4) vowel contrast ANOVA. All three main effects reached significance [group  $F(2, 57) = 175.3, p < 0.001$ ; ISI  $F(1, 57) = 53.6, p < 0.001$ ; vowel contrast  $F(3, 171) = 95.8, p < 0.001$ ] as did the three-way interaction [ $F(3, 171) = 2.4, p = 0.033$ ], which was explored through simple effects tests. As seen in Fig. 3, an important source of

the interaction was a differing pattern of between-group differences in the two ISI conditions. The group effect was significant for all four vowel contrasts at both ISI = 1000 ms [ $F(2, 57) = 9.7$  to 180.3,  $p < 0.001$ ] and ISI = 0 ms [ $F(2, 57) = 15.4$  to 158.6,  $p < 0.001$ ]. Tukey tests ( $\alpha = 0.05$ ) revealed that at ISI=1000 ms, the S-monolinguals' scores were significantly lower for all four contrasts than those obtained by the E-monolinguals and early learners, who did not differ significantly. For ISI=0 ms, the S-monolinguals' scores were lower for all four contrasts than the E-monolinguals'. Importantly, the early learners' scores were lower than the E-monolinguals' for two difficult contrasts, [i]-[eʰ] and [ʊ]-[oʊ]. Although the early learners' scores for difficult contrasts were closer to the E-monolinguals' than to the S-monolinguals' (see Fig. 3), these differences suggested that the early learners' perception of English vowels was not "functionally equivalent" (see Flege and MacKay, 2004) to that of the E-monolinguals.

The lower scores obtained for the easy contrast by the S-monolinguals than by the early learners and E-monolinguals might seem surprising given that all but one S-monolingual classified the English /i/ and /u/ stimuli as instances of two different Spanish vowels (*viz.* Spanish /i/ and /u/). The S-monolinguals' lower scores might be attributed to acoustic phonetic differences between the /i/s and /u/s found in Spanish and English (see Fig. 2).<sup>5</sup> Note, however, that a much smaller difference existed between the E-monolinguals and S-monolinguals for the easy contrast ( $\Delta = 7\%$ ) than for the three difficult contrasts ( $\Delta = 31\% - 45\%$ ).

The simple effect of contrast was significant for all six group  $\times$  ISI combinations [ $F(3, 57) = 4.3$  to 65.5,  $p < 0.01$ ], but the pattern of between-contrast differences varied as a function of group and ISI. The E-monolinguals obtained lower scores for [ʊ]-[oʊ] than [i]-[u] and [i]-[eʰ] at ISI = 1000, but [ʊ]-[oʊ] and [i]-[eʰ] did not differ at ISI=0. The S-monolinguals obtained higher scores for [i]-[u] than for all three difficult contrasts at both ISI=0 and ISI=1000, and higher scores for [a]-[ʌ] than [i]-[eʰ] at ISI=1000 ms. For early learners, higher scores were obtained for [i]-[u] than all three difficult contrasts at ISI=0, for [i]-[u] than [i]-[eʰ] and [ʊ]-[oʊ] at ISI=1000 ms, and for [a]-[ʌ] than [ʊ]-[oʊ] at both ISIs.

Tests of the simple effect of ISI revealed that significantly lower scores were obtained at ISI=0 than ISI = 1000 ms for 10 of the 12 contrast  $\times$  group combinations [ $t(19) = 2.1$  to 4.8,  $p < 0.05$ ]. The two exceptions were the highest and the lowest scores obtained, *viz.* the score obtained for the easy contrast by the E-monolinguals, and for the [i]-[eʰ] contrast by the S-monolinguals. The first exception might well have been a ceiling effect (with mean values in the two ISI conditions of 98% and 99% correct), the second a floor effect ( $M = 52\%$  and 54%).

The effect of ISI for the E-monolinguals—for whom all trials were presumably between-category trials—was unexpected. This finding suggests that ISI influences both the between- and within-category discrimination of natural vowels [as has previously been found for synthetic vowels, see Cowan and Morse (1986)]. Perhaps between-category dis-

TABLE III. Mean characteristics of the four early learners who obtained the highest discrimination scores, and the four obtaining the lowest scores. Ranges are in parentheses.

