Native Italian speakers' perception and production of English vowels

James Emil Flege[a)]
Department of Rehabilitation Sciences, University of Alabama at Birmingham, Birmingham, Alabama 35294

Ian R. A. MacKay
Department of Linguistics, University of Ottawa, Ottawa, Ontario K1S 5JS, Canada

Diane Meador
Speech and Language Sciences Research Laboratory, Juneau, Alaska

This study examined the production and perception of English vowels by highly experienced native Italian speakers of English. The subjects were selected on the basis of the age at which they arrived in Canada and began to learn English, and how much they continued to use Italian. Vowel production accuracy was assessed through an intelligibility test in which native English-speaking listeners attempted to identify vowels spoken by the native Italian subjects. Vowel perception was assessed using a categorial discrimination test. The later in life the native Italian subjects began to learn English, the less accurately they produced and perceived English vowels. Neither of two groups of early Italian/English bilinguals differed significantly from native speakers of English either for production or perception. This finding is consistent with the hypothesis of the speech learning model [Flege, in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues* (York, Timonium, MD, 1995)] that early bilinguals establish new categories for vowels found in the second language (L2). The significant correlation observed to exist between the measures of L2 vowel production and perception is consistent with another hypothesis of the speech learning model, viz., that the accuracy with which L2 vowels are produced is limited by how accurately they are perceived. © *1999 Acoustical Society of America.* [S0001-4966(99)04711-6]

PACS numbers: 43.71.Hw [JMH]

INTRODUCTION

This study examined the production and perception of English vowels by groups of native Italian speakers who differed according to their age of arrival (AOA) in Canada and amount of self-reported continued use of Italian. The study addressed three questions. The first was whether the subjects' accuracy producing and perceiving English vowels would diminish as AOA increased. The second was whether the subjects who began to learn English as young children (''early'' bilinguals) would perform more like subjects in a native English comparison group than would the subjects who began to learn English as young adults (''late'' bilinguals). The final question was whether differences in amount of native language (L1) use would affect the native Italian subjects' production or perception of English vowels.

A. Previous research

Research has shown that late bilinguals often produce vowels in a second language (L2) differently than do monolingual native speakers (e.g., Major, 1987; Flege, 1992a, b; Busà, 1992, 1995; Munro, 1993; Jun and Cowie, 1994; Flege *et al.*, 1997a). Early bilinguals may produce L2 vowels more accurately than late bilinguals, although they may not produce the full range of L2 vowels exactly like monolingual native speakers.

In a study by Flege (1992a), native English-speaking listeners attempted to identify native Spanish speakers' productions of four English vowels (/i ɪ ɛ æ/). The intelligibility scores obtained for late bilinguals' vowels, but not those obtained for the early bilinguals, differed from the native English speakers' intelligibility scores. Both the early and late bilinguals produced large temporal differences between /i/–/ɪ/ and /ɛ/–/æ/. However, there was more spectral overlap between pairs of vowels produced by the late bilinguals than by the early bilinguals or the English monolinguals. In a study by Munro *et al.* (1996), native English-speaking listeners used a scale ranging from 1 (''wrong vowel or very strong foreign accent'') to 5 (''correctly produced English vowel, no foreign accent'') to rate vowels spoken by 240 native Italian subjects with AOAs ranging from 2 to 22 years. The production of an English vowel was accepted as ''nativelike'' if it received a rating that fell within two standard deviations of the mean rating obtained for native English speakers. The later the native Italian subjects began to learn English, the less likely they were to be credited with nativelike productions of English vowels.

Although early bilinguals appear to produce L2 vowels accurately overall, their production of certain vowels may differ from those of native speakers. Flege (1998) re-examined the ratings obtained by Munro *et al.* (1996) for 72 early bilinguals having AOAs of 2–8 years and 72 late bilinguals having AOAs of 16–22 years. A mean rating was calculated for each subject's production of English /i ɛ ɑ/

[a)]Electronic mail: jeflege@uab.edu

(which each differ slightly from a vowel found in Italian) and /ɪ ʊ ɚ/. The second set of English vowels might be considered ''new'' in the sense that no such vowels occur contrastively in Standard Italian. The results of the reanalysis were consistent with two hypotheses of the speech learning model, or SLM (Flege, 1995). As predicted, the late bilinguals' vowels received lower ratings (i.e., were more foreign accented) than the early bilinguals' vowels. This supported the hypothesis that the later L2 learning begins, the less likely L2 learners are to establish new categories for vowels in the L2. The early bilinguals' productions of /i ɛ ɑ/ but not /ɪ ʊ ɚ/ received significantly lower ratings than native English speakers' vowels. The presence of a detectable foreign accent in the early bilinguals' productions of /i ɛ ɑ/ was attributed to the blockage of category formation for these English vowels by the presence of phonetically similar Italian vowels. This finding was consistent with another hypothesis of the SLM, namely that the likelihood of category formation for L2 vowels is directly related to degree of perceived cross-language phonetic similarity.

Previous findings for vowel perception broadly parallel those just reported for vowel production. Research examining vowel perception has shown that late bilinguals often differ from monolingual native speakers. However, the magnitude of native versus non-native differences seems to depend on the degree of perceived similarity between L2 vowels and the closest L1 vowel (Polka, 1995; Best *et al.*, 1996; Flege *et al.*, 1997a, 1998; see also Best and Strange, 1992). The relation between AOA and the perception of L2 vowels and consonants (or ''sounds,'' for short) is as yet uncertain. MacKay *et al.* (under review) found that the frequency with which native Italian subjects erred in identifying word-initial English consonants increased as AOA increased. Yamada (1995) found that the earlier native speakers of Japanese had arrived in the United States (US), and the longer they had lived in the US, the more accurately they perceived synthetic English /ɹ/ and /l/ tokens.

To our knowledge, no previous study has assessed the effect of AOA on L2 vowel perception (see Strange, 1995). However, the results of several studies suggest that early bilinguals may not perceive all L2 vowels in a nativelike fashion. Mack (1989) examined the identification and discrimination of the members of a synthetic English /i/–/ɪ/ continuum by English monolinguals and early bilinguals who had learned both of their languages (English and French) as young children. The two groups' vowel discrimination did not differ, but the location of their phoneme boundaries did differ. This led Mack to conclude that the early bilinguals' perception of English vowels ''approximated but did not match'' that of monolinguals (p. 187).

More recently, Pallier *et al.* (1997) examined native speakers of Spanish who had begun to learn Catalan by the age of 6 years, had lived in Barcelona since birth, used both Catalan and Spanish frequently, and were highly proficient in both languages. The task was to identify and discriminate the members of a synthetic Catalan /e/–/ɛ/ continuum. (Catalan and Spanish both have an /e/, whereas Catalan but not Spanish has an /ɛ/.) The grouped identification function obtained for native speakers of Catalan revealed a clear crossover from /e/ to /ɛ/, and a discrimination peak near the phoneme boundary. However, the identification function obtained for the Spanish/Catalan bilingual group was flat, and this group's discrimination function revealed no peak. The authors concluded from these findings that ''even early exposure... is not sufficient'' for many individuals to learn ''two new phonetic categories which overlap'' a single L1 vowel category (1997, B14).

Finally, Sebastián-Gallés and Soto-Faraco (in press) used a modified version of the gating procedure to examine the identification of vowels in naturally produced disyllabic Catalan nonwords. The subjects' task was to identify the first vowel in CVC(C)V words as /e/ or /ɛ/ (or as /o/ or /ɔ/ in another condition). Spanish/Catalan bilinguals needed to hear significantly longer portions of the stimuli than did native speakers of Catalan in order to identify the Catalan vowels correctly. The authors concluded that the malleability of phonemic categories is severely limited, and that even early and extensive exposure to an L2 may be insufficient to ''overcome the influence of L1 phonemic categories in the formation of new non-native categories.''

B. The present study

The present study examined the production and perception of English vowels by four groups of Italian/English bilinguals. The subjects in three of the groups were selected on the basis of AOA. AOA averaged 7 years for the group designated ''early,'' 14 years for the group designated ''mid,'' and 19 years for subjects in the ''late'' group. The subjects in these three groups reported using Italian 31% of the time, on the average. The native Italian subjects in a fourth group, designated ''early-low,'' consisted of individuals who were matched to the subjects in the early group for AOA but reported using Italian less ($M = 8\%$).

Experiment 1 of the present study assessed the production of ten English vowels (/i ɪ eᴵ ɛ æ u ʊ o ʌ ɒ/). An intelligibility score was obtained for each vowel token by determining the percentage of times it was heard as intended by a panel of native English-speaking listeners. Research has shown that foreign accents in an L2 become progressively stronger as AOA increases (e.g., Flege *et al.*, 1995). Other research has shown that the more bilinguals continue to use their L1, the stronger their foreign accent in the L2 will be (Flege *et al.*, 1997a; Piske and MacKay, 1999; Guion *et al.*, under review). Still other research has shown that a strong correlation exists between overall degree of perceived foreign accent and the accuracy with which specific L2 vowels are produced (see Flege, 1998, for review). These findings led to two predictions concerning the vowel production results that might be obtained in experiment 1. The first was that the native Italian subjects' intelligibility scores would decrease as AOA increased. The second prediction was that the early bilinguals who used Italian often would produce English vowels less accurately than would the early bilinguals who used Italian seldom (i.e., subjects in the early-low group).

Vowel perception was assessed in experiment 2 using a categorial discrimination test. This test assessed phonetic sensitivity to contrasts between two English vowels (/i/–/ɪ/,

/u/–/ʊ/, /æ/–/ʌ/, and /ɒ/–/ʌ/), between an English vowel and an Italian vowel (/æ/–/a/, /ʌ/–/a/, /ʊ/–/o/, and /eᶦ/–/e/), and between two Italian vowels (/u/–/o/, /e/–/a/, /u/–/i/). Standard Italian contains the vowels /i e ɛ a ɔ o u/ (Agard and DiPietro, 1964). The vowel systems acquired as children by the native Italian subjects examined here were likely to vary (e.g., Trumper, 1995). We can nevertheless be sure that all of them contained fewer vowels than English. This being the case, both vowels in some of the ''English–English'' and ''English–Italian,'' contrasts were likely to have been identified as instances of one Italian category, at least at early stages of L2 acquisition.

