Factors affecting the recognition of words in a second language*

DIANE MEADOR
Speech and Language Sciences Research Laboratory
Juneau, Alaska, USA
JAMES E. FLEGE
Department of Rehabilitation Sciences
University of Alabama at Birmingham
Birmingham, Alabama, USA
IAN R. A. MACKAY
Department of Linguistics
University of Ottawa
Ottawa, Ontario, Canada

This study examined the recognition of English words by groups of native speakers of Italian who differed in age of arrival in Canada and amount of continued native language use. The dependent variable was the number of words correctly repeated in English sentences presented in noise. Significantly higher word recognition scores were obtained for early than late bilinguals, and for early bilinguals who used Italian seldom than for early bilinguals who used Italian relatively often. A hierarchical regression analysis showed that the native Italian participants' ability to perceive English vowels and consonants accounted for a significant amount of variance in the word-recognition scores independently of age of arrival, amount of L1 use, and length of residence in Canada. The native language use effect was interpreted to have arisen from differences in the extent to which the early bilinguals' Italian phonetic system influenced the representations they developed for English vowels and consonants.

Recognizing words in a second language (L2) depends on a variety of factors. These include degree of familiarity with the L2 words (e.g., Koster, 1987), the preceding sentential or discourse context (e.g., Mayo, Florentine & Buus, 1997), and degree of semantic and phonological relatedness, if any, to words found in the native language (e.g., Flege, Frieda, Walley & Randazza, 1998). The primary purpose of this study was to examine still another factor: cross-language phonetic differences in vowels and consonants found in the L2 and the native language (L1). The participants examined here were adult native speakers of Italian who had learned English as an L2. The recognition of English words was assessed by having the participants repeat as many words as possible in English sentences presented in noise.

The phonetic inventories of English and Italian differ considerably. For example, Standard Italian has only 7 contrastive vowels as compared to about 15 in English (Agard & DiPietro, 1964). In word-initial position, English /p/, /t/ and /k/ are produced with more aspiration, and English /b/, /d/ and /g/ with less voicing lead, than their Italian counterparts. English /l/ is realized as a dorsal approximant rather than a trill, as in Italian. And some word-initial English consonants (/ð/, /z/, /θ/, /h/) do not occur in the initial position of Italian words. Far fewer Italian than English words end in a consonant. Moreover, word-final English stops are less likely to be released than their Italian counterparts; and some consonants found in the final position of English words (/v/, /ʃ/, /z/, /θ/, /tʃ/, /dʒ/, /ŋ/) do not occur in the final position of Italian words.

Cross-language phonetic differences such as those just mentioned lead to differences in segmental production and perception between native speakers of English and native speakers of Italian who have learned English as an L2. Several studies have examined the effect on English pronunciation of native Italian participants' age of first exposure to English, indexed by age of arrival (AOA) in Canada. The later the participants arrived in Canada, the stronger were their foreign accents in English sentences (Flege,

* This research was supported by grant DC00257 from the National Institute for Deafness and Other Communicative Disorders. The authors thank J. Prosperine and M. Pearse for help locating participants, Fr. M. Brodeur of St. Anthony's Church, and all of the participants. Thanks are also extended to R. Lanni for the analysis of extemporaneous Italian speech samples, T. Piske for comments on an earlier draft and to Upali Lockridge for editorial assistance.

Address for correspondence
James Emil Flege, Department of Rehabilitation Sciences, University of Alabama at Birmingham, VH 503, Birmingham, Alabama, USA 35294 **E-mail**: jeflege@uab.edu

Munro & MacKay, 1995a), the less accurately they produced certain English vowels (Munro, Flege and MacKay, 1996), and the less accurately they produced certain English consonants (Flege, Munro and MacKay, 1995b). Other studies have shown that the later native Italian participants arrived in Canada, the less accurately they discriminated certain pairs of English vowels (Flege, MacKay and Meador, 1999), and the less accurately they identified English consonants presented in noise (MacKay, Meador & Flege, under review).

Based on the research just cited, it seemed likely that the native Italian participants examined here would not perceive the full range of English vowels and consonants in a fully native-like manner. If so, and if native versus non-native differences in segmental perception affect word recognition, then AOA should also affect native Italian participants' recognition of English words. More specifically, individuals who began learning English in adolescence or adulthood (''late bilinguals'') should recognize English words less accurately in non-ideal listening conditions than individuals who began to learn English as children (''early bilinguals'').

The results of three previous studies supported the hypothesis that AOA affects the accuracy of L2 word recognition. The stimuli used in a study by Bott (1992) were English sentences such as *The happy girl feeds the thirty turtles*. The sentences were degraded phonetically by processing through a LPC (linear predictive coding) vocoder. The task of native Korean participants was to write down as much of each aurally presented sentence as possible. A reanalysis of the data reported in Table 4 of Bott (1992) revealed that groups of participants who arrived in the United States between the ages of 9 and 12 years and after the age of 12 years made significantly more word recognition errors than did participants having AOAs of 0–4 and 5–8 years ($p < .01$ by Tukey's test). Mayo et al. (1997) presented English sentences at varying signal-to-noise (S/N) levels to Spanish–English bilinguals. Early bilinguals were able to identify the final word in sentences at lower S/N levels than late Spanish–English bilinguals could. (Although the early bilinguals resembled native English speakers more than the late bilinguals, they were significantly less ''tolerant'' of the masking noise than the native English participants were.) Oyama (1982a) examined native Italian men who differed in AOA (6–10 vs. 11–15 vs. 16–20 years) and length of residence (LOR) in the USA (5–11 vs. 12–18 years). English sentences such as *Shepherds seldom lose their sheep* were presented at S/N ratios that ranged from about 6 dB, on the first presentation, to about 16.5 dB, on the fourth and final presentation. A significant correlation was found to exist between AOA and the number of words correctly repeated, even when variation in LOR was partialled out.