Variable	Context	4 Highest	4 Lowest
Age <sup>a</sup>	...	36 (22–50)	31 (20–39)
Age of exposure to E	...	3 (2–5)	8 (5–10)
Years of exposure to E	...	33 (21–45)	23 (10–34)
Years of S education <sup>b</sup>	...	0 (0)	3 (0–4)
Years of E education <sup>b</sup>	...	18 (13–20)	13 (10–17)
% E use—last 1 <sup>c</sup>	overall	75 (50–90)	68 (50–90)
% E use—last 5 <sup>c</sup>	overall	65 (50–80)	70 (50–90)
% E use—last 5 <sup>c</sup>	at home	75 (30–100)	58 (0–90)
% E use—last 5 <sup>c</sup>	with friends	90 (70–100)	60 (50–70)
% E use—first 5 <sup>c</sup>	in general	55 (50–70)	25 (10–40)
% E use—first 5 <sup>c</sup>	at home	48 (40–50)	8 (0–30)
% E use—first 5 <sup>c</sup>	with friends	88 (80–100)	43 (0–90)
% S-accented E <sup>c</sup>	last 5	15 (0–30)	30 (10–60)
% S-accented E <sup>c</sup>	first 5	20 (0–70)	33 (10–70)
E ability <sup>d</sup>	...	7 (7–7)	5 (4–6)

<sup>a</sup>Chronological age, in years.

<sup>b</sup>Years of education in Spanish (S) or English (E) medium schools.

<sup>c</sup>Self-estimated percentage in the last 1, last 5, or first 5 years in the U.S.

<sup>d</sup>Self-rated ability (1=poor, 7=good).

crimination was based to some extent on the ISI-sensitive context codes in addition to phonetic label codes.

### 3. Individual differences

The large difference in scores obtained by the E-monolinguals and S-monolinguals suggests that the AXB test used here was phonetically sensitive, perhaps by virtue of hindering use of context codes (Hessen and Schouten, 1992). In the ISI=0 ms condition, the scores obtained by individual E-monolinguals and S-monolinguals were virtually nonoverlapping for the three difficult contrasts. (The exception was a single S-monolingual outlier for [a]-[A]<sup>6</sup> and [ʊ]-[o<sup>ʊ</sup>].) Likewise, the scores obtained for individual early learners and S-monolinguals showed virtually no overlap at ISI=0 ms (the exceptions being one outlier in each group for [a]-[A] and [ʊ]-[o<sup>ʊ</sup>]), which provided direct evidence of perceptual learning by nearly all early learners.

Given the theoretical importance of this finding, it is worth considering in greater detail the scores obtained by individual early learners at ISI=0 ms. Averaged over all three difficult contrasts, the scores obtained for the 20 E-monolinguals ranged from 81% to 99% correct, as compared to a range of 64% to 98% for the early learners. The four lowest scores obtained for early learners fell below the E-monolinguals' range (but were nonetheless higher than those obtained for all but one S-monolingual), whereas the four highest-scoring early learners all exceeded the mean score obtained for the E-monolingual group.

Examination of the participants' questionnaires revealed, not surprisingly, that all four of the lowest-scoring early learners reported less proficiency in English than all four of the highest-scoring early learners. In fact, of all 20 early learners, only the four learners who scored below the native English range did not assign themselves the maximum value of "7" when self-rating their ability to speak English.

What enabled some early learners to discriminate English vowels so much better than others? As summarized in Table III, three of the lowest-scoring early learners were exposed to English later in life (i.e., had a later AOE) than all four highest-scoring early learners. Three of the highest-scoring learners reported using English more with friends in the past 5 years than all four of the lowest-scoring learners. And all four highest-scoring learners reported more use of English during their first 5 years of exposure to English than all four of the lowest-scoring learners. Finally, three of the four highest-scoring learners reported having attended school in the U.S. longer than all four of the lowest-scoring early learners. Although these data are only suggestive given the small number of participants (four per subgroup), the pattern seen here suggests that differences in the age of first exposure and amount of L2 use (especially during the first 5 years of L2 exposure) were responsible for important individual differences among the early learners.