It was more difficult to generate predictions concerning the discrimination scores than for the intelligibility scores obtained in experiment 1. For one thing, no previous study has examined the effect of AOA on vowel perception. Moreover, two existing theoretical models led to differing expectations. The spectral and temporal properties which define any L1 vowel category vary idiolectally and dialectally, and according to factors such as speaking rate and degree of stress. Thus, over time, the likelihood that typical realizations of an L2 vowel will fall within the range of tokens previously identified as being instances of an L1 category is likely to increase (Flege, 1992b). Also, as L1 categories become better defined during L1 development, they might ''assimilate'' L2 tokens more strongly (e.g., Best *et al.*, 1996; see also Kuhl, 1993).

According to the SLM, the formation of new phonetic categories (or functional equivalence classes; see, Kluender *et al.*, 1998) for L2 vowels becomes less likely with increasing age. By hypothesis, L2 vowels of a given degree of phonetic similarity to the closest L1 vowel will, with increasing age, become more likely to be heard as instances of an existing L1 category (i.e., to be functionally ''equated''). If this hypothesis is correct, then one would expect the early bilinguals examined here to discriminate L2 vowels—both from one another and from neighboring L1 vowels—more accurately than the late bilinguals. (This prediction assumes, of course, that instances of the two vowel categories being discriminated would, in early stages of exposure to the L2, be identified as instances of a single L1 category.)

The perceptual assimilation model (PAM) developed by Best and her colleagues (e.g., Best, 1995, 1999; Best *et al.*, 1996) does not predict an effect of AOA on L2 vowel discrimination. According to the PAM, the discriminability of any pair of vowels in an L2 will depend on the degree of perceived similarity of the L2 vowels to vowels in the L1. (As for the SLM, the perceived relation of phones drawn from two languages is assessed through listeners' classifications and ratings of goodness of fit to categories in the L1.) The PAM attributes ''malleability'' to the perceptual systems of adult L2 learners, and predicts that L2 vowel discriminability may improve as a function of L2 experience (Best and Strange, 1992). However, the PAM does not predict age-related differences in the perceived similarity of L2 vowels to the closest L1 vowel(s). Thus, the PAM leads to the expectation that AOA will not affect the discrimination of contrasts between two L2 vowels whose realizations are

likely to be identified as instances of a single L1 category.

The data obtained in experiments 1 and 2 are relevant to the issue of how L2 segmental production and perception are related. This topic has generated considerable discussion over the years (see, e.g., Strange, 1995; Flege, 1999a, b). According to some (e.g., Rochet, 1995), the accuracy with which L2 phonetic segments are produced is limited by how accurately they are perceived. However, Bever (1981) hypothesized that a ''critical'' period for L2 speech learning ends when the phonological system of the L1 is fully established because there is no longer a need to align speech production and perception. On this view, L2 production accuracy will not be constrained by perceptual accuracy. In support of this, native Japanese speakers of English have been observed to produce /ɹ/ and /l/ accurately, yet to differ from native speakers in perceiving English liquids (e.g., Goto, 1971; Sheldon and Strange, 1982). Were such a finding to generalize to the production and perception of English vowels by highly experienced speakers of an L2, it would undermine the SLM. According to the SLM, learners of all ages retain the capacity to align their production of L2 phonetic segments to long-term memory representations for vowels and consonants in the L2. On this view, many segmental production errors observed for experienced speakers of an L2 are likely to have a perceptual basis. This hypothesis leads to a final prediction, namely that if early bilinguals are found to produce L2 vowels accurately, then they will also perceive L2 vowels accurately.

The present article is organized as follows. The vowel production experiment will be presented first, then the vowel discrimination experiment. The relation between the intelligibility scores obtained in experiment 1 and the discrimination scores obtained in experiment 2 will then be assessed in the third section. This will be done through regression analyses, and by comparing the vowel perception accuracy of subgroups of native Italian subjects who differed in L2 vowel production accuracy.

I. GENERAL METHOD

A. Subjects

Seventy-two of the subjects were born in Italy and began to learn English when they immigrated to Canada. All but two of the native Italian subjects were from working-class backgrounds (as indicated by parental occupation). They had lived in Canada for an average of 35 years at the time of testing, the minimum being 18 years. Most of the native Italian subjects were members of, or socially connected to, a predominantly Italian Roman Catholic parish in Ottawa where testing was carried out. The 18 monolingual native English speakers (or ''native English,'' for short) who were tested were also long time residents of Ottawa. The mean age of the 90 subjects was 48 years (s.d.=6). Subjects in the five groups did not differ in age [$F(4,85)=0.05$, $p>0.10$]. All 90 subjects passed a pure tone hearing screening (defined using a 35 dB HL criterion at 500, 1000, 2000, and 4000 Hz in the best ear) prior to participating; none of them reported a history of auditory disorder.

TABLE I. Characteristics of the five groups of subjects. ''Age'' indicates mean chronological age, in years. ''AOA'' and ''LOR'' indicate the native Italian subjects' mean age of arrival and length of residence in Canada, in years, and ''%Italian'' their mean self-estimated percentage use of Italian. Standard deviations are in parenthesis.

	Gender	Age	AOA	LOR	% Italian
Native	9m,9f	48 (7)	···	···	···
Early-low	9m,9f	48 (5)	7 (3)	40 (5)	8 (6)
Early	8m,10f	47 (6)	7 (2)	40 (6)	32 (16)
Mid	8m,10f	48 (6)	14 (1)	34 (7)	20 (11)
Late	8m,10f	48 (6)	19 (1)	28 (5)	41 (23)
M	···	48 (6)	12 (6)	35 (7)	25 (19)

As summarized in Table I, three groups of native Italian subjects differed primarily according to AOA (early—7, mid—14, late—19 years). Subjects in the early-low group were matched to those in the early group for AOA, but reported using Italian less (8% vs 32%). As summarized by MacKay *et al.* (under review), subjects in the early-low group seldom used Italian except when visiting relatives. They were less likely than were subjects in the early group to report using Italian at home, at work, on the telephone, or at social gatherings. There are reasons to think that the self-reports of Italian use were valid and reliable. There was a strong inverse correlation between the native Italian subjects' self-reported use of English and Italian ($r = -0.96$, $df = 70$, $p < 0.01$). There was also a strong correlation between the Italian use self-reports given by 62 subjects in this study and those given by these same subjects when they participated in the Flege *et al.* (1995) study [$r(60) = 0.84$, $p < 0.01$]. Also, as reported by MacKay *et al.* (under review), subjects in the early-low group made more grammatical and lexical errors when speaking Italian extemporaneously than did subjects in the early group.

As expected from previous research (e.g., Flege *et al.*, 1995; Yamada, 1995), the native Italian subjects' AOAs were correlated inversely with length of residence (LOR) ($r = -0.62$, $df = 70$, $p < 0.01$), leading to a significant effect of group on LOR [$F(3,68) = 16.1$, $p < 0.01$]. Also, the earlier the native Italian subjects had arrived in Canada, the less schooling they had received in Italy. The difference between groups (early-low: 1.8 years, early: 1.9 years, mid: 6.6 years, late: 8.5 years) was significant [$F(3,68) = 46.3$, $p < 0.01$].

All but 3 of the 72 native Italian subjects came from southern Italy. As discussed later, the vowel systems of southern Italian dialects differ in terms of the number and phonetic quality of contrastive vowels. Differences in the native vowel systems might have influenced the results obtained here, either for production or perception. However, the place of birth—and therefore by extension the native dialect—of the native Italian subjects did not appear to vary systematically across the four groups.

B. Procedure

The subjects participated in several experiments in addition to those reported here during two 1-h sessions. The other experiments involved repeating meaningful English sentences; speaking extemporaneously in English and Italian; identifying English consonants presented in noise (MacKay *et al.*, under review); and repeating semantically unpredictable English sentences (Meador *et al.*, under review). Also, as reported by MacKay *et al.*, the subjects repeated nonwords that were formed by splicing together 2–5 Italian CV syllables. This was done to assess phonological short-term memory. The dependent measure was the number of nonwords that were repeated correctly.

II. EXPERIMENT 1

The purpose of this experiment was to evaluate the native Italian subjects' productions of ten English vowels, /i ɪ eᴵ ɛ æ ɔ ʌ ɒ ʊ u/. This was done by determining the percentage of times that native English-speaking listeners heard the target vowels as intended.

A least some of the English vowels were likely to differ from vowels found in the native dialect of the native Italian subjects. Standard Italian is usually described as having seven vowels, /i e ɛ a ɔ o u/ (e.g., Agard and DiPietro, 1964). However, the number and nature of vowels in southern Italian dialects differ. For example, Romito and Trumper (1989) described a southern Italian dialect having eight vowels (/i y e ɛ æ ʌ ɔ u/) and another with five vowels (/i ɛ a o u/). Trumper (1995) observed that /a/ can be implemented with variants ranging from [æ] to [ɑ] in southern Italy. Further, southern Italians living in North America may speak a ''central-southern standard'' form of Italian in addition to their native dialect (Milani, 1996, p. 480). Although we cannot be certain what vowels the native Italian subjects' possessed when they arrived in Canada, we can be confident that considerable phonetic learning would be required for them to produce all 10 of the English vowels accurately.

A. Method

The technique used here to elicit the production of English words was similar to that employed by Bradlow *et al.* (1997). The subjects were provided with both visual and auditory prompts. The visual prompts consisted of a written list of four-word sequences, each of which contained one of the 10 target vowels of interest (e.g., *read, deed, heed, bead* for the vowel /i/). The auditory prompts, which were presented via a loudspeaker, consisted of digitized versions of the four-word sequences that had been spoken by an adult male native speaker of English. The sequences were presented in the same order in which they appeared on the written list. The subjects were told to say the four words in each sequence after hearing the entire sequence. Auditory prompts were used in addition to orthography to reduce the likelihood of spelling pronunciations (e.g., production of the /ɪ/ in ''bid'' as [i] because ''i'' in written Italian is pronounced [i]). However, the availability of native-produced models might conceivably have led to an overestimation of the native Italian subjects' accuracy in producing certain English vowels.

