Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Phonetics 33 (2005) 263–290
Journal of Phonetics
www.elsevier.com/locate/phonetics

A developmental study of English vowel production and perception by native Korean adults and children

Kimiko Tsukada[a,*], David Birdsong[b], Ellen Bialystok[c], Molly Mack[d], Hyekyung Sung[e], James Flege[f]

[a] *MARCS Auditory Laboratories, University of Western Sydney, Locked Bag 1797, Penrith South DC NSW 1797, Australia*
[b] *University of Texas at Austin, USA*
[c] *York University, Canada*
[d] *University of Illinois, USA*
[e] *Defense Language Institute, USA*
[f] *University of Alabama at Birmingham, USA*

Received 16 July 2003; received in revised form 16 August 2004; accepted 13 October 2004

Abstract

This study examined the production and perception of English vowels by native Korean (NK) learners of English on two occasions separated by about 1 year. A preliminary experiment revealed that NK adults classified some pairs of contrastive English vowels using two different Korean vowels whereas other pairs showed classification overlap, implying they would be difficult for Korean learners of English to discriminate. In two subsequent experiments, NK adults and children differing in length of residence in North America (3 vs. 5 years; 4 groups of 18 each) were compared to age-matched native English (NE) speakers. In Experiment 2, NK children were found to discriminate English vowels more accurately than NK adults but less accurately than NE children. In Experiment 3, English words containing /i ɪ eɪ ɛ æ ɑ ʌ/ were elicited using a picture-naming task. Some vowels produced by NK children were heard as intended significantly more often than vowels produced by NK adults. Acoustic analyses revealed that NK children produced significantly larger between-vowel contrasts than NK adults but did not differ from NE children.

*Corresponding author. Tel.: +61 2 9772 6230; fax: +61 2 9772 6326.
E-mail address: k.tsukada@uws.edu.au (K. Tsukada).

0095-4470/$ - see front matter
doi:10.1016/j.wocn.2004.10.002

Taken together, the results suggested that although children are more successful than adults in learning the phonetic properties of second-language vowels, they might continue to differ from age-matched native speakers in certain respects as uncovered by the vowel discrimination test in Experiment 2.

1. Introduction

The notion that "earlier is better" with respect to second language (L2) speech learning is well accepted (see Singleton, 1989 for a review). As discussed by Snow (1987, page 192), it is widely believed that children are able to learn an L2 "quickly, automatically, effortlessly, and to a level indistinguishable from that of native speakers" whereas L2 learning in adulthood is conceived as something that is "slow, effortful, and often less than perfectly successful". The purpose of the present study was to evaluate this view of adult–child differences in L2 speech learning. It did so by examining the discrimination and production of English vowels by native Korean (NK) adults and children who were learning English as an L2.

Most evidence that children are better able than adults to learn L2 speech is indirect, inasmuch as it comes from studies employing a retrospective developmental design in which comparisons are made between groups of adult participants differing in age of first exposure to an L2. Participants first exposed to their L2 as children ("early learners") typically show more native-like performance than participants first exposed to the L2 in late adolescence or adulthood ("late learners"). Significant early-late differences have been observed for the production and perception of L2 consonants (e.g., Flege, Munro, & MacKay, 1995; Yamada, 1995; Mackay, Flege, Piske, & Schirru, 2001), the production and perception of L2 vowels (e.g., Munro, Flege, & MacKay, 1996; Flege, MacKay, & Meador, 1999a; Piske, Flege, MacKay, & Meador, 2002; Flege, Schirru, & MacKay, 2003), and overall degree of foreign accent (e.g., Seliger, Krashen, & Ladefoged, 1975; Oyama, 1976; Yeni-Komshian, Flege, & Liu, 2000).

Most studies just mentioned compared groups of adult immigrants who differed in age of arrival (AOA) in a predominantly L2-speaking country. Given that chronological age was not controlled, length of residence (LOR) was necessarily confounded with the participants' AOA (see Flege, 1998, for review). Also, the early learners had generally lived in an L2-speaking country for many years at the time of testing. Thus, another limitation of a retrospective developmental design is that it provides little insight into adult–child differences in early stages of L2 learning.

More direct evidence of adult–child differences in L2 speech learning could be obtained by directly comparing adults to children. If children are really better able than adults to learn L2 speech, they should show greater improvement in performance than adults over time in a longitudinal study. An effect of chronological age might also be obtained by comparing groups of adults and children who differed in LOR. The traditional view of adult–child differences suggests that children would show a greater effect of apparent time (i.e., show a greater effect of LOR) than adults.