### IV. GENERAL DISCUSSION

This study evaluated the discrimination of English vowels by native Spanish adults who were first exposed to English in childhood (early learners). The study used a variant of the categorical AXB discrimination test that appeared to hinder vowel discrimination based on low-level or context memory codes (as opposed to language-specific phonetic codes, see experiment 1). The findings relate to the more general question of the extent to which the speech perception system remains plastic following attunement to the L1 sound system in early childhood.

The two English vowels in each of three "difficult" contrasts were both heard by Spanish monolinguals (S-monolinguals) as instances of a single Spanish vowel category. These contrasts were expected to be difficult to discriminate for the S-monolinguals and also for the early

learners unless they had developed new long-term memory representations (phonetic categories) for one or both vowels in the contrasts. The two vowels making up a fourth contrast were heard by S-monolinguals as distinct Spanish vowels. This contrast was expected to be easy to discriminate for Spanish speakers with or without perceptual learning of English vowels.

The AXB test was administered at two ISIs, 1000 vs. 0 ms, and was highly phonetically sensitive in the ISI = 0 ms condition in that listeners with English-specific phonetic categories (the E-monolinguals) obtained scores near ceiling for the difficult English contrasts whereas listeners without such categories (the S-monolinguals) obtained scores that were at or near chance. Given this characteristic of the discrimination test, it was well suited to reveal perceptual learning by the early learners. The early learners obtained significantly higher scores than the S-monolinguals at both ISIs and did not differ from the E-monolinguals for any contrast at ISI = 1000 ms. However, the early learners obtained lower scores than the E-monolinguals for two difficult contrasts at ISI = 0 ms, suggesting that their long-term memory representations for English vowels were not identical to the E-monolinguals'.

Significant differences between early learners and L2 native speakers have often been interpreted as evidence for a loss of perceptual plasticity following attunement to the L1 sound system (e.g., Bosch *et al.*, 2000; Pallier *et al.*, 1997; Sebastián-Gallés and Soto-Faraco, 1999). However, as mentioned in the Introduction, there are *a priori* reasons to expect the perceptual systems of early learners to differ from those of L2 native speakers, and thus to expect differences in performance in sensitive tests such as the one used here. It is for this reason that including the S-monolingual group was important for considering whether the early learners' perceptual system remained plastic. Except for one outlier in each group, the scores obtained for individual early learners and S-monolinguals for the three difficult contrasts were non-overlapping at ISI = 0 ms. In our opinion, this finding provides evidence of extensive perceptual learning in early learners and is inconsistent with the view that the speech perception system loses plasticity (e.g., Pallier *et al.*, 1997; Sebastián-Gallés and Soto-Faraco, 1999), and that new phonetic categories become unlikely (Bosch *et al.*, 2000) following attunement to the L1 sound system.

Research with early learners has tended to focus on between-group differences with a view to determining whether the speech perception system does, or does not, remain plastic. However, as usually holds true in L2 perception research (e.g., Sebastián-Gallés *et al.*, 2005), we observed important individual differences among the early learners tested here. This suggests that an important question to ask in future research is: "What factors influence early learners' L2 speech perception?" A detailed analysis of the scores obtained by individual early learners identified several factors that will need to be examined in greater detail in future research. Just four (of 20) early learners obtained average scores for the three difficult contrasts that fell below the E-monolingual range at ISI = 0 ms. These four early learners differed from the four highest-scoring early learners in sev-

eral potentially important ways: The lowest scoring early learners reported using English less often, especially during their first years of English use, than the highest scoring early learners. This raises the question of how much L2 input early learners must receive, and when, if they are to show a native-like perception of L2 phonetic segments.

A related question is whether the *kind* L2 input early learners receive affects their performance. For example, in a study examining Spanish speakers who had begun learning their L2 (Catalan) in Barcelona by the age of 4 years, Sebastián-Gallés *et al.* (2005) obtained scores ranging from chance level to values well above the native speaker mean for a vowel contrast not found in the L1 (Spanish). The authors noted that the numerous native Spanish speakers in Barcelona tend to realize both Catalan /e/ and /ɛ/ as an [e]-quality vowel, raising the possibility that the intersubject variability might be explained by how often the early learners had heard their L2 spoken by fellow native speakers of Spanish who neutralize the Catalan contrast.