The subjects' production of the English words was recorded using a head-mounted microphone (Shure SM10A) and a portable DAT tape recorder. The final word in each sequence (viz., *bead, bid, bade, bed, bad, booed, bode, bud, god, hood*) was digitized at 22.05 kHz, then normalized for peak intensity. Not all of the words began with /b/, and some of the words were likely to differ in familiarity to the listeners who later evaluated them (compare, e.g., *bed* to *bade*). Either or both factors might have led to biases. Therefore, the 900 words (90 subjects × 10) were edited to obscure the identity of the initial and final consonants (and, thus, lexical identity) while leaving intact cues to vowel identity insofar as possible. All portions of the waveform following the complete constriction of the final /d/ (defined on the basis of changes in waveform amplitude and/or shape) were removed. Any prevoicing present in words beginning with a stop consonant (/b/ or /g/) was removed, as was the /h/ in *hood*. A weighting function was then applied to an interval at the beginning of the remaining signal.[1]

The stimuli prepared in this way were presented for identification to six young adult native speakers of English who had been born and raised in the Ottawa, Ontario region. Some of the listeners knew French, but none of them was proficient in that or any language other than English. All passed a pure-tone hearing screening before participating. The listeners were tested one at a time at the Phonetics Laboratory of the University of Ottawa. The stimuli were presented via a loudspeaker at a self-selected comfortable level using a notebook computer. The interval between each response and presentation of the next stimulus was fixed at 1.0 s.

The listeners used keywords to identify the vowel in each stimulus. The vowels were randomly presented in two counterbalanced sets to restrict the number of keywords needed to represent the English vowels that were likely to be heard. The keywords offered for /i ɪ eᴵ ɛ æ/ were *heed, hid, hayed, head, had, hot*, and *hut*. The keywords offered for /u ʊ o ʌ ɒ/ were *who'd, hood, hoed, hut, hot, head*, and *had*. The listeners were told to identify each vowel by clicking one of seven keywords shown on the screen. Pilot work had been undertaken to identify the vowels most likely to be heard. However, given the possibility that listeners might hear some other English vowel, they were given a list of additional keywords that represented all of the English vowels not represented by a keyword shown on the screen. A supplementary keyword was used just 17 times.[2]

B. Results

Vowels spoken by the native English speakers were identified as intended at a higher average rate ($M = 94\%$ correct) than were vowels spoken by subjects in the early group ($M = 92\%$), mid group ($M = 87\%$), or late group ($M = 67\%$). Averaged across these four groups, the rates at which individual vowels were correctly identified ranged from 83% (for /ɪ ɛ ɒ/) to 94% (for /i/).

The vowels heard instead of the target vowels were tabulated. When vowels spoken by native speakers of Italian were not heard as intended, they were most frequently heard as a vowel that was slightly lower in vowel space than the target vowel (/eᴵ/-for-/i/, /ɛ/-for-/ɪ/, /ɛ/-for-/eᴵ/, /æ/-for-/ɛ/, /ɒ/-for-/ʌ/, /ɒ/-for-/o/, /o/-for-/ʊ/, and /ʊ/-for-/u/). Exceptions to this general rule were the two low target vowels (/æ/, /ɒ/) which were most frequently heard as slightly higher vowel (/ɛ/-for-/æ/ and /ʌ/-for-/ɒ/).

A total of 720 percent correct scores (72 subjects × 10 vowels) were computed. Each score was based on the identification responses of six native English listeners. Following arcsine transformation (Kirk, 1968), the intelligibility scores were submitted to a (4) group by (10) vowel ANOVA. The ANOVA yielded significant main effects of group [$F(3,68) = 23.9$, $p < 0.01$] and vowel [$F(9,612) = 9.6$, $p < 0.01$] and a significant two-way interaction [$F(27,612) = 1.7$, $p = 0.018$].

As summarized in Table II, the simple effect of group was significant for all vowels except /eᴵ/, /ɛ/, and /ɒ/. Turkey's tests were used to determine which pairwise differences between groups were significant for the seven vowels showing a significant group effect. Six vowels spoken by subjects in the late group (viz., /i ɪ u ʊ o ʌ/) received intelligibility scores that were significantly lower than the native English groups' scores, and one vowel spoken by subjects in the mid group (/ʌ/) received significantly lower scores ($p < 0.05$). No vowel spoken by the early group received significantly lower scores than those of the native English speakers ($p > 0.10$).

The 540 scores obtained for the native English speakers and the two groups of early bilinguals were submitted to a 3 (group) × 10 (vowel) ANOVA. The aims of this analysis were to determine if the early bilinguals who seldom spoke Italian (early-low) produced English vowels more accurately than did the early bilinguals who spoke Italian relatively often (early), and if either group differed from the native English group. The main effect of vowel was significant [$F(9,459) = 8.7$, $p < 0.01$], but not the main effect of group [$F(2,51) = 1.6$, $p > 0.10$] or the two-way interaction [$F(18,459) = 1.2$, $p > 0.10$].

C. Discussion

In agreement with Munro *et al.* (1996), the later the native Italian speakers arrived in Canada, the less accurately they produced English vowels. Six vowels produced by subjects in the late group (/i ɪ u ʊ o ʌ/) received significantly lower intelligibility scores than did the native English subjects' productions of these vowels. (A seventh vowel, /æ/, received scores that were lower than those obtained for vowels spoken by subjects in the early and mid groups.) Just one vowel spoken by subjects in the mid group was significantly less intelligible than the native English speakers' vowels. However, none of the English vowels spoken by either group of early bilinguals received a significantly lower intelligibility score than did vowels spoken by the native English group. In agreement with the results of Flege (1992b), this suggests that early bilinguals who are highly experienced in their L2 may produce L2 vowels in a nativelike fashion.

III. EXPERIMENT 2

As discussed in the Introduction, the results of several recent studies suggested that early bilinguals may perceive

TABLE II. The mean intelligibility scores obtained for ten English vowels spoken by a group of native English ("NE") subjects and three group of native Italian subjects differing in age of arrival to Canada ("early," "mid," "late"). Standard deviations are in parentheses. Superscript a and b indicate significance at the 0.01 and 0.001 levels; respectively. An 0.05 level was used for the Tukey's *post hoc* tests.

	Group				$F(3,68)$	Tukey's test
	NE	Early	Mid	Late		
/i/	99 (4)	95 (13)	99 (4)	84 (28)	4.03[a]	late<NE, mid
/ɪ/	95 (16)	97 (6)	81 (34)	57 (41)	9.26[b]	late<NE, early, mid
/eᴵ/	95 (20)	99 (4)	89 (24)	81 (36)	2.26	
/ɛ/	91 (19)	86 (20)	83 (29)	70 (32)	2.46	
/æ/	84 (31)	89 (22)	95 (14)	62 (40)	4.29[a]	late<early, mid
/ɒ/	94 (10)	83 (32)	86 (29)	72 (32)	1.75	
/ʌ/	86 (12)	80 (23)	48 (41)	48 (44)	6.39[b]	mid, late<NE, early
/o/	99 (4)	100 (0)	95 (11)	72 (41)	6.55[b]	late<NE, early, mid
/ʊ/	97 (6)	94 (10)	97 (6)	56 (40)	14.51[b]	late<NE, early, mid
/u/	94 (13)	98 (5)	91 (17)	68 (36)	7.37[b]	late<NE, early, mid

L2 vowels differently than monolingual native speakers do (Mack, 1989; Pallier *et al.*, 1997; Sebastián-Gallés and Soto-Faraco, in press). However, when taken together with the production results obtained in experiment 1, the speech learning model (Flege, 1995) predicts that the early bilinguals examined here would perceive English vowels in a nativelike fashion. This was tested by assessing the native Italian subjects' phonetic sensitivity to differences between pairs of English and/or Italian vowels.

Four of the 11 vowel contrasts examined in this experiment involved two English vowels (/æ/–/ʌ/, /ʌ/–/ɒ/, /i/–/ɪ/, /u/–/ʊ/). Four others involved one English and one Italian vowel (/æ/–/a/, /ʌ/–/a/, /ʊ/–/o/, /eᴵ/–/e/), and the remaining three involved two Italian vowels (/u/–/o/, /e/–/a/, /u/–/i/). Five different individuals with different-sounding voices produced tokens of each vowel category. The average frequency of the first and second formants ($F1$,$F2$) at the acoustic midpoint of the five tokens of each category are shown in Fig. 1.[3]

The vowel stimuli were examined in a categorial discrimination test using a modified oddity format. The three stimuli presented on each trial were always spoken by different talkers, and so were physically different. Half of the trials (called "different" trials) contained an odd item out, and the remaining half (called "catch" trials) contained three instances of a single category.

The subjects were offered four response alternatives rather than the two alternatives offered in ABX and AXB tests (e.g., Gottfried, 1984; Best *et al.*, 1996). They were told to identify the serial position of an odd item out, if they judged one vowel to different from the other two, by clicking a button marked "1," "2," or "3." They were told to click a button marked "none" if they heard three different examples of one vowel (i.e., for the catch trials). The presence

of both different trials (where one vowel differed from the remaining two) and catch trials permitted the calculation of an unbiased measure of sensitivity (A') to the phonetic difference between the vowel contrasts examined here. To obtain a high score, the subjects had to respond to relevant phonetic differences while ignoring auditorily accessible differences (e.g., in voice quality) that were not phonetically relevant.

As mentioned earlier, the vowel systems of southern Italian dialects may vary considerably. For example, Trumper (1995) observed that /a/ can be implemented with variants ranging from [æ] to [ɑ] in southern Italy.
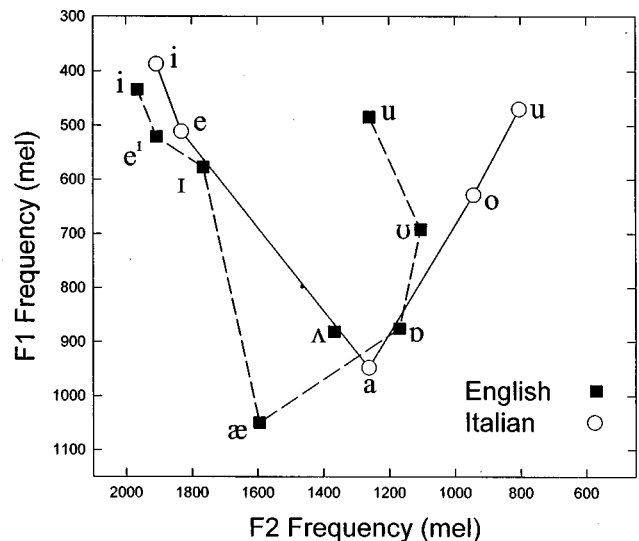

FIG. 1. The mean frequency of the first and second formants ($F1$,$F2$) in five naturally produced tokens each of English and Italian vowels, measured at the acoustic midpoint. All of the vowels shown except Italian /i/ were used as stimuli in a vowel discrimination experiment.