The second hypothesis tested here was that native Italian participants who continued to use their L1 often would identify English words less accurately than those who used their L1 relatively seldom. This hypothesis was also motivated by recent research examining Italian–English bilinguals living in Canada. Flege et al. (1995a) found that amount of self-reported use of English accounted for about 15% of the variance in foreign accent ratings for English sentences. Flege, Frieda and Nozawa (1997) compared degree of foreign accent for two groups of Italian–English bilinguals who were matched for AOA in Canada (*mean* = 7 years) but differed in self-reported use of Italian. The participants in both groups were found to have detectable foreign accents. However, those who used Italian relatively often had significantly stronger foreign accents than did those who used Italian seldom.[1] Piske & MacKay (1999) observed a similar effect of L1 use on foreign accent for both early and late Italian–English bilinguals. Finally, MacKay et al. (under review) found that amount of continued L1 use affected early Italian–English bilinguals' identification of English consonants presented in noise.

The results of other studies, however, suggested that an effect of L1 use might not be observed in the present study. Flege et al. (1999) did not observe an effect of L1 use on early Italian–English bilinguals' discrimination of English vowels presented in the quiet. Bahrick, Hall, Goggin, Bahrick & Berger (1994) administered written lexical decision and vocabulary recognition tests and an aural comprehension test in both Spanish and English. The participants were 801 Spanish–English bilinguals who arrived in the United States between the ages of 10 and 26 years. The participants who used English relatively often tended to obtain higher scores on the English than the Spanish tests. A larger percentage of early-arriving than late-arriving participants obtained higher scores on the English than on the Spanish aural comprehension test (25% vs. 13%). The effect of English use on the aural comprehension scores was found to be ''negligible'' (1994, 275). The authors observed that their participants' use of Spanish and English were likely to be inversely related (see Flege et al., 1995a). If so, then variation in amount of L1 use might not have affected the

[1] Guion, Flege & Loftin (1999) observed a similar effect of L1 use on degree of L2 foreign accent for Quichua–Spanish bilinguals in Ecuador.

native Spanish participants' aural comprehension scores either.

The 72 native Italian participants examined here were assigned to one of four groups based on AOA and self-reported percentage use of Italian. Their LOR in Canada ranged from 18–48 years (*mean* = 35 years, *SD* = 7). Thus, although LOR was not examined in ANOVAs comparing the four groups, it was of interest to determine if LOR would affect the participant's recognition of English words. Oyama (1982a) observed an inverse correlation ($r = -0.39$, $p < .01$) between native Italian participants' LOR in the USA and their word recognition scores. However, we hypothesized that the longer the native Italian participants had lived in Canada and spoken English, the better they would recognize English words.

In addition to testing the three hypotheses already mentioned, we tested the hypothesis that a significant correlation would exist between the native Italian participants' perception of English vowels and/or consonants and their ability to recognize English words. A number of investigators have suggested that native versus non-native differences in segmental perception might affect L2 word recognition (Greene, Pisoni and Gradman, 1985; Nábělek and Donahue, 1984; Mayo et al., 1997). More specifically, Oyama (1982a, 45) hypothesized that some of the word recognition errors observed in her study might have been due to "individual phonemic confusions" arising from cross-language interference. Work by Bradlow and her colleagues (Bradlow & Pisoni, 1998; Bradlow, Koch, Krause & Pisoni, 1999) indirectly supported this hypothesis. Bradlow presented English words for open-set identification to 20 non-native adults who had lived in the United States for 1–8 years. The 36 "easy" words had fewer lexical neighbors differing by a single phoneme than did the 36 "hard" words (e.g., *bat* vs. *pat, sat, fat, cat, bad, back, bass, beet, boat, boot* and so on). The intent of the lexical manipulation was to vary how accurate segmental representations needed to be to distinguish words from other minimally paired words in the English lexicon. By hypothesis, more fine-grained (and accurate) representations were needed for the recognition of the hard than the easy words. Like native speakers of English, the non-native participants identified fewer hard than easy words. However, the non-native participants showed a much larger difference between the two lexical sets than native English speakers did (25% vs. 4%), even when analysis was restricted to words that were highly familiar.

The phoneme confusion hypothesis (Oyama, 1982a) was tested by examining in regression analyses the relation between the word recognition scores obtained here and previously obtained measures of the native Italian participants' perception of English phonetic segments. English consonant perception scores were drawn from a study that examined the native Italian participants' identification of 18 English consonants in the initial and final position of bisyllabic non-words presented in noise (MacKay et al., under review). Vowel perception scores were drawn from a study (Flege et al., 1999) that examined the native Italian participants' categorial discrimination of pairs of English vowels, such as /i/–/ɪ/, and English–Italian pairs such as /æ/–/a/. (The test was "categorial" in the sense that multiple natural tokens of each vowel category of interest were presented.)

In summary, the present study examined the effects of age of first exposure to English (indexed by AOA), percentage use of Italian and LOR in Canada on the recognition of English words by native speakers of Italian. Effects of the two variables used in participant selection (AOA and L1 use) were evaluated in ANOVAs. Another aim of the study was to test the hypothesis that an inaccurate perception of L2 vowels and consonants is associated with L2 word recognition difficulty. This was tested in regression analyses which examined the relation between English segmental perception scores obtained for the native Italian participants in previous studies (Flege et al., 1999; MacKay et al., to appear) and the word recognition scores obtained here.

Method

Participants

Seventy-two of the 90 participants examined here were born in Italy but had lived for many years in the Ottawa, Ontario region (*mean* = 35 years, *minimum* = 18 years). All but two of the native Italian participants were from working-class backgrounds, as indicated by parental occupation. Most of the native Italian participants were recruited through the predominantly Italian Roman Catholic parish in Ottawa where testing was carried out. The remaining 18 participants were monolingual native speakers of English who were also long-time residents of Ottawa. The mean age of the 48 female and 42 male participants was 48 years (SD = 6). Participants in the 5 groups did not differ in age, $F(4,85) = .05$, $p > .10$. All 90 participants 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; no participant reported a history of auditory disorder.

As summarized in Table 1, each of the five groups

Table 1. *Mean characteristics of the five groups of participants; standard deviations are in parenthesis*

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

Note: Age, chronological age, in years; AOA, age of arrival in Canada, in years; LOR, length of residence in Canada, in years; % Italian, self-reported percentage use of Italian.

consisted of 18 individuals. Three of the four native Italian groups were selected solely on the basis of AOA. The fourth native Italian group, ''Early-low'', was then formed by recruiting participants with the same AOA as those in the Early group but who used Italian less (Early-low: 8%, Early: 32%). The native Italian participants' AOAs were correlated inversely with LOR, r (70) = −0.62, $p < .01$, which led to a significant effect of group on LOR, F (3,68) = 16.1, $p < .01$. Also, the earlier the native Italian participants 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, $p < .01$.