A review of the existing literature provides only partial support for the expectations just outlined. One of the few direct adult–child comparisons in the literature is that of Snow and Hoefnagel-Höhle (1978), who examined vowel discrimination and word production by native English (NE) adults and 3–15 year-old NE children who were learning Dutch as an L2. The NE adults showed somewhat more native-like performance than the NE children did after 6 months of residence in The Netherlands (T1), whereas the children tended to outperform the adults after 10–11 months of residence (T2).

The results obtained by Snow and Hoefnagel-Höhle (1977,1978) might be interpreted to mean that the NE children were more successful than the NE adults in learning L2 Dutch. However, there is reason for caution in reaching such a conclusion. The children were enrolled in Dutch public schools where they needed to use Dutch for everyday activities, whereas the adults seldom used Dutch because English is widely spoken by Dutch adults. The adult–child difference observed at T2 might therefore be attributed, at least in part, to a greater opportunity and/or need to use Dutch by the NE children than by the NE adults. Also, the discrimination and production of phonetic segments in the native language changes during childhood (e.g., Flege & Eefting, 1986; Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004). Given that the NE adults and NE children were not compared to age-matched groups of native Dutch participants, it was therefore uncertain if the NE children's performance was indeed more Dutch-like at T2 than that of the NE adults.

In keeping with the traditional view of adult–child differences, research examining the production and perception of L2 consonants has suggested that LOR may exert a somewhat stronger influence on L2 performance for children than adults. Yamada and colleagues (Yamada, Strange, Magnuson, Pruitt, & Clark, 1994; Yamada, 1995) observed a positive correlation between LOR and the production and perception of /r l/ by young native Japanese (NJ) adults. However, these results could not be interpreted with certainty because the participants' LOR was strongly correlated with their AOA in the US. Aoyama et al. (2004, under review) tested NJ adults and children after 0.5 years (T1) and 1.5 years (T2) of residence in the US. The NJ children but not adults showed a significant improvement in discriminating English consonants from T1 to T2. Both groups' production of English liquids improved over the 1-year study interval, but the NJ children showed more improvement than the NJ adults did (especially for /ɹ/), and only the NJ children showed an improved production of fricatives. It was noteworthy, however, that although the NJ children showed evidence of greater learning than the NJ adults, they nonetheless differed significantly from NE children.

Williams (1979) tested native Spanish children who differed in chronological age (8–10 vs. 14–16 years) and LOR in the US (< 0.5 vs. 1.5–2 vs. 3–3.5 years). The identification of synthetic labial stops differing in voice onset time (VOT) was significantly more English-like for children with an LOR greater than 1.5 years than for children having an LOR less than 0.5 years. The children with an LOR of 3–3.5 years tended to produce English /p/ with longer (and thus more English-like) VOT values than did children with an LOR less than 0.5 years. However, even the children who had lived in the US for more than 3 years produced /p/ with significantly shorter VOT values than NE children did.

Research examining overall degree of foreign accent has suggested that LOR effects may be greater for children than adults. Flege and Fletcher (1992) observed significantly milder foreign accents for native Spanish adults having an LOR in the US of 14 than 1 year, whereas Flege (1988) failed to observe a significant difference between groups of Chinese adults having mean LORs of 5 vs. 1 year. Oyama (1976) recorded young Italian adults who had arrived in the US by the age of 15 years. No difference was found to exist between participants having LORs of 5–11 and 12–18 years. In a study by Asher and García (1969), English sentences produced by native

Spanish children who arrived in the US prior to 13 years of age were classified "native", "near-native", "slight foreign accent", or "definite foreign accent" by NE-speaking listeners. None of the 56 children were judged to be "native". This finding, which agrees with the results of studies employing a retrospective developmental design (e.g., Guion, Flege, & Loftin, 2000; Piske, Flege, & MacKay, 2001; Yeni-Komshian et al., 2000), is not consistent with the traditional view of children's L2 speech learning ability. However, Asher and García (1969) obtained more "near-native" classifications for children having an LOR of 5–8 than 1–4 years. And, more recently, Aoyama et al. (2003) tested NJ adults and children 6–14 years of age after 0.5 and 1.5 years of residence in the US. These authors found that the NJ children but not the adults were less foreign-accented at T2 than T1.