Given this, and the lack of empirical data pertaining to the perceived similarity of vowels in southern Italian dialects and Canadian English, it was not possible to make *a priori* predictions concerning the relative degree of discriminability of the 11 vowel contrasts examined here. However, several expectations could be derived from the acoustic values presented in Fig. 1. Among these was that the native Italian subjects would tend to identify the English /i/ and /eᶦ/ stimuli as instances of Italian /i/ and /e/, but they might not identify the fronted English /u/ stimuli as instances of Italian /u/. Other expectations were that the native Italian subjects would tend to identify the English /ɒ/ and /ʌ/ stimuli as instances of Italian /a/, and thus have difficulty with the /ʌ/–/ɒ/ contrast.

Cross-language comparisons of English and Italian (Agard and DiPietro, 1964) led us to expect that at least some native Italian subjects would have difficulty discriminating the English–English pairs /i/–/ɪ/ and /u/–/ʊ/ owing to the absence of lax vowels in Italian. The observation by Munro *et al.* (1996) that certain Italian subjects' productions of English /ʊ/ were heard as /o/, plus the acoustic data in Fig. 1, led us to expect difficulty for the English /ʊ/–Italian /o/ contrast. However, whether any of these predictions would actually be fulfilled was likely to depend on the nature of vowels in the subjects' native dialect of Italian, and also how (or if) the perceived relation between L1 and L2 vowels changes over time.

We had some general expectations concerning the three subsets of vowel contrasts examined here. We expected lower discrimination scores for the E–I (English versus Italian) contrasts than for the E–E (English versus English) contrasts. This was based on the assumption that the native Italian subjects would be more likely to identify both members of E–I contrasts as being instances of a single L1 category than to do so for both members of E–E contrasts.

We were less certain, however, how the native Italian subjects would respond to the I–I (Italian versus Italian) contrasts. The native Italian subjects might obtain near-perfect discrimination scores for the I–I contrasts because they involved two different L1 vowels. However, the SLM (Flege, 1995) leads to a different expectation, namely, that some native Italian subjects—especially those in the late group—might not obtain high scores for the I–I contrasts. According to the SLM, the subjects in the late group should be less likely than the early group subjects to form categories for English vowels. The SLM posits that when a category is not established for an L2 vowel, then the L1 category used to process tokens of the L2 vowel will change over time as the result of a process called ''assimilation.'' For example, if a category was not established for English /ʌ/, and if tokens of this vowel were judged to be instances of Italian /a/, then the extent of the Italian /a/ category might broaden so as to encompass instances of both English /ʌ/ and Italian /a/. If the late groups' Italian vowels were affected in this way more often than the early groups' Italian vowels, then the late groups' scores for the I–I contrasts might be lower than the early groups' scores.

A. Method

1. Stimuli

The English vowel stimuli were derived from words spoken by five female native speakers of English from Ottawa in a previous study (Munro *et al.*, 1996). The words (*bait, boot, back, but, bock, beat, bit, book*) were spoken at the end of a carrier phrase (*Now I say_*). A native speaker of Canadian English identified as intended all of the English vowels used as stimuli.

The Italian stimuli were derived from nonwords spoken by five female speakers of Italian living in Padua, Italy, who could be considered speakers of Standard Italian. Vowels spoken by these speakers were used as stimuli, rather than speakers of a southern Italian dialect, due to the lack of a cogent rationale for choosing one southern Italian dialect over another. The Italian nonword stimuli, which were formed by inserting /i e a o u/ into a /b_do/ frame, were produced immediately following the auditory presentation of four real words containing the vowel of interest (e.g., ''*rido, fido, lido, nido*'' for /i/). The identifiability of the Italian vowels could not be verified in a relevant fashion, as was done for the English stimuli. This is because the Italian vowels were representative of a northern Italian dialect, whereas the subjects tested here spoke varieties of southern Italian.

The English words and Italian nonwords were digitized at 22.05 kHz. Prevoicing, if any, was removed from the initial /b/ tokens, as was everything following the complete constriction of the postvocalic consonants. The edited signals were then linearly ramped off over the final 30 ms of the remaining signal. The 15 Italian stimuli were longer ($M = 243$ ms) than the 40 English stimuli were ($M = 180$ ms). Bohn (1995) hypothesized that non-native listeners may respond to differences in duration if they fail to perceive a spectral difference. The temporal difference between the Italian and English stimuli was reduced to prevent subjects from discriminating E–I contrasts on the basis of duration only. This was done by deleting ten nonsuccessive glottal pulses (mean duration=49 ms) from the middle of the Italian vowel stimuli. The 55 stimuli were then normalized for peak intensity.

2. Procedure

The interstimulus interval between the three stimuli presented on each trial was 1.2 s. Each vowel contrast was tested by eight different trials. The odd item out in these trials occurred with equal frequency in all three possible serial positions. All 13 vowel categories were tested by four catch trials ($N = 52$) consisting of three physically different realizations of a single vowel category.

The stimuli were presented using a notebook computer. As mentioned, the subjects were told to click a button marked ''1,'' ''2,'' or ''3'' if they heard one vowel that differed from the other two vowels, or ''no'' if they heard three instances of a single vowel. The subjects were familiarized with the task before the experiment began through visually presented arrays of geometric figures (e.g., two circles and a square to illustrate the concept of oddity; three triangles differing in size to illustrate a configuration that
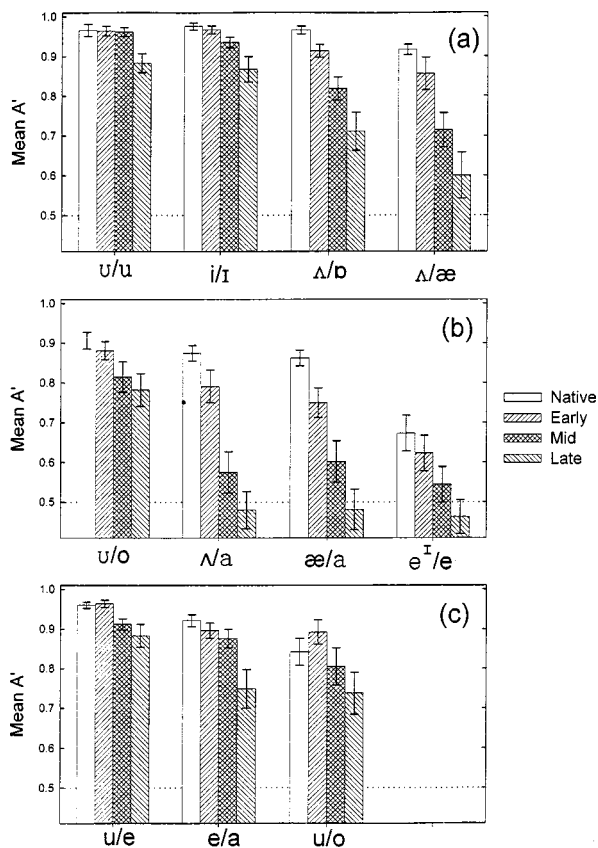
FIG. 2. The mean sensitivity (A′) of the subjects in four groups to phonetic differences between two English vowels (a), between English vowels and neighboring Italian vowels (b), or between two Italian vowels (c). The subjects were native speakers of English, and three groups of Italian/English bilinguals who differed according to the age of first extensive exposure to English (see text). The error bars bracket ∓ one SE.

required a response of ''no''). They received feedback during a practice session with I–E vowel contrasts not included in the test (/æ/–/u/, /i/–/a/). Feedback was not provided during the experiment, however.

An unbiased measure of phonetic sensitivity, A′, was calculated for each of the 11 vowel contrasts based on the proportions of hits and false alarms (see Snodgrass *et al.*, 1985).[4] There was a maximum of eight possible hits for each contrast (i.e., eight opportunities to correctly select an odd item out in different trials), and a maximum of eight possible false alarms. (For the /i/–/ɪ/ contrast, for example, there were four catch trials for /i/ and four catch trials for /ɪ/.) An A′ score of 1.0 indicated perfect sensitivity (i.e., correct responses to all eight catch trials and to all eight different trials). An A′ score of 0.5, on the other hand, indicated a lack of phonetic sensitivity.

B. Results

1. The Effect of AOA

As shown in Fig. 2(a), the native English subjects obtained higher A′ scores for all four English–English (E–E) contrasts than did the three groups of native Italian subjects differing in AOA. However, the size of native versus non-native differences was larger for certain contrasts (/ʌ/–/ɒ/, /ʌ/–/æ/) than for others (/u/–/ʊ/, /i/–/ɪ/). The native Italian

subjects' scores decreased as AOA increased. Subjects in the late group obtained lower scores than did subjects in the mid group who, in turn, obtained lower scores than did subjects in the early group.

The A′ scores obtained for the E–E vowel contrasts were submitted to a mixed-design (4) group by (4) contrast ANOVA. The significant interaction it yielded [$F(9,204) = 4.84, p < 0.01$] was explored through simple effects tests. The effect of group was found to be significant for all four vowel contrasts ($p < 0.01$). Tukey's *post hoc* tests revealed that the late group obtained lower scores for all four E–E contrasts than the native English and Early group did ($p < 0.05$). Also, the late group obtained lower scores than the mid group for /u/–/ʊ/, and the mid group obtained lower scores than the native English group did for /ʌ/–/ɒ/ and /ʌ/–/æ/ ($p < 0.05$).

The scores obtained for the English–Italian (E–I) contrasts are shown in Fig. 2(b). Once again, the native English group obtained higher scores than the early, mid, and late groups, whose scores decreased as AOA increased. The size of between-group differences was larger for /ʌ/–/a/ and /æ/–/a/ than for /ʊ/–/o/ and /eɪ/–/e/, yielding a group by vowel contrast interaction [$F(9,204) = 3.07, p < 0.01$]. The simple effect of group was significant for all four E–I contrasts ($p < 0.01$). The late group obtained lower scores than the native English group for /ʌ/–/a/, /æ/–/a/, and /eɪ/–/e/; the late group obtained lower scores than the early group for /ʌ/–/a/ and /æ/–/a/; and the mid group obtained lower scores than the native English group for /ʌ/–/a/ and /æ/–/a/ ($p < 0.05$ by Tukey's test).