Language use patterns, of course, are related to language dominance. The Spanish-Catalan bilinguals examined by Sebastián-Gallés and Soto-Faraco (1999) reported higher proficiency in their L1 Spanish than their L2 Catalan. This might explain why they differed from Catalan native speakers in perceiving Catalan speech sounds. It is noteworthy that L2 perception by clearly L2-dominant early learners examined by Flege and MacKay (2004) did not differ from that of L2 native speakers. If the L1 phonetic subsystem in bilinguals becomes weak through lack of use, the possibility exists that effects of cross-language phonetic interference will be minimized, and performance in the L2 will eventually become native-like.

Indirect support for this hypothesis comes from recent research on Korean adults adopted by French families between the ages of 2 and 9 years, which suggested that early attunement to the L1 sound system may disappear following years of disuse (Pallier *et al.*, 2003; Ventureyra *et al.*, 2004). This interesting research raises the possibility that the L2 performance of early bilinguals with little or no continued L1 use for many years might become indistinguishable from that of native speakers. If so, it would suggest that perceptual learning is limited by language use or dominance patterns rather than a loss of plasticity arising from maturational constraints or irreversible effects of prior learning on subsequent learning. This hypothesis would be strongly supported if clearly L2-dominant early bilinguals with little L1 use could be shown to perform like native speakers in their L2, but show below-native performance in their L1.

The highest-scoring and lowest-scoring early learners also differed in terms of age of first exposure to English in the U.S. ( $M=3$  vs. 8 years). Prior research of "age" effects has focused on differences between early and late L2 learners. However, the differences between the highest- and lowest-scoring early learners in experiment 2 suggested that age effects might exist among participants traditionally classified as "early" learners. Given the small sample size, however, there is a need for additional research, in which factors confounded with age are controlled. If L2 discrimination is indeed better for those first exposed to the L2 at the age of 3

than 8 years, this might be due to a differing effect of L1-on-L2 phonetic interference arising from differences in the state of development of L1 phonetic categories (e.g., Flege, 2003).

In summary, support for the existence of extensive perceptual plasticity in early learners was provided by the finding that most early learners showed nativelike discrimination of three difficult English vowel contrasts which were discriminated at near-chance levels by Spanish monolinguals. A few early learners obtained scores below the native English range, indicating that early exposure to an L2 is insufficient in itself to guarantee nativelike discrimination of L2 vowels. Additional research will be needed to clarify which factors influence early learners' perception of L2 phonetic segments, and to what extent bilinguals can show nativelike performance in both of their languages.

## ACKNOWLEDGMENTS

This research was supported by NIH Grant No. DC000257. Additional support was provided to the first author by the Danish Research Council for the Humanities and by the Oticon Foundation. We wish to thank A. Bradlow and three anonymous reviewers for very helpful comments on our original submission, and O.-S. Bohn and G. Morrison for comments on a previous draft.

<sup>1</sup>In the blocked condition, eight extra trials were presented before each 64-trial block as a "warm-up." In mixed and mixed/*F0* conditions, the 256-trial block was preceded by 32 extra trials. All four vowel contrasts were presented in the extra trials, which were not analyzed. All participants received a different randomization of the stimuli.

<sup>2</sup>The simple effect of contrast was significant for all three groups (*F* values ranging from 43.6 to 61.5,  $p < 0.001$ ). Tukey tests showed that higher scores were obtained for the easy contrast than for all three difficult contrasts in each testing condition ( $p < 0.05$ ). For participants assigned to the blocked and mixed/*F0* conditions, higher scores were obtained for difficult 1 than for difficult 2 and 3; in the mixed condition, higher scores were obtained for difficult 1 and 3 than for difficult 2.