The responses to questionnaire items pertaining to where Italian was used suggested that the participants in the Early-low group tended to use Italian mostly when visiting relatives. They were less likely than participants in the Early group were to report using Italian at home, at work, on the telephone or at social gatherings. The self-reports of percentage L1 use appear to have been reliable. A correlation existed between the estimates of Italian use given by 62 participants in the present study and the Italian use estimates given by the same participants in a study (Flege et al., 1995a, b) carried out three years earlier, r (60) = .84, $p < .01$. There was a strong inverse correlation between estimates of English and Italian use on separate portions of the language background questionnaire, r (70) = −0.96, $p < .01$. This suggested that the language use estimates were valid, as well. We nonetheless carried out two additional tests to evaluate the validity of the participants' estimates of Italian use.

If participants in the Early-low group really did use Italian less than those in the Early group, one would expect them to speak Italian more poorly than participants in the Early group. Competence in Italian was evaluated by examining the frequency of errors in the extemporaneous production of Italian. A native Italian speech–language pathologist who was blind as to the group membership of the participants orthographically transcribed recorded Italian speech samples. She noted all grammatical errors involving noun phrase agreement, subject–verb agreement, prepositions, verb tense, auxiliary verbs and pronouns. Lexical errors were also noted. (Divergences from standard Italian that could be attributed to dialect influences were not counted as errors.)

The total number of content words produced by participants in the four native Italian groups (Early-low: 156, Early: 162, Mid: 206, Late: 162) did not differ significantly, F (3,68) = 1.72, $p > .10$. The percentage of errors made by each participant was calculated by dividing the number of errors observed for each participant by the number of content words produced. The difference in the mean percentage of errors obtained for the four native Italian groups (Early-low: 14%, Early: 11%, Mid: 9%, Late: 7%) was significant, F (3,68) = 8.21, $p < .01$. A Tukey's test revealed that the participants in both the Early-low group and the Early group made more errors than did participants in the Late group. However, only participants in the Early-low group made significantly more errors than participants in the Mid group did ($p < .05$). This last finding is consistent with the belief that participants in the Early-low group used Italian less often than those in the Early group.

As mentioned in the Introduction, variation in amount of L1 use has been associated with variation in degree of perceived foreign accent in an L2. We therefore obtained foreign accent ratings for the participants' production of three short English sentences (*The red book was good; I can read this for you; He turned to the right*). These sentences were digitized, then presented randomly to five native speakers of English (four from Ontario, one from Wisconsin). The listeners rated each sentence three times using a scale ranging from ''strongest foreign accent'' (1) to ''native English – no accent'' (9). The difference in mean ratings obtained for participants in the five groups (native English: 7.1, Early-low: 6.2, Early: 5.7, Mid: 4.8, Late: 2.8) was significant, F (4,85) = 51.3, $p < .01$. Sentences produced by participants in the Early, Mid and Late groups, but not those produced by participants in the Early-low group

received significantly lower ratings than did the native English participants' sentences ($p < .05$ by Tukey's test). The finding that participants in the Early group but not those in the Early-low group had a detectable foreign accent is consistent with the view that participants in the Early group used Italian more than those in the Early-low group.

Stimuli

The 10 sentences shown in Table 2 were used as stimuli. Each sentence had the form noun phrase + verb + noun phrase. The noun phrase always consisted of article + adjective + noun. The article was always "the", and the verb was always in the past tense. Each sentence contained nine syllables and five content words. Half of the noun phrases contained monosyllabic adjectives and bisyllabic nouns, whereas the remaining noun phrases contained bisyllabic adjectives and monosyllabic nouns.

Native and non-native speakers have been found to differ in their use of the preceding sentence context when identifying the final word in sentences presented in noise (Florentine 1985a, b; Mayo et al. 1997). Such contextual effects, if obtained in the present study, might obscure the effect of native versus non-native differences in segmental perception on L2 word recognition. We therefore decided to examine semantically unpredictable sentences. This reduced the likelihood that a word could be "guessed" from the surrounding, especially preceding, context (see Benoit, Grice & Hazan, 1996). Take, for example, the sentence, *The blond dentist ate the heavy bread.* A word such as "dentist" is less likely than, say, "actress" to occur with "blond"; and "heavy" is less likely than "fresh" to co-occur with "bread." Admittedly, however, the sentences used here did not eliminate every possible co-occurrence pattern.

Previous research suggested that L2 words will be less familiar to non-native speakers than to native speakers, and that such differences might influence L2 word recognition (e.g., Kalikow, Stevens & Elliott, 1977; Connine, Mullinex, Shernoff & Yelen, 1990; de Jonge, Garber & Pisoni, 1991; Flege, Takagi & Mann, 1996). We therefore attempted to use only words that were likely to be known and familiar to the native Italian participants. The subjective familiarity of the words in the test sentences ranged from 5.5–7.0 ($M = 6.9$) on a seven-point scale for young adult native speakers of English. In fact, all but six of the 50 words found in the sentences received the maximum subjective familiarity rating of 7 in a study by Nusbaum, Pisoni & Davis (1984). The frequency range of the words used in the stimulus sentences ranged from 3–1635 ($M = 140$) per million (Kučera and Francis, 1967). Work by Bradlow & Pisoni (1998) suggested that neighborhood density might affect the non-native participants' ability to recognize words in the stimulus sentences. That is, words having many phonological neighbors (e.g., *bat)* might be more difficult to recognize than words having fewer neighbors, such as *jeep* (see, e.g., Luce, 1986). The neighborhood density of the words examined here ranged from 0–39 ($M = 10.8$). The average text frequencies of the words' phonological neighbors ranged from 0–1134 ($M = 11.0$).

Table 2. *The semantically unpredictable sentences used as stimuli*

The blond dentist ate the heavy bread.
The nervous boy saw the hollow tree.
The happy clerk bought the new sweater.
The tall teacher touched the big flower.
The worried girl caught the small button.
The fat pilot threw the little rock.
The busy nurse found the silver key.
The tired writer smelled the rotting leaves.
The clumsy man heard the fourth whistle.
The brilliant child wrote the short sonnet.