The literature has not yet provided a comparison of the effect of LOR on adults' and children's L2 vowel production and perception. No previous study, to our knowledge, has assessed the effect of LOR for children. As for adults, Flege, Bohn, and Jang (1997) compared groups of nonnative adults differing in L1 (Korean, Mandarin, Spanish, German) and LOR in the US (overall *means* = 1 vs. 7 years). Tokens of English vowels (/i/, /ɪ/, /ɛ/, /æ/) produced by the participants with an LOR of 7 years were slightly more intelligible than vowels produced by the participants with an LOR of 1 year; however, the LOR effect reached significance only for /ɪ/. (Similar results were obtained in acoustic analyses.) Vowel perception was assessed by manipulating F1 frequency in synthetic continua ranging from /ɛ/-to-/æ/ and /i/-to-/ɪ/. In most instances, the effect of the F1 manipulation was greater for the participants having a relatively long than short LOR.

The present study compared the discrimination and production of English vowels by NK adults and children learning English in North America and age-matched groups of NE speakers. The NK participants had all migrated from Korea to either the US or Canada. All of them were living in predominantly English-speaking communities where English was needed for everyday interactions. Half of the 72 NK participants were adults and half were children. The NK adults and children, as well as the NE adults and children ($n = 18$ each), were tested on two occasions separated by about 1 year (T1, T2). To evaluate the effect of LOR, the NK adults and children were subdivided into groups differing in LOR, yielding four NK groups of 18 each.

The effect of chronological age on L2 speech learning was evaluated in three ways. First, NK groups differing in LOR (3 vs. 5 years) were compared. The traditional view of children's L2 learning ability leads to the expectation that LOR differences would exert a greater effect for the NK children than for the NK adults. Second, performance at T1 and T2 was compared. The traditional view leads to the expectation that if the NK children had not already achieved a native-like levels of performance at T1, they would show a greater improvement from T1 to T2 than the NK adults would. Finally, the two groups of NK children and the two groups of NK adults were compared to the NE children and NE adults, respectively. The traditional view leads to the expectation that the NK adults but not children would differ from age-matched NE controls, regardless of time of testing or LOR.

The study was organized as follows. Naturally produced English vowels were classified by NK adults in Experiment 1 in order to determine which pairs of contrastive English vowels might prove difficult for NK participants to discriminate. Experiment 2 tested the discrimination of five English vowel contrasts by the NK and NE participants. The Experiment 1 results suggested that the NK participants would have difficulty discriminating /i/-/ɪ/, /eᴵ/-/ɛ/, /ɛ/-/æ/ and /ɑ/-/ʌ/ but not

/i/-/ɑ/ (because these last vowels would likely be heard as distinct Korean vowels by NK speakers). A picture-naming task was used in Experiment 3 to elicit the production of English words. Vowel production accuracy was assessed through phonetic classification and acoustic analysis.

2. Experiment 1

The purpose of this experiment was to identify four pairs of contrastive English vowels that would likely be difficult for the NK participants tested in Experiment 2 to discriminate, and one other pair of English vowels that would be relatively easy for them to discriminate and thus could serve as a control to ensure that the Experiment 2 discrimination task was understood.

It is generally agreed that nonnatives' discrimination of L2 vowels depends importantly on how the L2 vowels are classified with respect to vowels in their L1. For example, according to Best's Perceptual Assimilation Model (Best, 1993, 1995), instances of contrastive L2 vowel categories that are identified as instances of a single L1 vowel category will be relatively difficult to discriminate whereas instances of contrastive L2 vowels that are mapped onto two different L1 vowels will be discriminated more accurately. The aim of this experiment, therefore, was to determine how the English vowel tokens used in Experiment 2 would be classified in terms of Korean vowel categories.

Korean is usually analyzed as having either eight vowels, /i e ɛ a ʌ o ɨ u/ (Lee & Zhi, 1987; Hong, 1988; Lee, 1993; Magen & Blumstein, 1993; Kang, 1996; Park, 1997) or ten vowels (if /y/ and /ø/ are included; see Yang, 1992, 1996; Sohn, 1999; Lee & Ramsey, 2000). The phonemic status of Korean /e/ and /ɛ/ depends on age (Hong, 1988; Park, 1997; Lee & Ramsey, 2000). Many Koreans have only one phoneme in the mid-front region of the vowel space due to the recent merger of /e/ and /ɛ/ (caused by the raising of /ɛ/, see Kang, 1996, and Lee & Ramsey, 2000).

Another on-going sound change in Korean pertains to vowel duration. Korean is traditionally regarded as having phonological length distinctions (Lee & Zhi, 1987; Sohn, 1999; Lee & Ramsey, 2000), but such distinctions are disappearing (Magen & Blumstein, 1993; Park, 1997; Sohn, 1999; Lee & Ramsey, 2000). This may increase the difficulty that NK learners of English have in discriminating certain pairs of English vowels such as /i/ and /ɪ/ because such vowels differ in duration as well as vowel quality.