<sup>3</sup>The effect for within-category trials might be explained by a lack of time to build up context codes in the 0-ms condition. The effect for between-category trials is surprising assuming that discrimination was based on label codes. However, because the vowel stimuli were drawn from a synthetic continuum with little acoustic difference between the steps, the participants might have relied on context codes in addition to label codes in between-category discrimination. Also, listeners may depend more on context codes for synthetic than natural stimuli (see Hessen and Schouten, 1992).

<sup>4</sup>The mean *F2* of the [u] tokens in experiment 2 was 1545 Hz, which is somewhat higher (indicating fronting) than previously published *F2* values for American English [u], e.g., 1238 Hz in a /bVt/ context (Bradlow, 1995) or 870 Hz in an /hVd/ context (Peterson and Barney, 1952). Possibly, the preceding [l] had a fronting effect on [u].

<sup>5</sup>Another possibility is that the S-monolinguals' scores for the easy contrast were reduced somewhat by anxiety because they were unsure about the correct response in the majority of trials (the "difficult" trials). Whatever the explanation, it is worth noting that in previous cross-language research, similar levels of AXB scores have been obtained for instances of a clear "two-category" consonant contrast (Best *et al.*, 2001, p. 787).

<sup>6</sup>Inspection of the classification results revealed that this S-monolingual participant—unlike the other S-monolinguals—consistently classified English [a] and [A] as different Spanish vowels (*viz.* [a] and [e]), which likely explains his high discrimination scores for this English contrast.

Best, C. T., Faber, A., and Levitt, A. G. (1996). "Perceptual assimilation of non-native vowel contrasts to the American English vowel system," *J. Acoust. Soc. Am.* **99**, 2602.

- Best, C. T., McRoberts, G. W., and Goodell, E. (2001). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," *J. Acoust. Soc. Am.* **109**, 775–794.
- Bosch, L., Costa, A., and Sebastián-Gallés, N. (2000). "First and second language vowel perception in early bilinguals," *Eur. J. Cogn. Psychol.* **12**, 189–221.
- Bradlow, A. R. (1995). "A comparative acoustic study of English and Spanish vowels," *J. Acoust. Soc. Am.* **97**, 1916–1924.
- Cowan, N., and Morse, P. A. (1986). "The use of auditory and phonetic memory in vowel discrimination," *J. Acoust. Soc. Am.* **79**, 500–507.
- Dupoux, E., Peperkamp, S., and Sebastián-Gallés, N. (2001). "A robust method to study stress 'deafness'," *J. Acoust. Soc. Am.* **110**, 1606–1618.
- Flege, J. E. (1995). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York Press, Timonium, MD), pp. 233–277.
- Flege, J. E. (2003). "Assessing constraints on second-language segmental production and perception," in *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and N. O. Schiller (Mouton de Gruyter, Berlin), pp. 319–355.
- Flege, J. E., and MacKay, I. R. A. (2004). "Perceiving vowels in a second language," *Stud. Second Lang. Acquis.* **26**, 1–34.
- Flege, J. E., MacKay, I. R. A., and Meador, D. (1999). "Native Italian speakers' perception and production of English vowels," *J. Acoust. Soc. Am.* **106**, 2973–2987.
- Flege, J. E., Munro, M. J., and Fox, R. A. (1994). "Auditory and categorical effects on cross-language vowel perception," *J. Acoust. Soc. Am.* **95**, 3623–3641.
- Fujisaki, H., and Kawashima, T. (1970). "Some experiments on speech perception and a model for the perceptual mechanism," *Annu. Rep. Eng. Rese. Inst. (Tokyo)* **29**, 207–214.
- Grønnum, N. (1998). *Fonetik og Fonologi—Almen og Dansk (Phonetics and Phonology—General and Danish)* (Akademisk Forlag, Copenhagen).
- Grøsjean, F. (1989). "Neurolinguists, Beware! The bilingual is not two monolinguals in one person," *Brain Lang.* **36**, 3–15.
- Harnsberger, J. D. (2001). "On the relationship between identification and discrimination of non-native nasal consonants," *J. Acoust. Soc. Am.* **110**, 489–503.
- Hessen, A. v., and Schouten, M. E. (1992). "Modeling phoneme perception: II. A model of stop consonant discrimination," *J. Acoust. Soc. Am.* **92**, 1856–1868.
- Iverson, P., Kuhl, P. K., Yamada, R. A., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). "A perceptual interference account of acquisition difficulties for non-native phonemes," *Cognition* **87**, B47–B57.
- Kluender, K. R., Lotto, A. J., Holt, L. L., and Bloedel, S. L. (1998). "Role of experience for language-specific functional mappings of vowel sounds," *J. Acoust. Soc. Am.* **104**, 3568–3582.
- Kuhl, P. K. (2000). "A new view of language acquisition," *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11850–11857.
- Kuhl, P. K., and Iverson, P. (1995). "Linguistic experience and the 'Perceptual Magnet Effect'," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 121–154.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). "Linguistic experience alters phonetic perception in infants by 6 months of age," *Science* **255**, 606–608.
- Long, M. H. (1990). "Maturational constraints on language development," *Stud. Second Lang. Acquis.* **12**, 251–285.
- Macmillan, N. A., Goldberg, R. F., and Braida, L. D. (1988). "Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua," *J. Acoust. Soc. Am.* **84**, 1262–1280.
- Massaro, D. W. (1970). "Perceptual auditory images," *J. Exp. Psychol.* **85**, 411–417.
- McClelland, J. L. (2001). "Failures to learn and their remediation: A Hebbian account," in *Mechanisms of Cognitive Development: Behavioral and Neural Approaches*, edited by J. L. McClelland and R. S. Siegler (Erlbaum, Mahwah, NJ), pp. 97–211.
- Pallier, C., Bosch, L., and Sebastián-Gallés, N. (1997). "A limit on behavioral plasticity in speech perception," *Cognition* **64**, B9–B17.
- Pallier, C., Colomé, A., and Sebastián-Gallés, N. (2001). "The influence of native-language phonology on lexical access: Exemplar-based *vs.* abstract lexical entries," *Psychol. Sci.* **12**, 445–449.
- Pallier, C., Dahan, S., Poline, J. B., LeBihan, D., Argenti, A. M., Dupoux, E., and Mehler, J. (2003). "Brain imaging of language plasticity in adopted