Scores for the three Italian–Italian (I–I) contrasts are shown in Fig. 2(c). The average scores for the native English, early, and mid groups (means = 0.92, 0.91, and 0.86, respectively) were higher than those obtained for the late group ($M = 0.79$). An ANOVA examining the I–I scores yielded a significant main effect of group [$F(3,68) = 7.2, p < 0.01$] and a significant main effect of vowel contrast [$F(2,136) = 16.2, p < 0.01$], but not a significant two-way interaction [$F(6,136) = 0.97, p > 0.10$]. A Tukey's test examining the average scores obtained for the I–I contrasts revealed that the late group obtained significantly lower scores than the native English and early groups did ($p < 0.05$).

2. L1 use

The scores obtained for the E–E contrasts by the subjects in the native English, early-low, and early groups differed little (means = 0.96, 0.95 and 0.93, respectively). The E–E scores were submitted to a (3) group × (4) vowel contrast ANOVA. It yielded a nonsignificant main effect of group [$F(2,51) = 3.15, p = 0.051$] and a nonsignificant two-way interaction [$F(6,153) = 1.02, p > 0.10$]. The main effect of vowel contrast was significant [$F(3,153) = 16.3, p < 0.01$] because the scores for /ʌ/–/æ/ were lower than those for /i/–/ɪ/, /u/–/ʊ/, and /ʌ/–/ɒ/ ($p < 0.05$ by Tukey's test).

The mean scores obtained for the E–I contrasts by the native English, early-low, and early groups differed little (means = 0.83, 0.79, and 0.76, respectively). The ANOVA examining the E–I scores yielded a nonsignificant main effect of group [$F(2,51) = 2.52, p = 0.09$] and a nonsignificant

two-way interaction [$F(6,153) = 1.45$, $p > 0.10$]. The main effect of vowel contrast was significant [$F(3,153) = 35.7$, $p < 0.01$] because the scores for /e$^{\text{I}}$/–/e/ were lower than those for /æ/–/a/, /ʌ/–/a/, and /ʊ/–/o/, and the scores for /æ/–/a/ were lower than those for /ʊ/–/o/ ($p < 0.05$ by Tukey's test).

Finally, an ANOVA examining the I–I scores yielded a nonsignificant main effect of group [$F(2,51) = 0.12$, $p > 0.10$] and a nonsignificant two-way interaction [$F(4,102) = 1.19$, $p > 0.10$]. The main effect of vowel contrast was significant, however [$F(2,102) = 22.1$, $p < 0.01$], because the scores for /e/–/a/ and /u/–/o/ were lower than those for /u/–/e/ ($p < 0.05$).

3. Comparison across contrast types

The analyses just presented examined the scores obtained for the E–E, E–I, and I–I contrasts separately. Here, average scores for the three contrast types were submitted to a (5) group by (3) contrast type ANOVA, which yielded a significant two-way interaction [$F(8,170) = 215.0$, $p < 0.01$]. The simple main effect of group was significant for all three contrast types, and the simple main effect of contrast type was significant for each group ($p < 0.01$).

Tukey's tests revealed that the interaction had two sources. First, a different pattern of between-group differences was obtained for the three contrast types. Subjects in the late group obtained significantly lower scores for E–E contrasts than did subjects in the native English, early, early-low, and mid groups, and the mid group obtained lower scores than the native English group ($p < 0.01$). The mid and late groups obtained significantly lower E–I scores than did the native English, early-low, and early groups ($p < 0.01$). Finally, subjects in the late group, but not those in the mid group, obtained significantly lower I–I scores than did the native English, early, and early-low groups ($p < 0.01$).

A second source of the two-way interaction was a differing pattern of between-contrast-type differences for the native English and native Italian subjects. All five groups obtained significantly lower scores for the E–I contrasts than for the E–E and I–I contrasts ($p < 0.01$). However, only the native English group obtained lower scores for the I–I than E–E contrasts (0.91 vs 0.96, $p < 0.01$).

4. Results for English /e$^{\text{I}}$/ and Italian /e/

It is unlikely that new phonetic categories will be established for L2 vowels that are highly similar to an L1 vowel (Flege, 1995). The results of this study shed light on the magnitude of cross-language vowel differences needed to trigger category formation.

The English /e$^{\text{I}}$/ and Italian /e/ stimuli were located in roughly the same portion of vowel space and, after editing (see above), differed little in duration (/e/ = 168 ms, /e$^{\text{I}}$/ = 150 ms). However, the English /e$^{\text{I}}$/ tokens showed more formant movement from the beginning to the end of the vowel (*mean F*1: 466 to 375 Hz, *mean F*2: 2429 to 2642 Hz) than did the Italian /e/ tokens (*mean F*1: 424 to 404 Hz, *mean F*2: 2434 to 2525 Hz).[5] The A′ scores obtained for the /e$^{\text{I}}$/–/e/ contrast were submitted to a series of independent *t*-tests to determine if such small cross-language phonetic differences were discriminable. The aim was to determine if the A′ score obtained for any group was significantly greater than 0.50 (a score which, theoretically, represents a complete lack of sensitivity).

The scores obtained for the native English and early-low groups (*means* = 0.67 and 0.69, respectively) significantly exceeded the chance level (Bonferroni $p < 0.01$), whereas the scores obtained for the early, mid, and late groups (0.62, 0.54, 0.46) did not. The early-low group's above-chance discrimination might be interpreted to mean that some subjects established a phonetic category for /e$^{\text{I}}$/. However, such an interpretation is clearly not reasonable for the native English subjects, who did not speak Italian. Their above-chance discrimination of /e$^{\text{I}}$/–/e/ might instead be attributed to the identification of the Italian /e/ tokens as /ɛ/ (see Gottfried, 1984).

C. Discussion

The perception results obtained here parallel those obtained earlier for vowel production. The later the native Italian subjects arrived in Canada, the less accurately they perceived English vowels. Eight contrasts involving an English vowel or vowels were examined. Subjects in the late group (mean AOA = 19 years) received significantly lower discrimination scores than the native English subjects did for all eight. Subjects in the mid group (mean AOA = 14 years) discriminated four contrasts less accurately than the native English subjects did. These differences are notable in that subjects in the mid and late groups had lived in Canada for an average of 31 years. It thus appears that certain native versus non-native differences in vowel perception persist.

None of the eight contrasts involving an English vowel were discriminated less accurately by either of two early bilingual groups (mean AOAs = 7 years) than by the native English group. This finding, which suggests that early bilinguals who are highly experienced in their L2 may perceive L2 vowels in a nativelike fashion, diverges from the results of several recent studies (Mack, 1989; Pallier *et al.*, 1997; Sebastián-Gallés and Soto-Faraco, in press). The success of the early bilinguals examined here might be attributed to the establishment of long-term memory representations (or "phonetic categories") for certain L2 vowels (Flege, 1995). Neither /ɪ/ nor /ʊ/ occur systematically in Italian. The fact that subjects in the late group, but not those in the early and mid groups, differed from the native English speakers in discriminating English /i/–/ɪ/ and /u/–/ʊ/ suggests that at least some subjects in the early and mid groups established phonetic categories for /ɪ/ and /ʊ/.

Other findings obtained here suggest indirectly that subjects in the Early group may have established a phonetic category for a low and/or a mid-central English vowel. Busà (1995) found that, in speech production, Italian late learners of English tended to neutralize differences between low and mid-central English vowels. This is probably because (Standard) Italian has an /a/ but no /æ/, /ɒ/ or /ʌ/. Subjects in the mid and late groups, but not those in the early group, obtained significantly lower discrimination scores than the native English group for /æ/–/a/, /ʌ/–/a/, /ʌ/–/ɒ/ and /ʌ/–/æ/.

The results obtained for the Italian–Italian (I–I) contrasts pose an interpretive challenge. The first issue to consider is why the native English subjects obtained scores that were well above chance ($M = 0.91$). We suspect that the native English subjects identified Italian vowels in the /u/–/e/, /e/–/a/, and /u/–/o/ contrasts in terms of two different English vowels, but that the Italian vowels did not represent a good ''fit'' to their long-term memory representations (see, e.g., Best, 1995; Best *et al.*, 1996). This might explain why, of the five groups examined, only the native English group obtained significantly lower scores for the I–I than the E–E contrasts.

The second issue to consider is why the late group obtained lower scores for the I–I contrasts ($M = 0.79$) than did the native English and early groups. Perhaps shortening the Italian vowel stimuli (see Sec. I) lessened the identifiability of these stimuli, or reduced their perceived goodness of fit to vowels in the L1 system of members of the late group. Another possibility is that the Italian vowel stimuli represented a poor fit to the late group's long-term memory representations because they were produced by speakers of a northern rather than a southern variety of Italian. A problem exists with both potential explanations, however. There is no reason to think that the effect of editing, or of cross-dialect differences, should have been greater for the late group than for the early group. This is because, as mentioned in Sec. I, the place of origin in Italy of subjects in the four groups seemed to be evenly distributed.

The categorial discrimination test used here placed a heavy load on working memory. Thus, another possible explanation is that some members of the late group had a deficient working memory. This possibility can probably be ruled out, however. As reported by MacKay *et al.* (under review), the subjects were asked to repeat nonwords that were formed by splicing together 2–5 Italian CV syllables. This was done to assess phonological short-term memory. The dependent variable was the number of nonwords repeated correctly. The late group's repetition scores did not differ from the scores obtained for any other group. In fact, the late groups' scores were nonsignificantly higher than those obtained for the native English and early groups. Moreover, when the nonword repetition scores were used as covariates in ANCOVAs, the same results as those reported earlier were obtained.