A native speaker of English read the sentences at a moderate rate. They were digitized at 22.04 kHz with 16–bit amplitude resolution, then normalized to 50% of the full-scale amplitude. Copies of a single 3.0-s pink noise segment were then adjusted to 40%, 20%, 10% and 5% of the full-scale value. The levels of the noise segments were selected on the basis of pilot tests to yield response functions that avoided both ceiling and floor effects. The noise segments were ramped on over the first 30 ms, and ramped off over the final 20 ms to avoid clicks. A 200-ms segment of silence was then appended to the beginning of the noise segments. Finally, copies of the sentence waveform files, whose durations ranged from 2.1–2.5 s ($M = 2.3$ s), were added to the pink noise files. This procedure yielded 4 sets of 10 sentences each having nominal S/N ratios of −6 dB, 0 dB, +6 dB and +12 dB. The rationale for basing the S/N estimates on peak intensities rather than RMS (root mean square) values is that the sentences contained segments that varied greatly in intensity as well as the number and duration of silent intervals (e.g., the closure interval of /k/ in *key*).

Procedure

The participants were tested one at a time in a quiet room. The sentences were presented via a loud-

speaker at a self-selected comfortable listening level. The presentation level was established individually for each participant before the experiment began using two practice sentences.

Previous research (e.g., McAllister, 1990) has shown that non-native speakers are less "tolerant" of masking noise than are native speakers when faced with the need to comprehend connected speech materials. However, given our desire to replicate and extend the study by Oyama (1982a), the study was not optimally designed to further test the hypothesis that the addition of noise affects native and non-native speakers differently. Each of the 10 test sentences were presented 4 times in a row. The order of the four S/N levels for each sentence was fixed, not counterbalanced. The first presentation of each sentence was at the −6 dB S/N level, the second presentation was at 0 dB, the third presentation was at +6 dB and the final presentation was at +12 dB. The participants were told that they would hear each sentence four times in a row, with less noise on each successive presentation. They were asked to repeat as many words as possible after each presentation of a sentence and to guess if necessary.

The participants took part in three tasks before the sentence repetition experiment reported here. In the first task, they repeated aurally presented English sentences. (The sentences were later rated for degree of perceived foreign accent; see above.) Next, the participants repeated isolated English words containing a variety of English vowels (see Flege et al., 1999). Finally, they responded extemporaneously to short questions presented via a loudspeaker. An analysis of their responses to the Italian questions was reported in the Method section.

No attempt was made to vary language "set" (see Bohn and Flege, 1993, and references therein) or language "mode" (Grosjean, to appear). The experiments were carried out in English by a native English-speaking experimenter who does not speak Italian. Identical testing procedures were used for all four groups of native Italian participants. One procedural difference existed between the native Italian participants and the native English participants. In the extemporaneous language production task, the Italian–English bilinguals responded to four questions posed in Italian and four questions posed in English. The English monolinguals, on the other hand, heard all eight questions, but they responded only to the English questions.

Analysis

The participants' vocal responses were recorded for later analysis using a directional headset microphone (Shure SM10A). A native speaker of English (DM), who was blind as to the participants' group membership, auditorily evaluated the five content words found in each sentence (e.g., the five underlined words in *The blond dentist ate the heavy bread*). The observed errors included productions of a word that differed from the target word by a single phoneme (e.g., "blind" for *blond*). Semantic substitutions were also observed (e.g., "climbed" for *saw* in *The nervous boy saw the hollow tree;* "tall" for *big*). A word was not counted as an error if a first attempt was later corrected (e.g., "long . . . blond!" for *blond*), or if it differed from the target word solely in terms a minor articulatory error (e.g., [æpiy] for *happy*, or [fort] for *fourth*). We reasoned that such errors were just as likely to be segmental production errors as to be segmental perception errors. (As already mentioned, [h] does not occur in the initial position of Italian words, and [θ] does not occur in the final position of Italian words.)

The weighting scheme used by Oyama (1982a) was not used here. Instead, we simply tabulated the numbers of content words that were produced correctly by each participant in response to sentences presented at the four S/N levels (−6, 0, +6, +12 dB). The percentage of words correctly repeated on the four successive presentations of the sentences (i.e., at each of the four S/N levels) was then calculated. Each of these percentages was based on a maximum of 50 possible correct repetitions (5 content words x 10 sentences).

None of the scores was 0%, but 14 (3.8%) of the scores – all for the fourth and final presentation of the sentences – were 100%. Accordingly, we transformed the percentage correct scores using the arcsin transformation recommended by Kirk (1968). However, the pattern of results obtained in ANOVAs examining the transformed scores was the same as those obtained when the raw percentage correct scores were examined. Therefore, only the latter ANOVAs will be reported in the next section.

Results

Effect of AOA

The purpose of the analysis presented in this section was to examine the effect of AOA on the recognition of English words. Just three of the four native Italian groups – those differing primarily in AOA – will be examined here. (The results obtained for the early bilinguals who seldom used Italian will be examined in the following section.)

Figure 1(a) shows the word recognition scores obtained for the native speakers of English and the