- adults: Can a second language replace the first?" *Cereb. Cortex* **13**, 155–161.
- Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 585–594.
- Polka, L. (1995). "Linguistic influences in adult perception of non-native vowel contrasts," *J. Acoust. Soc. Am.* **97**, 1286–1296.
- Polka, L., and Bohn, O.-S. (1996). "A cross-language comparison of vowel perception in English-learning and German-learning infants," *J. Acoust. Soc. Am.* **100**, 577–592.
- Scovel, T. (2000). "A critical review of the critical period research," *Annu. Rev. Appl. Linguist.* **20**, 213–223.
- Sebastián-Gallés, N., and Soto-Faraco, S. (1999). "Online processing of native and non-native phonemic contrasts in early bilinguals," *Cognition* **72**, 111–123.
- Sebastián-Gallés, N., Echeverría, S., and Bosch, L. (2005). "The influence of initial exposure on lexical representation: comparing early and simultaneous bilinguals," *J. Mem. Lang.* **52**, 240–255.
- Stager, C. L., and Werker, J. F. (1997). "Infants listen for more phonetic detail in speech perception than in word-learning tasks," *Nature (London)* **388**, 381–382.
- Steinlen, A. (2005). *The Influence of Consonants on Native and Non-native Vowel Production*, Language Development (Gunter Narr Verlag, Tübingen).
- Ventureyra, V. A., Pallier, C., and Yoo, H.-Y. (2004). "The loss of first language phonetic perception in adopted Koreans," *J. Neurolinguist.* **17**, 79–91.
- Watson, C. S. (2005). "Some comments on informational masking," *Acta. Acust. Acust.* **91**, 502–512.
- Werker, J. F., and Tees, R. C. (1984). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," *Infant Behav. Dev.* **7**, 49–63.
- Yamada, R. A. (1995). "Age and acquisition of second language speech sounds. Perception of American English /t/ and /l/ by native speakers of Japanese," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 305–320.