As mentioned in the Introduction, still another explanation for why the late group obtained relatively low scores for the I–I contrasts is that their long-term memory representations for Italian vowels changed as the result of learning English. More specifically, some of their Italian categories may have changed by assimilating a neighboring English vowel. Assimilation of this kind is predicted to occur only in the absence of category formation for an L2 vowel (which is predicted to occur more often for late than early bilinguals; Flege, 1995). This hypothesis will need to be tested by comparing the discrimination of I–I contrasts by Italian monolinguals and by early and late bilinguals who are highly experienced in English. If the hypothesis is correct, then late but not early bilinguals should obtain lower I–I discrimination scores than Italian monolinguals. If such a finding were

TABLE III. Simple correlations between the intelligibility scores (a measure of vowel production accuracy) and the discrimination scores obtained for the 72 native Italian subjects. E-E, I-I, and E-I indicate contrasts between two English vowels, two Italian vowels, or between one English and one Italian vowel. ''L1 use'' refers to the percentage of self-reported use of Italian; ''AOA'' and ''LOR'' indicate age of arrival and length of residence in Canada, in years. Superscript a and b indicate significance at the 0.01 and 0.001 levels, respectively.

	E–E	E–I	I–I
Simple correlation	0.64[b]	0.62[b]	0.40[b]
L1 use partialled out	0.59[b]	0.57[b]	0.41[b]
AOA partialled out	0.37[b]	0.30[a]	0.17
LOR partialled out	0.48[b]	0.46[b]	0.28

obtained, it would provide a perceptual analog to the observation that production of an L1 consonant may become less accurate as the production of a corresponding L2 consonant improves (Flege, 1987).

IV. THE RELATION BETWEEN PRODUCTION AND PERCEPTION

According to the speech learning model (Flege, 1995), phonetic segments in an L2 can be produced in a nativelike fashion only if they are perceived in a nativelike fashion. The results presented in Sec. II indicated that the early Italian/English bilinguals produced English vowels in a nativelike fashion. This finding thus led to the prediction that, contrary to the findings of several vowel perception studies (Mack, 1989; Pallier *et al.*, 1997; Sebastian-Gallés and Soto-Faraco, in press), the early bilinguals examined here would also perceive English vowels accurately. This prediction was confirmed in Sec. III. The aim of the analyses presented here was to further explore the relation between L2 vowel production and perception.

The first specific question addressed here was whether a significant correlation existed between the intelligibility scores assessing the 72 native Italian subjects' English vowel production accuracy and the average discrimination scores they obtained for English–English (E–E), English–Italian (E–I), and Italian–Italian (I–I) contrasts. As shown in Table III, there was a significant correlation between the intelligibility scores and all three sets of vowel discrimination scores. However, the correlation between the intelligibility scores and the E–E discrimination scores was significantly stronger than was the correlation between the intelligibility scores and the I–I discrimination scores [$X(1) = 11.8$, $p < 0.01$]. There was also a stronger correlation between the intelligibility scores and the E–I discrimination scores than between the intelligibility and I–I discrimination scores [$X(1) = 8.5$, $p < 0.01$]. This last finding suggests that degree of L2 vowel production accuracy is related more closely to how L2 vowels are perceived than to how vowels in the L1 are perceived.

Partial correlations were also computed. As shown in Table III, the correlations between the intelligibility scores, on the one hand, and the E–E and E–I discrimination scores, on the other hand, remained significant when variation in L1 use, AOA, and length of residence in Canada were partialled out. However, the correlations between the intelligibility

TABLE IV. The discrimination (A′) scores obtained for three contrasts involving the English vowel /ʌ/ for the native English subjects ($n = 18$) and for native Italian subjects who produced /ʌ/ relatively well ($n = 41$) or poorly ($n = 29$; see text). Standard deviations are in parentheses.

Vowel contrast	Native English	Good producers	Poor producers
/ʌ/ vs English /æ/	0.92 (0.06)	0.80 (0.20)	0.73 (0.23)
/ʌ/ vs Italian /a/	0.87 (0.08)	0.73 (0.21)	0.57 (0.24)
/ʌ/ vs English /ɒ/	0.97 (0.04)	0.87 (0.15)	0.82 (0.16)

scores and the I–I discrimination scores became nonsignificant when AOA and length of residence were partialled out.

A multiple regression analysis was able to account for just 57% of the variance in the intelligibility scores ($p < 0.01$). AOA accounted for 48% of the variance at step 1, the E–E discrimination scores accounted for 5% of the variance at step 2, and years of residence in Canada accounted for 4% of the variance at step 3. It is possible that an imprecise measurement of L2 vowel production accuracy was responsible for the small amount of variance accounted for. More likely, perhaps, ceiling effects limited how much variance was accounted for. Of the 720 vowel tokens spoken by the native Italian subjects, 81% were heard as intended by at least five of the six listeners. The vowel most often misidentified was /ʌ/. Five or more listeners correctly identified the /ʌ/ productions of 41 native Italian subjects, whereas four or fewer listeners identified the /ʌ/ productions of the remaining 31 (10 each in the mid and late groups, 6 in the early group, and 5 in the early-low group).

If L2 vowel production is limited by L2 vowel perception accuracy, then the 41 subjects who produced /ʌ/ relatively well should also have perceived /ʌ/ relatively well. Difficulty perceiving /ʌ/ would most likely manifest itself in the discrimination of English /ʌ/ from Italian /a/, for this is likely to be the Italian vowel that is perceptually closest to English /ʌ/. (Unfortunately, no objective assessment of this inference exists at present.)

As summarized in Table IV, a series of ANOVAs examined the discrimination scores obtained for the native Italian subjects who produced /ʌ/ poorly ($N = 31$), those who produced /ʌ/ relatively well ($N = 41$), and the native English speakers ($N = 18$). The effect of group on the /ʌ/–/a/ discrimination scores was significant [$F(2,87) = 12.0, p < 0.01$]. A Tukey's test revealed that the native English group obtained significantly higher /ʌ/–/a/ discrimination scores than did the poor but not the good producers of /ʌ/. Also, the good producers obtained significantly higher /ʌ/–/a/ discrimination scores than did the poor producers of /ʌ/ ($p < 0.01$). The effect of group was also significant for the /ʌ/–/æ/ and /ʌ/–/ɒ/ contrasts [$F(2,87) = 5.5$ and $6.2, p < 0.01$], but the good and poor producers' discrimination of these two English–English contrasts did not differ significantly.

V. GENERAL DISCUSSION

One experiment reported here examined the accuracy with which native speakers of Italian produced English vow-

els. The dependent variable was the percentage of correct identifications by native English-speaking listeners of the target English vowels. The later the native Italian subjects arrived in Canada, the less often their English vowels were heard as intended. The intelligibility scores obtained for the late group (mean age of arrival in Canada = 19 years) were significantly lower than were the intelligibility scores obtained for a group of native English speakers for six of the 10 vowels examined (/i ɪ u ʊ o ʌ/). The intelligibility scores obtained for the mid group (mean AOA = 14 years) were lower than the native English groups' scores for just one vowel (/ʌ/). Two groups of early Italian/English bilinguals matched for AOA ($M = 7$ years) but differing in L1 use (early-low = 8%, early = 32%) were also examined. No vowel produced by either of these groups was significantly less intelligible than vowels spoken by the native English group. Nor, contrary to prediction (see the Introduction), did the two groups of early bilinguals differ significantly from one another.

Parallel perception results were obtained in a second experiment. A categorial discrimination test was used to assess the native Italian subjects' phonetic sensitivity to differences between four pairs of English vowels, and four English versus Italian vowel contrasts. Subjects in the late group received significantly lower discrimination scores for all eight vowel contrasts than did the native English subjects. Subjects in the mid group received significantly lower scores than did the native English subjects for four vowel contrasts. However, neither group of early bilinguals differed significantly from the native English group for any vowel contrast.

Not surprisingly, the vowel intelligibility scores and the discrimination scores obtained for the 72 Italian/English bilinguals were correlated, even when effects of the variables used to select them (age of arrival in Canada, percent use of Italian) were partialled out. These results are consistent with, but do not in themselves prove, the claim that accuracy in producing phonetic segments in an L2 is limited by how accurately the phonetic segments are perceived (e.g., Flege, 1999a,b).

Bradlow *et al.* (1997) also obtained evidence of a link between production and perception. These authors administered identification training to native Japanese adults using naturally produced English /ɹ/ and /l/ tokens. The subjects' perception improved as the result of the training and so too did their production of English liquids in the absence of any explicit production training. However, the size of gains in perception and production were uncorrelated, and one subject who showed a perceptual gain showed little evidence of improved production. This led Bradlow *et al.* (1997) to conclude that perceptual learning is not a ''necessary or sufficient condition'' for improved production (p. 2307), which may lag behind perceptual changes (see Flege, 1995, for a similar suggestion). If some perceptual changes are never ''transported'' to production (Bever, 1981), this might explain the lack of perfect correlation between production and perception observed in the present study. Another possible explanation for the modest production–perception correlations observed here was that errors were made in measuring

the subjects' vowel production accuracy, their vowel perception accuracy, or both (see Flege, 1999b).

The AOA effect on speech production observed here replicates the findings of Munro *et al.* (1996) for Italian/English bilinguals living in Ottawa. However, this is apparently the first time that an effect of AOA has been observed for the perception of vowels in an L2 (see Strange, 1995, for discussion). The effects of AOA on both vowel production and perception are notable because the native Italian subjects had been living in Canada for an average of 36 years (range: 18–48 years) when tested, and most of them used English more than Italian. Given that AOA indexed the age at which the native Italian subjects were first exposed extensively to English, these findings might be taken as support for the existence of a critical period for L2 speech learning (e.g., Scovel, 1988, 1969; Patkowski, 1989). However, an alternate account of these findings can be derived from the SLM (Flege, 1995, 1998). According to the SLM, age-related effects on L2 speech production and perception arise primarily from changes in how the L1 and L2 phonetic systems interact, not from a neurologically triggered loss in the ability to learn speech. The nature of L1–L2 interactions, according to the SLM, varies as a function of the state of development of the L1 phonetic system when L2 learning begins.

The lack of a vowel production difference between the early bilinguals and the native English speakers diverges from a previous finding. As mentioned in the Introduction, Flege (1998) found that mean ratings obtained for early Italian/English bilinguals' productions of English /i ɛ ɑ/, but not their productions of English /ɪ ʊ ɚ/, were significantly lower than the ratings obtained for native English speakers' productions of the same vowels. The difference between the early bilinguals' productions of /ɪ ʊ ɚ/ and /i ɛ ɑ/ was attributed to the formation of phonetic categories by the early bilinguals for /ɪ ʊ ɚ/ (which lack phonetic counterparts in Italian), but the lack of category formation for /i ɛ ɑ/ due to the presence of phonetically similar vowels in Italian. The intelligibility scores obtained here were sufficiently sensitive to reveal differences between the native English group and the mid and late groups. Still, the Flege (1998) results suggest the possibility that the early bilinguals might have been found to differ from the native English speakers, at least for certain English vowels, had a more sensitive technique been used to assess their production of English vowels.