2.1. Method

2.1.1. Stimuli

The English vowels examined here (and subsequently in Experiment 2) were produced by adult male NE speakers. Eight NE speakers produced /i ɪ eᶦ ɛ æ ɑ ʌ u/ in a /b_bo/ context within a carrier phrase. The /bVbo/ nonwords were digitized (22.05 kHz, 16-bit resolution) and normalized for peak amplitude. All portions of the speech signal preceding release of the initial /b/ and following complete constriction of the intervocalic /b/ were removed to control for unwanted variation in portions of the stimuli other than the vowels of interest. The resulting /bVb/ stimuli were then evaluated by four NE-speaking listeners (2 male and 2 female), who rated each token for goodness using a 5-point scale. The tokens selected as stimuli were the five tokens of each vowel receiving the highest ratings. The selected tokens of all eight target vowel categories were produced by five different speakers.

Table 1 presents the mean duration as well as the mean first and second formant frequencies (F_1, F_2 at the vowel midpoint) of the English vowels examined here. It is noteworthy that three pairs of vowels that were adjacent in vowel space (/i/-/ɪ/, /ɛ/-/æ/, /ɑ/-/ʌ/) differed in terms of duration, and that one other pair of vowels that were adjacent in terms of midpoint formant F1–F2 values, /eɪ/-/ɛ/, differed in terms of degree of formant movement, as well (which is not shown). Acoustic values have been published for Korean vowels (see, e.g., Yang, 1996), but not for vowels produced in the context examined here (/b_b/), making direct comparison inappropriate (see, e.g., Strange, Verbrugge, Shankweiler, & Edman, 1976).

2.1.2. Listeners

NK speakers (3 male, 5 female) with a mean age of 31.4 years (*range* = 24–38 years) who had lived only briefly in an English-speaking country (*range* = 6–11 months) participated. All reported having normal hearing, and all had studied English in Korea (*range* = 6–10 years).

2.1.3. Procedure

The participants were tested individually in a quiet room after completing a language background questionnaire. Written instructions in Korean were given. The 40 stimuli (8 English categories × 5 tokens) described earlier were randomly presented five times each via headphones at a self-selected comfortable level. The participants identified the stimuli in terms of one of the eight vowels of Korean (/i e ɛ a ʌ o ɨ u/) by clicking Hangul characters displayed on the computer screen. The interval between each response and presentation of the next stimulus was 1 s. A response was required for each stimulus. The participants were told to guess if uncertain. They could replay a stimulus but not change a response once given. Twenty extra tokens presented for familiarization at the beginning of the experiment were not analyzed.

Table 1
Acoustic characteristics of the stimuli used in Experiment 1

Stimuli	Duration	F1	F2
/i/	113 (9.7)	285 (25.7)	2288 (204.8)
/ɪ/	99 (17.7)	442 (17.2)	1696 (117.0)
/eɪ/	156 (15.0)	457 (27.5)	1890 (150.4)
/ɛ/	115 (16.9)	549 (26.4)	1575 (48.8)
/æ/	157 (27.8)	701 (69.1)	1554 (53.7)
/ɑ/	161 (21.0)	722 (53.8)	1148 (51.5)
/ʌ/	98 (15.1)	574 (16.6)	1199 (63.3)
/u/	131 (24.4)	315 (16.4)	1228 (92.5)

Note: Duration in ms. F1 and F2 in Hz. Standard deviations are in parentheses.

2.2. Results and discussion

Table 2 shows the percentages of times, based on a total of 200 judgments (8 listeners × 5 tokens × 5 replicate judgments), that tokens of the eight English target vowels were classified as an instance of each Korean vowel category. The results obtained here led us to infer that the NK participants in Experiment 2 would have little or no difficulty in discriminating the /i/ stimuli from the /ɑ/ stimuli, whereas they would experience some degree of difficulty in discriminating /i/-/ɪ/, /eɪ/-/ɛ/, /ɛ/-/æ/ and /ɑ/-/ʌ/.

The /i/ and /ɑ/ tokens were unambiguously identified as instances of two different Korean vowels, /i/ and /a/. This finding suggested that even though the English /i/ and /ɑ/ stimuli might differ phonetically from Korean /i/ and /a/, NK learners of English will generate two different phonetic codes when hearing English /i/ and /ɑ/. Accordingly, /i/-/ɑ/ served as a control contrast in Experiment 2 to verify that all participants understood the discrimination task and possessed a working memory that was adequate for the task.