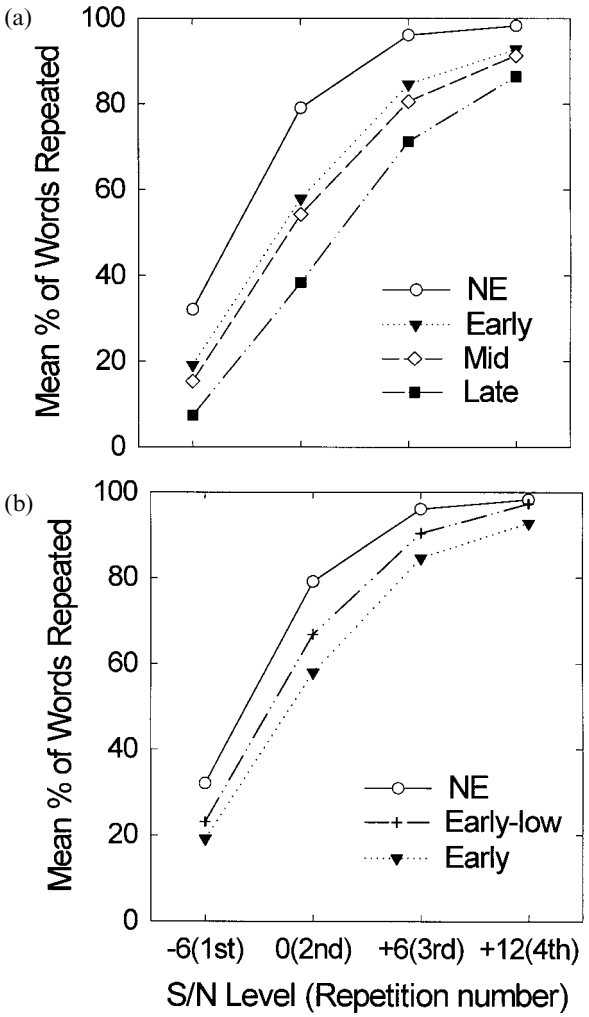

Figure 1. (a) The mean percentage of words repeated correctly by native speakers of English (NE) and groups of native Italian speakers who differed in age of arrival to Canada; (b) the mean scores obtained for the NE speakers and two groups of early Italian–English bilinguals who differed in amount of L1 use (see text)

participants in the Early, Mid and Late groups. Averaged over the four S/N levels, the native English speakers repeated more words (M = 77%) than did participants in the Early group (M = 64%), the Mid group (M = 60%) or the Late group (M = 51%). As expected, the percentage of words repeated correctly by participants in all four groups increased systematically across the four presentations of the sentences (i.e., with increasingly higher S/N levels).

The scores in Figure 1(a) were submitted to a (4) group x (4) S/N level ANOVA. It yielded significant main effects of group, F (3,68) = 31.8, $p < .01$, and S/N level, F (3,204) = 1768, $p < .01$, and a significant interaction, F (9,204) = 9.7, $p < .01$. Tests of simple main effects suggested that the interaction arose primarily from a different pattern of between-group differences at the four S/N levels. The effect of group was significant at all four S/N levels ($p < .01$). At the −6 dB, 0 dB and +6 dB S/N levels (i.e., on the first three presentations of the sentences), native English speakers obtained higher scores than did the participants in all three native Italian groups ($p < .05$ by Tukey's test). However, at the +12 dB S/N level (i.e., on the final presentation of the sentences), the native English speakers obtained significantly higher scores than participants in the Mid and Late groups but not the Early group ($p < .05$). (Tukey's tests also revealed that the Early group's scores were higher than the Late group's at the −6, 0, and +6 dB levels, and the Mid group's scores were higher than the Late group's at the 0 dB and +6 dB levels ($p < .05$). However, no significant differences between the native Italian groups were observed at the +12 dB level.)

A different effect of noise across the four groups also contributed to the significant two-way interaction. The effect of noise was highly significant for all four groups ($p < .01$). Tukey's tests revealed that all pair-wise differences between scores obtained for the four S/N levels (presentations) was significant for the Early, Mid, and Late groups ($p < .05$). However, the difference in scores obtained at the +6 and +12 dB S/N levels was non-significant for the native English group.

Effect of L1 use

Figure 1(b) shows the word recognition scores obtained for the native speakers of English and the two groups of early bilinguals who differed in amount of L1 use. Once again, the percentage of words repeated correctly increased systematically across the four successive presentations of the sentences (i.e., as the S/N levels increased from −6 dB to +12 dB). As mentioned earlier, the native English speakers repeated more words overall (M = 77%) than the participants in the Early group did (M = 64%). The participants considered here for the first time – those in the Early-low group – obtained a mean score (69%) that was intermediate to the mean scores obtained for participants in the native English and Early groups.

The word recognition scores were examined in a (3) group x (4) S/N level ANOVA. It yielded significant main effects of group, F (2,51) = 30.1, $p < .01$, and S/N level, F (3,153) = 1768, $p < .01$ and a significant two-way interaction, F (6,153) = 6.63, $p < .01$. The results of simple effects tests suggested that the interaction arose primarily from a different pattern of between-group differences at the four S/N

levels. The effect of group was significant at all four S/N levels ($p < .01$). Tukey's tests revealed that the native English speakers repeated significantly more words than did both groups of early bilinguals at the −6, 0 and +6 dB levels. However, they repeated significantly more words than did participants in the Early group, but not those in the Early-low group, at the +12 dB level ($p < .05$). The Early-low group did not differ significantly from the Early group at the +12 dB level (the final presentation), according to the Tukey's test. However, the Early-low participants repeated significantly more words than participants in the Early group did at the −6, 0 and +6 dB levels ($p < .05$).

The effect of noise was highly significant for all three groups ($p < .01$). Tukey's tests revealed that all pair-wise differences between the four S/N levels were significant for the Early-low and Early groups. However, as already mentioned, the difference between the +6 and +12 dB levels was non-significant for the native English group.

Regression analyses

Analyses presented in the last two sections showed that AOA and L1 use affected the native Italian participants' ability to recognize words in semantically unpredictable English sentences. One possible interpretation of these findings is that the native Italian participants recognized fewer words than the native English participants did because the English words were relatively less familiar to them (see Flege et al., 1996). Another possibility, however, is that difficulty identifying English vowels and consonants impeded the native Italian participants' recognition of English words. The aim of the analyses presented here was to evaluate the second interpretation. We did this by examining the relation between the word recognition scores obtained here and segmental perception scores obtained for the 72 native Italian participants in two previous studies.

One set of segmental perception scores used in the regression analyses was taken from a study by MacKay et al., under review). This study examined the identification of 18 English consonants (/m n s r l p t k b d g f v ʃ tʃ dʒ θ z/ in the initial and final positions of bisyllabic non-words (e.g., /ˈbɑdo/, /hodas/) spoken by a native speaker of English. The stimuli containing initial and final consonants were presented in counterbalanced order to the participants via headphones at +12, +6, 0 and −6 dB S/N levels. The participants identified each consonant using orthographically presented keywords. The percentages of errors each participant made in the identification of initial and final consonants were then computed. These percentage error scores were each based on 72 forced-choice judgments (18 consonants x 4 S/N levels).