The lack of a vowel perception difference between the early bilinguals and the native English speakers also appears to diverge from previous findings. As mentioned in the Introduction, several studies identified differences between early bilinguals and monolingual native speakers (Mack, 1989; Pallier *et al.*, 1997; Sebastián-Gallés and Soto-Faraco, in press). The apparent difference between studies may be due to the use of different testing procedures and/or stimuli. The present study employed a categorial discrimination test and made use of multiple natural tokens of each vowel category of interest. Mack's (1989) study, on the other hand, examined the identification and discrimination of the members of a synthetic English /i–ɪ/ continuum. Pallier *et al.* (1997) examined the identification and discrimination of the members of a synthetic Catalan /e/–/ɛ/ continuum. Finally,

Sebastián-Gallés and Soto-Faraco (in press) examined the identification of increasingly longer portions of naturally produced vowels using a modified version of the gating task.

Another possible source of the apparent difference between perceptual studies examining early bilinguals is a difference in the subjects examined. All of the early bilinguals examined by Mack (1989) had learned their L2 (either French or English) by the age of 8 years. It is likely that all of them used English more than French, for all of these individuals were living in Providence, RI, at the time of testing. The Spanish/Catalan bilinguals examined by Pallier *et al.* (1997) began to learn Catalan in Barcelona, Spain by the age of 6 years. These subjects used both Catalan and Spanish frequently, had always lived in Barcelona, and were said to be highly proficient in both of their languages. Sebastián-Gallés and Soto-Faraco (in press) also examined Spanish/Catalan bilinguals. The subjects in the present study appear to have been older (mean age=48 years) than the predominantly college-aged subjects examined in the studies just cited, and were likely to have used their L2 for a longer period of time. Their use of the L1 (Italian) appears to have been restricted because the Italian-speaking community in Ottawa is small.

The subjects tested in Barcelona by Pallier *et al.* and by Sebastián-Gallés and Soto-Faraco may have had more opportunity and need to use their L1 (Spanish) than the early bilinguals examined here. If so, then the L1 system of the Spanish/Catalan bilinguals may have exerted a stronger influence on long-term memory representations developed for L2 (Catalan) vowels than the L1 system of the Italian/English bilinguals exerted on their representations for L2 (English) vowels. Support for this inference was provided by a recent study (Flege and MacKay, under review) which examined the categorial discrimination of nine pairs of English vowels. Early Italian/English bilinguals who used Italian relatively often were found to perceive English vowels significantly less accurately than did AOA-matched group of early Italian/English bilinguals who seldom used Italian.

Even if the SLM is correct in its claim that early bilinguals often establish new categories for certain L2 sounds, it is important to note that their L2 categories may differ from monolinguals.' Grosjean (1989) suggested that bilinguals can not function exactly like ''two monolinguals in one person'' because they use two languages systems, not one. According to the SLM (Flege, 1995), the L1 and L2 phonetic systems exist in a ''common phonological space.'' The SLM proposes two mechanisms that might lead to differences between monolinguals' and bilinguals' phonetic categories. One mechanism is an age-related restriction on the use of features, or feature combinations, that are exploited in the L2 but not the L1. A multi-dimensional scaling (MDS) study by Fox *et al.* (1995) revealed that native Spanish speakers of English used fewer dimensions when perceiving vowels than did native speakers of English. McAllister *et al.* (1999) examined the perception of distinctions between phonologically long versus short Swedish vowels by native speakers of an L1 that possesses phonological length distinctions (viz., Estonian) and of L1s that do not (English and Spanish). The native Estonian subjects performed much like native speak-

ers of Swedish. However, many of the native English subjects, and even more of the native Spanish subjects, were unable to differentiate familiar Swedish words based on their phonological length specification. Flege *et al.* (1996) found that highly experienced Japanese speakers of English identified /ɹ/ tokens at nativelike rates. However, their identification of English liquids, unlike that of native English subjects, was affected by lexical familiarity. This suggested that the native Japanese subjects' representations for English /ɹ/ may have differed from the native English speakers,' perhaps because it was based on a different array of features (see Yamada, 1995, Fig. 3).

Another mechanism proposed by the SLM that might lead to differences in early bilinguals' L2 vowel is categories dissimilation. By hypothesis, a category established for an L2 vowel in the same portion of vowel space occupied by an L1 vowel may ''deflect away'' from the existing L1 category to preserve phonetic contrast within a common L1–L2 phonological space. Such a process, if confirmed, would be analogous to aspects of historical sound change (e.g., Martinet, 1955) and to between-dialect differences. For example, Moulton (1945) observed that in Swiss German dialects without an /æ/ category, /a/ is produced with central or even fronted variants. However, in dialects having an /æ/, /a/ has been deflected backward in vowel space. A case study by Mack (1990) illustrates the effect predicted for L2 speech. Voice onset time (VOT) was measured in stop consonants produced by a 10-year-old who spoke French at home and English elsewhere. The child produced /b d g/ with short-lag VOT values in both French and English, but he produced /p t k/ with much longer average VOT values in English than in French (108 vs 66 ms). The child may have increased VOT in English /p t k/ beyond values typical for English to distinguish these stops from French /p t k/. He may have increased VOT in French /p t k/ beyond values typical for French in order to distinguish these stops from the short-lag productions of /b d g/ that are found in English but not French.

The findings obtained in the present study are consistent with the hypothesis that early bilinguals establish long-term memory representations (or ''phonetic categories'') for certain L2 vowels. This conclusion has implications for another model that deals with cross-language perception, the perceptual assimilation model (PAM) of Best and colleagues (e.g., Best, 1995, 1999; Best *et al.*, 1996). The PAM generates predictions concerning the relative degree of discriminability of pairs of foreign-language phones. The subjects are typically infants, or adults who do not speak the languages(s) from which the foreign phones have been drawn. Predictions are derived from measures of the degree of perceived similarity of the foreign phones under examination to sounds in the L1. The PAM was extended to L2 learning in a study which examined the perception of English liquids and glides by adult native speakers of Japanese (Best and Strange, 1992). Subjects who had a great deal of conversational experience in English were found to perceive English /ɹ/ and /l/ tokens more like native speakers of English than did Japanese adults with less such experience. The authors concluded that English-language experience led to a ''reorganization of

perceptual assimilation of non-native phones'' (p. 327).

Importantly, the PAM differs from the SLM in the predicted effect of variations in perceived similarity on the likelihood of category formation. According to the SLM, the likelihood of category formation increases linearly as perceived cross-language similarity decreases. The PAM, on the other hand, predicts a U-shaped function. That is, category formation is predicted to be more likely for L2 sounds judged to be moderately similar to an L1 category than for L2 sounds that are highly similar, or else so dissimilar as to not be classified as an instance of an L1 category (Best and Strange, 1992, p. 327). Also, the PAM makes no predictions regarding the effect of age of L2 learning on the discrimination of L2 vowels. However, if interpreted within the PAM framework, the AOA effect obtained here would seem to suggest one of two things. One is that the perceived relation between sounds in the L1 and L2 differs as a function of the age of first exposure to an L2. The other possibility is that the age of L2 learning later affects the extent to which cross-language assimilation patterns can be reorganized.

In summary, the age at which Italian/English bilinguals were first exposed to English influenced the accuracy with which they produced and perceived English vowels, even though they were all highly experienced in English. The study provided no evidence that early Italian/English bilinguals differed from monolingual native speakers of English in producing or perceiving English vowels. The early bilinguals' performance in both domains was nativelike, at least as assessed in the present study. The results obtained here are consistent with the hypothesis that early bilinguals establish new phonetic categories for certain L2 vowels. However, additional research using more sensitive procedures is needed to further examine early bilinguals' vowel production and perception in order to determine how, or if, their phonetic categories differ from those established by monolinguals.

[1]The first half of the weighting function attenuated the signal to zero, and the second half linearly ramped the signal from 0% to 100% of amplitude. Pilot work indicated that the shortest interval that would yield unidentifiable initial consonants, once the weighting function had been applied, varied as a function of word duration. A 30-ms interval was used for 29 stimuli that were shorter than 140 ms. A 46-ms weighting function was applied to the 737 stimuli having a duration of 140–300 ms; and a 60-ms weighting function was applied to the 224 stimuli that were longer than 300 ms. Auditory assessment by two of the authors revealed that none of the initial or final consonants was clearly identifiable.

[2]When the same four-word sequences were presented a second time, the subjects were required to insert the target vowel (e.g., /i/ in *read, deed, heed, bead*) into a /b_do/.frame, thereby forming a nonword. These pro-

ductions will not be assessed here. The listeners who identified the vowels, plus four others, also rated the vowels using a six-point scale prior to the identification experiment. Thus, they were highly familiar with the range of vowels to be identified. The goodness ratings will be presented in a separate study together with an analysis of vowels spoken in extemporaneous speech (Piske *et al.*, 1999).

[3]The mean values shown in Fig. 1 were obtained using the linear predictive coding analysis function (autocorrelation method) of Multispeech (Kay Elemetrics, Inc.). A 25-ms Hamming window was placed at the acoustic midpoint of each vowel token; 14 or 16 coefficients were calculated.

[4]If the proportion of hits (H) equaled the proportion of false alarms (FA), then A′ was set to 0.5. If H exceeded FA, then $A' = 0.5 + ((H - FA)*(1 + H - FA))/((4*H)*(1 - FA))$. If FA exceeded H, then $A' = 0.5 - ((FA - H)*(1 + FA - H))/((4*FA)*(1 - H))$.

[5]The frequencies just reported were obtained by placing a 25-ms Hamming window at points located 20% and 80% into the periodic portion of the edited stimuli.

Agard F., and DiPietro R. (**1964**). *The Sounds of English and Italian* (Univ. of Chicago, Chicago).

Best, C. (**1995**). ''A direct realist perspective on cross-language speech perception,'' in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, edited by W. Strange (York, Timonium, MD).