The second set of segmental perception scores were drawn from a study (Flege et al., 1999) which used a categorial discrimination test to assess the discrimination of the contrast between four pairs of English vowels (such as the vowels in "beat" and "bit"), and the contrast between four English vowels and a neighboring Italian vowel. All stimuli were presented in the quiet. The participants' task was to identify the serial position of the "odd item out" from among the three stimuli presented on each trial, or else to respond "none" if three instances of a single vowel category were presented. The participants' phonetic sensitivity to differences between each of the eight vowel contrasts was estimated by calculating A′ scores (see Snodgrass, Levy-Berger & Haydon, 1985). The A′ scores were based on the proportion of "hits" (correct identifications of an odd item out in trials containing one token drawn from a different category than the other two tokens) and "false alarms" (incorrect choices of an odd item out in trials containing three instances of a single category). To obtain a high score, the participants had to respond to relevant phonetic differences while ignoring auditorily accessible differences (e.g., in voice quality) that were phonetically irrelevant. The A′ scores examined here were an average obtained for all eight vowel contrasts examined.

Table 3 shows simple correlations that were computed between nine variables and the overall word recognition scores (based on a maximum of 200 possible correct repetitions of words). Two of the nine correlations – those involving the frequency of errors in extemporaneous Italian (see Method section) and chronological age – were non-significant. Four significant inverse correlations were obtained. The earlier the native Italian participants arrived in Canada, the less they used Italian, the fewer errors they made identifying word-initial English consonants in noise, and the fewer errors they made identifying word-final English consonants in noise, the more words they recognized in the present study. Three significant positive correlations were obtained. The better (i.e., less foreign-accented) was the native Italian participants' overall pronunciation of English sentences, the more accurately they discriminated English vowels, and the longer they had lived in Canada, the more words the native Italian participants recognized. The significant positive correlation observed here between LOR and word recognition scores runs counter to the significant inverse correlation obtained by Oyama (1982a). However, the correlation between the foreign accent ratings and the

Table 3. *Simple correlations for 72 native speakers of Italian; the vowel discrimination and consonant identification scores were drawn from previous studies (see text); an asterisk indicates significance at the .001 level*

		2	3	4	5	6	7	8	9	10
1	Chronological age (years)	.12	.68*	.04	.12	−.08	.00	.24	.01	−.07
2	Age of arrival in Canada (years)		−.62*	.39*	−.49*	−.78*	−.72*	.43*	.30	−.64*
3	Length of residence in Canada (years)			−.26	.44*	.51*	.51*	−.10	−.21	.41*
4	Self-reported % use of Italian				−.30	−.48*	−.26	.22	.22	−.41*
5	% errors in extemporaneous Italian					.34	.20	.02	.13	.19
6	Foreign accent in English sentences						.62*	−.38	−.29	.69*
7	Vowel discrimination accuracy (A′)							−.41*	−.36	.66*
8	% initial consonant ID errors								−.53*	−.59*
9	% final consonant ID errors									−.49*
10	% word recognition errors									

Note: 6, mean rating on a nine-point scale that ranged from "strongest foreign accent" to "native English-no foreign accent"; 8, percentage of errors identifying word-initial English consonants in noise; 9, percentage of errors identifying word-final English consonants in noise; 10, percentage of errors made in repeating words in English sentences presented in noise (*maximum* = 200).

word recognition scores obtained here is consistent with the results obtained in two studies by Oyama (1982a, 2b).

As in previous studies (e.g., Oyama, 1982a, b; Bahrick et al., 1994), the native Italian participants' AOA in Canada was inversely correlated with LOR. The correlation between the AOA and the word recognition scores remained significant when LOR was partialled out, r (69) = −0.55, $p < .01$, but the correlation between LOR and the word recognition scores became non-significant when variation in AOA was partialled out, r (69) = .02, $p > .10$. This supports the conclusion (Oyama, 1982a) that AOA is a more important predictor of L2 word recognition than LOR is.

A step-wise multiple regression analysis was carried out to examine the relation between the word recognition scores and three measures of segmental phonetic perception (average vowel discrimination accuracy, percentage of initial consonant identification errors and percentage of final consonant identification errors). As summarized in Table 4, this analysis accounted for 56% of the variance in the 72 native Italian participants' word recognition scores, F (2,69) = 43.3, $p < .01$. The vowel discrimination scores accounted for 43% of the variance, and the word-initial consonant identification scores accounted for 12%.

The finding just reported might be taken as support for the hypothesis that the native Italian participants' perception of English phonetic segments affected their ability to recognize English words. However, this is not a necessary conclusion, for the segmental perception scores were correlated with LOR and AOA (see Table 4). A hierarchical regression analysis was therefore undertaken. The variables entered at Step 1 were AOA, LOR, percentage use of Italian and chronological age. The percentage of errors made in extemporaneous production of Italian was entered at Step 2. The Italian error scores were used to index the strength of the native Italian participants' L1 system. We reasoned that the stronger the L1 system was, the more it might influence the native Italian participants' perception of English phonetic segments. The three segmental perception scores were entered at Step 3. Entering the

Table 4. *The results of a step-wise regression analysis examining the relation between three measures of segmental perception and the word recognition scores obtained for the 72 native Italian participants; an asterisk indicates significance at the .01 level*

Predictor variables	*Beta*	R^2	R^2 change	*F*
Vowel discrimination accuracy (A′)	.499	.433	.433	32.2*
% initial consonant ID errors	−.386	.557	.124	19.3*

Table 5. *The results of a hierarchical regression analysis examining the word recognition scores obtained for the 72 native Italian participants; an asterisk indicates significance at the .01 level*

Step	Predictor variables	*Beta*	R^2	R^2 change	*F*
1	Age of Arrival	−.263	.415	.415	68.7*
2	% initial consonant ID errors	−.345	.538	.123	20.4*
3	vowel discrimination accuracy (A′)	.328	.589	.051	8.4*

segmental perception scores last provided a way to test the hypothesis that L2 segmental perception is related to L2 word recognition independently of other potentially important factors.

As summarized in Table 5, the hierarchical regression analysis accounted for 59% of the variance in the word recognition scores, F (3,68) = 32.5, $p < .01$. As expected (see above), AOA but not LOR accounted for a significant amount of variance at Step 1. The Italian error scores did not account for additional variance at Step 2. The segmental perception scores accounted for a significant additional amount (17%) of variance at Step 3.

Discussion

The present study replicated the AOA effect obtained in previous studies (Oyama, 1982a; Bott, 1992; Mayo et al. 1997). Early bilinguals who arrived in Canada as children were found to obtain higher word recognition scores than late bilinguals who arrived in late adolescence or early adulthood. The effect of the participants' AOA in Canada on the word recognition scores might be taken as support for the existence of a "sensitive period" that "profoundly affects the perception of speech" (Mayo et al., 1997, 692). According to some (e.g., Scovel, 1988), AOA might be taken as an index of the learner's state of neurological development at the time L2 learning begins. On this view, the ability to learn L2 speech – including the vowels and consonants found in an L2 – decreases after a critical period is passed. According to others (e.g., Flege, 1995), AOA indexes the state of development of the learner's native language (L1) phonetic categories at the time L2 learning begins. On this latter view, the ability to establish new categories for phonetic segments encountered in an L2 decreases with age because of an increased likelihood that existing L1 categories will assimilate L2 vowels and consonants (see Best and Strange, 1992).

Another possibility is that the observed AOA effect arose, at least in part, from differences in lexical familiarity. As mentioned in the Introduction, Bradlow (Bradlow and Pisoni, 1998; Bradlow et al., 1999) has shown that lexical variables may affect non-native speakers' recognition of English words presented in isolation. Non-native participants recognized English words having just a few, infrequent lexical neighbors ("easy" words) more often than English words having many frequent neighbors ("hard" words).

The sentences examined in the present study were made up of words that are highly familiar to young adult monolingual speakers of English (Nusbaum et al., 1984). It is nevertheless likely that these words were less familiar to the native Italian participants than they would be for native English speakers of the same age (see, e.g., Flege et al., 1996). To the extent that this was true, it may have contributed to the native versus non-native differences observed here, or to differences between the native Italian groups differing in AOA. Pronunciation may also have played a role. The relative ease or difficulty in pronouncing an L2 word may influence whether it is learned or not (Rodgers, 1969; Ellis and Beaton, 1993a, b).

The results obtained here do not rule out a critical (or sensitive) period interpretation, or a lexical interpretation. However, the present study demonstrated that at least one other factor plays a role in determining how well L2 words will be recognized in non-ideal listening conditions. Early bilinguals who spoke Italian relatively often (designated the "Early" group) obtained significantly lower word recognition scores than the native speakers of English did for all four presentations of the sentences. However, the early bilinguals who spoke Italian seldom (the "Early-low" group) differed from the native English speakers only for the first three presentations of the sentences (i.e., at the −6, 0 and +6 dB S/N levels). More importantly, the Early group obtained signifi-

cantly lower scores than the Early-low group did for the first three presentations of the English sentences.

The two groups of early bilinguals were matched for age of first exposure to English. This means that the difference between the Early and Early-low groups cannot be attributed to a critical or sensitive period for L2 speech learning that derives from normal neurological development (Scovel, 1988). Of course, the native Italian participants' self-reported use of Italian was inversely related to their self-reported use of English. One might therefore attribute the "L1 use" effect obtained here to variation in English-language experience. We think that this is unlikely, however, because the native Italian participants had lived in Canada for at least 18 years. Moreover, if the L1 use effect was a covert "L2 practice" effect, one would expect to observe an effect of LOR in Canada. However, as in previous work (e.g., Oyama, 1982a), variation in LOR had little effect on the word recognition scores obtained here. We can also be fairly confident that the significant difference between the two groups of early bilinguals was not due to lexical factors. This is because the participants in the Early and Early-low groups had lived in Canada for an average of 40 years and had received nearly all of their education in English-speaking Canadian schools.

The observed L1 use effect may therefore have been due to differences in the extent to which the native Italian participants' L1 phonetic system affected the representations they developed for English vowels and consonants. This interpretation assumes that the more the early bilinguals used Italian, the more representations for Italian phonetic segments influenced the representations they established for English phonetic segments (or the less likely the early bilinguals were to establish independent representations for English vowels and consonants). This working hypothesis will need to be further validated, of course, before it can be accepted.

Previous work (Oyama, 1982a; Nábělec & Donahue, 1984; Greene et al., 1985; Mayo et al., 1997; Bradlow and Pisoni, 1998; Bradlow et al., 1999) led us to hypothesize that the native Italian participants' accuracy in perceiving English vowels and consonants would be related to their recognition of English words. This was tested by regressing segmental perception scores obtained for the native Italian participants in two previous studies (Flege et al., 1999; MacKay et al., under review) onto the word recognition scores obtained here. The native Italian participants' ability to identify word-initial English consonants (in noise) and to discriminate English vowels (in the quiet) accounted for 57% of the variance in the word recognition scores in a step-wise regression analysis. A second, more conservative hierarchical regression analysis was also undertaken. In it, the segmental perception scores were entered following other variables that might affect the word recognition scores. This analysis revealed that the native Italian participants' ability to identify word-initial English consonants and to discriminate English vowels accounted independently for 17% of the variance in the word recognition scores.

This is the first time, as far as we know, that the perception of phonetic segments in an L2 has been directly linked to the recognition of L2 words. The finding is notable in view of how experienced in English the native Italian participants were. They had lived in Canada for an average of 35 years at the time of testing (*range*: 18–48 years) and reported using English 73% of the time on average (*range*: 10–98%). In fact, 80% of the native Italian participants reported using English as often, or more often than Italian; and 74% of them indicated that they would choose English if they could retain only one of their two languages. Future research should be carried out to determine if L2 segmental perception accounts for an even larger percentage of the variation in L2 word recognition scores for individuals who are less experienced in their L2 than the native Italian participants were.

In summary, native Italian participants' recognition of English words was assessed by having them repeat as many words in English sentences as possible. Three important findings were obtained. Significantly higher word recognition scores were obtained for early than late bilinguals. Early bilinguals who used Italian seldom obtained significantly higher word recognition scores than did early bilinguals who used Italian relatively often. Finally, a regression analysis showed that the native Italian participants' ability to identify English consonants and discriminate English vowels accounted for a significant amount of variance in the word-recognition scores independently of AOA and L1 use. The L1 use effect was interpreted to have arisen from differences in the extent to which the early bilinguals' Italian phonetic system influenced the representations they developed for English vowels and consonants.

References

Bahrick, H., Hall, L., Goggin, J., Bahrick, L. & Berger, S. (1994). Fifty years of language maintenance in bilingual Hispanic immigrants. *Journal of Experimental Psychology: General* 123, 264–283.

Benoit, C., Grice, M. & Hazan, V. (1996). The SUS test: a method for the assessment of text-to-speech synthesis intelligibility using semantically unpredictable sentences. *Speech Communication* 18, 381–392.