Best, C. (**1999**). ''Development of language-specific influences on speech perception and production in pre-verbal infancy,'' in *Proceedings of the 14th International Congress of Phonetic Sciences*, edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Bailey (Department of Linguistics, University of California at Berkeley), August 1999, pp. 1261–1263.

Best, C., and Strange, W. (**1992**). ''Effects of phonological and phonetic factors on cross-language perception of approximants,'' J. Phonetics **20**, 305–330.

Best, C., Faber, A., and Levitt, A. (**1996**). ''Perceptual assimilation of nonnative vowel contrasts to the American English vowel system,'' J. Acoust. Soc. Am. **99**, 2602(A).

Bever, T. (**1981**). ''Normal acquisition processes explain the critical period for language learning,'' in *Individual Differences in Language Learning Aptitude*, edited by K. Diller (Newbury House, Rowley, MA), pp. 176–198.

Bohn, O.-S. (**1995**). ''Cross-language speech perception in adults: First language transfer doesn't tell it all,'' in *Speech Perception and Linguistic Experience, Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 273–304.

Bradlow, A., Pisoni, D., Akahane-Yamada, R., and Tohkura, Y. (**1997**). ''Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production,'' J. Acoust. Soc. Am. **101**, 2299–2310.

Busà, M. G. (**1992**). ''On the production of English vowels by Italian speakers with different degrees of accent,'' in *New Sounds 92, Proceedings of the 1992 Amsterdam Symposium on the Acquisition of Second-language Speech*, edited by J. Leather and A. James (Dept. of English, Univ. of Amsterdam), pp. 47–63.

Busà, M. G. (**1995**). *L'inglese Degli Italiani* (Unipress, Padua, Italy).

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

Flege, J. E. (**1992a**). ''Speech learning in a second language,'' in *Phonological Development: Models, Research, and Implications*, edited by C. Ferguson, L. Menn, and C. Stoel-Gammon (York, Timonium, MD), pp. 565–604.

Flege, J. E. (**1992b**). ''The intelligibility of English vowels spoken by British and Dutch talkers,'' in *Intelligibility in Speech Disorders*, edited by R. Kent (Benjamins, Amsterdam), pp. 157–232.

Flege, J. E. (**1995**). ''Second-language speech learning: Theory, findings, and problems,'' in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, edited by W. Strange (York, Timonium, MD), pp. 233–272.

Flege, J. E. (**1998**). ''The role of subject and phonetic variables,'' in *Papers From the 34th Annual Meeting of the Chicago Linguistic Society, II*, edited by K. Gruber, D. Higgins, K. Olsen, and T. Wysocki (Chicago Linguistic Society, Chicago).

Flege, J. E. (**1999a**). ''Age of learning and second-language speech,'' in *New Perspectives on the Critical Period Hypothesis for Second Language Acquisition*, edited by D. P. Birdsong (Lawrence Erlbaum, Hillsdale, MD) (to appear).

Flege, J. E. (**1999b**). ''The relation between L2 production and perception,'' in *Proceedings of the 14th International Congress of Phonetic Sciences*, edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Bailey (Department of Linguistics, University of California at Berkeley), pp. 1273–1276.

Flege, J. E., and Mackay I. R. A. (in preparation).

Flege, J. E., Bohn, O.-S., and Jang, S. (**1997a**). ''Effects of experience on nonnative subjects' production and perception of English vowels,'' J. Phonetics **25**, 437–470.

Flege, J. E., Frieda, E. M., and Nozawa, T. (**1997b**). ''Amount of native-language (L1) use affects the pronunciation of an L2,'' J. Phonetics **25**, 169–186.

Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Downs-Pruitt, J. (**1998**). ''Categorial discrimination of English and Japanese vowels and consonants by native Japanese and English subjects,'' in *Proceedings of the 135th Meeting of the Acoustical Society of America, IV*, edited by P. Kuhl and L. Crum (Acoustical Society of America, New York), pp. 2973–2974.

Flege, J. E., Munro, M. J., and MacKay, I. R. A. (**1995**). ''Factors affecting degree of perceived foreign accent in a second language,'' J. Acoust. Soc. Am. **97**, 3125–3134.

Flege, J. E., Takagi, N., and Mann, V. (**1996**). ''Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/,'' J. Acoust. Soc. Am. **99**, 1161–1173.

Fox, R. A., Flege, J. E., and Munro, M. J. (**1995**). ''The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis,'' J. Acoust. Soc. Am. **97**, 2540–2551.

Goto, H. (**1971**). ''Auditory perception by normal Japanese adults of the sounds 'l' and 'r,' '' Neuropsychologia **9**, 317–323.

Gottfried, T. L. (**1984**). ''Effect of consonant context on the perception of French vowels,'' J. Phonetics **12**, 91–114.

Grosjean, F. (**1989**). ''Neurolinguists, Beware! The bilingual is not two monolinguals in one person,'' Brain Lang **36**, 3–15.

Guion, S. G., Flege, J. E., and Loftin, J. D. (under review). ''The effect of L1 use on pronunciation in Quichua-Spanish bilinguals,'' to appear in J. Phonetics.

Jun, S.-A., and Cowie, I. (**1994**). ''Interference for 'new' and 'similar' vowels in Korean speakers of English,'' Ohio State Univ. Working Papers, Vol. 43, pp. 117–130.

Kirk, R. (**1968**). *Experimental Design: Procedures for the Behavioral Sciences* (Brooks/Cole, Belmont, CA).

Kluender, K., Lotto, A., Holt, L., and Bloedel, S. (**1998**). ''Role of experience for language-specific functional mappings of vowel sounds,'' J. Acoust. Soc. Am. **104**, 3568–3582.

Kuhl, P. (**1993**). ''Innate predispositions and the effects of experience in speech perception: The Native Language Magnet Theory,'' in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, edited by D. de Boysson-Bardies, S. deSchonen, P. Jusczyk, P. MacNeilage, and J. Morton (Kluwer Academic, The Hague), pp. 259–274.

Mack, M. (**1989**). ''Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals,'' Percept. Psychophys. **46**, 187–200.

Mack, M. (**1990**). ''Phonetic transfer in a French-English bilingual child,'' in *Language Attitudes and Language Conflict*, edited by P. H. Nelde (Dümmler, Bonn, Germany).

MacKay, I. R. A., Meador, D., and Flege, J. E. (under review). ''Factors affecting the identification of consonants in a second language.''

Major, R. (**1987**). ''Phonological similarity, markedness, and rate of L2 acquisition,'' Studies Sec. Lang. Acquis. **9**, 63–82.

Martinet, A. (**1955**). *Economie des Changements Phonétiques* (Franke, Berne).

McAllister, R., Flege, J. E., and Piske, T. (**1999**). ''The Acquisition of Swedish long vs short vowel contrasts by native speakers of English and Spanish,'' in *Proceedings of the 14th International Congress of Phonetic Sciences*, edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Bailey (Department of Linguistics, University of California at Berkeley), pp. 751–754.

Meador, D., Flege, J. E., and MacKay, I. R. A. (under review). ''Nonnative processing of English sentences presented in noise.''

Milani, C. (**1996**). ''Language contact among North American people of Italian origin,'' in *Language Contact Across the North Atlantic*, edited by P. Sture Ureland and I. Clarkson (Max Niemeyer, Tubingen), pp. 479–501.

Moulton, W. (**1945**). ''Dialect geography and the concept of phonological

space,'' Word **18**, 23–32.

Munro, M. J. (**1993**). ''Production of English vowels by native speakers' of Arabic: Acoustic measurements and accentedness ratings,'' Lang. Speech **36**, 39–66.

Munro, M. J., Flege, J. E., and MacKay, I. R. A. (**1996**). ''The effects of age of second-language learning on the production of English vowels,'' Appl. Psycholing. **17**, 313–334.

Pallier, C., Bosch, L., and Sebastián-Gallés, N. (**1997**). ''A limit on behavioral plasticity in speech perception,'' Cognition **64**, B9–B17.

Patkowski, M. (**1989**). ''Age and accent in a second language: A reply to James Emil Flege,'' Appl. Ling. **11**, 73–89.

Piske, T., and MacKay, I. A. R. (**1999**). ''Age and L1 Use effects on degree of foreign accent in English,'' in *Proceedings of the 14th International Congress of Phonetic Sciences*, edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Bailey (Department of Linguistics, University of California at Berkeley), pp. 1433–1436.

Piske, T., Flege, J. E., and MacKay, I. A. R. (**1999**). ''Non-natives' production of vowels in conversational speech,'' J. Acoust. Soc. Am. **105**, 1033(A).

Polka, L. (**1995**). ''Linguistic influences in adult perception of non-native vowel contrasts,'' J. Acoust. Soc. Am. **97**, 1286–1296.

Rochet, B. (**1995**). ''Perception and production of second-language speech sounds by adults,'' in *Speech Perception and Linguistic Experience*, edited by W. Strange (York, Timonium, MD).

Romito, L., and Trumper, J. (**1989**). ''Un problema della coarticulazione: L'isocronia rivistata,'' Acts of the 27th National Conference of the Italian Acoustical Association, pp. 449–455.

Scovel, T. (**1969**). ''Foreign accents, language acquisition, and cerebral dominance,'' Lang. Learning **19**, 245–253.

Scovel, T. (**1988**). *A Time to Speak. A Psycholinguistic Inquiry Into the Critical Period for Human Speech* (Newbury House, Cambridge, MA).

Sebastián-Gallés, N., and Soto-Faraco, S. (in press). ''On-line processing of native and non-native phonemic contrasts in early bilinguals,'' Cognition.

Sheldon, A., and Strange, W. (**1982**). ''The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception,'' Appl. Pycholing. **3**, 243–261.

Snodgrass, J., Levy-Berger, G., and Haydon, M. (**1985**). *Human Experimental Psychology* (Oxford U. P., Oxford).

Strange, W. (**1995**). ''Phonetics of second-language acquisition: Past, present, future,'' in *Proceedings of the ICPhS 95*, edited by K. Elenius and P. Branderud (Arne Stombergs, Stockholm), pp. 77–83.

Trumper, J. (**1995**). ''L'influenza di eventi macrosismici su alcune discontinuità linguistiche (Calabria),'' in *Saggi Dialettologici in Area Italo-Romanza*, edited by G. B. Pellegrini (CNR, Padova), pp. 89–103.

Yamada, R. (**1995**). ''Age and acquisition of second language speech sounds: Perception of American English /ɹ/ 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.