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

Bohn, O.-S. & Flege, J. E. (1993). Perceptual switching in Spanish/English bilinguals: evidence for universal factors in stop voicing judgments. *Journal of Phonetics*, 21, 267–290.

Bott, S. M. (1992). Speech intelligibility and bilingualism: the effects of age of acquisition. Unpublished Ph.D. thesis, University of Illinois, Dept of Linguistics.

Bradlow, A. R., Koch, D., Kraus, N. & Pisoni, D. B. (1999). Effects of talker and lexical variability on speech perception by "special populations." *Proceedings of the 14th International Congress of Phonetic Sciences*, 1953–1956.

Bradlow, A. R. & Pisoni, D. B. (1998). Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors. *Research on Spoken Language Processing, Progress Report* 22, 73–195. Indiana University, Dept of Psychology.

Connine, C., Mullinex, J., Shernoff, E. & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology* 16, 1084–1096.

de Jonge, C., Garber, E. & Pisoni, D. B. (1991). Assessing non-native speakers' familiarity of spoken and written words in English: a first report. *Research on Speech Perception* 17, 319–339. Indiana University, Dept of Psychology.

Ellis, N. C. & Beaton, A. (1993a). Factors affecting the learning of foreign vocabulary: imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology* 46A, 533–558.

Ellis, N. C. & Beaton, A. (1993b). Psycholinguistic determinants of foreign vocabulary learning. *Language Learning* 43, 559–617.

Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience*, pp. 233–277. Timonium, MD: York Press.

Flege, J. E. (1998). The role of subject and phonetic variables in L2 speech acquisition. In M. Gruber, D. Higgins, K. Olsen & T. Wysocki (eds.), *Papers from the 34th Annual Meeting of the Chicago Linguistic Society, Vol. II, The Panels*, pp. 213–232. Chicago: Chicago Linguistic Society.

Flege, J. E., Frieda, E. & Nozawa, T. (1997). Amount of native-language (L1) use affects pronunciation of an L2. *Journal of Phonetics* 25, 169–186.

Flege, J. E., Frieda, E. M., Walley, A. C. & Randazza, L. (1998). Lexical factors and segmental accuracy in second-language speech production. *Studies in Second Language Acquisition* 20, 155–188.

Flege, J. E., MacKay, I. R. A. & Meador, D. (1999). Native Italian subjects' production and perception of English vowels. *Journal of the Acoustical Society of America* 106, 2973–2987.

Flege, J. E., Munro, M. J. & MacKay, I. R. A. (1995a). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97, 3125–3134.

Flege, J. E., Munro, M. J. & MacKay, I. R. A. (1995b). Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16, 1–26.

Flege, J. E, Takagi, N. & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *Journal of the Acoustical Society of America* 99, 1161–1173.

Florentine, M. (1985a). Non-native listeners' perception of American-English in noise. *Proceedings of Inter-noise* 85, 1021–1024.

Florentine, M. (1985b). Speech perception in noise by fluent, non-native listeners. *Proceedings of the Acoustical Society of Japan* H–85–16.

Greene, B. G., Pisoni, D. B. & Gradman, H. L. (1985). Perception of synthetic speech by non-native speakers of English. *Research on Speech Perception, Progress Report No. 11*. Indiana University, Dept of Psychology.

Grosjean, F. (to appear). The bilingual's language modes. In J. L. Nicol and T. D. Langendoen (eds.), *Language processing in the bilingual*. Oxford: Blackwell.

Guion, S., Flege, J. E. & Loftin, J. (1999). The effect of L1 use on foreign accent ratings in Quichua–Spanish bilinguals. *Proceedings of the 14th International Congress of Phonetic Sciences*, 1471–1474.

Kalikow, D., Stevens, K. & Elliott, L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America* 61, 1337–1351.

Kirk, R. (1968). *Experimental design: procedures for the behavioral sciences*. Belmont, CA: Brooks/Cole.

Koster, C. J. (1987). *Word recognition in foreign and native language*. Dordrecht: Foris.

Kučera, F. & Francis, W. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Luce, P. (1986). Neighborhoods of words in the mental lexicon. *Research on Speech Perception, Technical Report 6*. Indiana University, Dept of Psychology.

MacKay, I. R. A., Meador, D. & Flege, J. E. (under review). The identification of English consonants by native speakers of Italian.

Mayo, L., Florentine, M. & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech and Hearing Research* 40, 686–693.

McAllister, R. (1990). Perceptual foreign accent: L2 user's comprehension ability. In J. Leather (ed.), *Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second-language Speech*. University of Amsterdam, Dept of English.

Munro, M. J., Flege, J. E. & MacKay, I. R. A. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics* 17, 313–334.

Nábělek, A. K. & Donahue, A. M. (1984). Perception of consonants in reverberation by native and non-native

listeners. *Journal of the Acoustical Society of America* 75, 632–634.

Nusbaum, H. C., Pisoni, D. B. & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: measuring the familiarity of 20,000 words. *Research on Speech Perception, Progress Report No. 10*. Indiana University, Dept of Psychology.

Oyama, S. (1982a). The sensitive period and comprehension of speech. In S. Krashen, R. Scarcella & M. Long (eds.), *Child–adult differences in second language acquisition*, pp. 39–52. Rowley, MA: Newbury House.

Oyama, S. (1982b). A sensitive period for the acquisition of a non-native phonological system. In S. Krashen, R. Scarcella & M. Long (eds.), *Child–adult differences in second language acquisition*, pp. 20–38. Rowley, MA: Newbury House.

Piske, T. & MacKay, I. R. A. (1999). Age and L1 use effects on degree of foreign accent in English. *Proceedings of the 14th International Congress of Phonetic Sciences*, 1433–1436.

Rodgers, T. S. (1969). On measuring vocabulary difficulty: an analysis of item variables in learning Russian–English vocabulary pairs. *Review of Applied Linguistics* 7, 327–343.

Scovel, T. (1988). *A time to speak: a psycholinguistic inquiry into the critical period for human speech*. Cambridge: Newbury House.

Snodgrass, J., Levy-Berger, G. & Haydon, M. (1985). *Human experimental psychology*. Oxford: Oxford University Press.

Received August 17, 1998 Revision accepted December 28, 1999