An inspection of the results in Table 2 suggested that the NK participants in Experiment 2 might have difficulty discriminating /i/-/ɪ/, /ɛ/-/æ/, and /ɑ/-/ʌ/. The instances of one English vowel in each of these pairs was consistently classified as a single Korean vowel. The "consistent" members of these pairs, /i/, /æ/ and /ɑ/, were classified in 86–100% of instances as Korean /i/, /a/ and /a/, respectively. The other members of these pairs—/ɪ/, /ɛ/ and /ʌ/, respectively—were classified in terms of multiple Korean vowels, including the Korean vowel associated with the "consistent" member of the contrast. For example, English /i/ was always classified as Korean /i/ while English /ɪ/ was identified as Korean /i/ (32%) and also Korean /e/ (30%), /ɛ/ (18%) and /ɨ/ (12%).

Guion, Flege, Akahane-Yamada, and Pruitt (2000) suggested that the "overlap" of classifications in a perception assimilation experiment (i.e., how often a single L1 category was used for instances of contrasting L2 categories) might predict L2 discrimination accuracy (see also Flege & MacKay, 2004). If so, then one might expect Korean learners of English to have little to

Table 2
The mean percentage of classifications of English vowels as one of the eight vowels of Korean in Experiment 1

	Response category							
English stimuli	이 /i/	에 /e/	애 /ɛ/	아 /a/	어 /ʌ/	오 /o/	으 /ɨ/	우 /u/
/i/	**100**							
/ɪ/	**32**	30	18		4		12	5
/eɪ/	3	43	**55**					
/ɛ/		**27**	**27**	22	24			
/æ/			6	**86**	6			
/ɑ/				**98**				
/ʌ/				**56**	35	6	4	
/u/							38	**62**

Note: Percentages less than 2% are not shown. Each percentage is based on 200 (8 listeners × 5 tokens × 5 repetitions) judgments. Modal response categories are boldfaced.

moderate difficulty in discriminating /i/-/ɪ/ and /ɛ/-/æ/ in Experiment 2. However, the partial overlap responses seen here, as well as the inconsistency of responses for one member of each contrast, suggested that /i/-/ɪ/ and /ɛ/-/æ/ would be more difficult to discriminate than the control contrast, /i/-/ɑ/. Support for this inference was provided by Polka (1995), who reported that NE adults had difficulty in discriminating a German vowel contrast that included a vowel they were unable to relate consistently to an English vowel category. Also, previous studies have provided evidence that NK adults have difficulty in identifying and/or classifying English /i/-/ɪ/ and /ɛ/-/æ/ (Flege et al., 1997; Ingram & Park, 1997).[1]

The NK participants examined here did not classify either the /eɪ/ or the /ɛ/ tokens consistently in terms of a single Korean vowel. Both vowels were mapped onto multiple Korean vowel categories with a high degree of overlap, leading to the expectation of at least some discrimination difficulty.

3. Experiment 2

The purpose of this experiment was to compare the discrimination of English vowels by NK adults and children to that of age-matched NE adults and children. It examined the discrimination of four English vowel contrasts (/i/-/ɪ/, /eɪ/-/ɛ/, /ɛ/-/æ/, /ɑ/-/ʌ/) that were likely to be difficult for Koreans to discriminate, and one other contrast (/i/-/ɑ/) that was likely to be easy for them to discriminate. The stimuli were the naturally produced /bVb/ tokens examined in Experiment 1, all of which were judged by NE-speaking listeners as "good" instances of their intended category. The participants were tested twice with an average 1.2 year-interval between the two times of testing (T1, T2).

3.1. Method

3.1.1. Participants

A total of 108 participants were assigned to one of six groups ($n = 18$ each). Seventy-two participants were NK speakers who had immigrated to North America from Korea; the remaining 36 were NE speakers. As summarized in Table 3, half of the NK and NE participants were adults and half were children. The participants were recruited and tested at five different locations in North America (25 at the University of Alabama at Birmingham; 22 at Stanford University in Palo Alto, California; 22 at the University of Illinois in Champaign-Urbana; 21 at the University of Texas at Austin; and 18 at York University in Toronto, Ontario) because it was impossible to find all of the participants meeting our selection criteria at a single location.

The NK participants were selected on the basis of chronological age and LOR. The NK children were required to be between 9 and 17 years of age at T1, and the NK adults were required to be between the ages of 23–41 years. Half of the NK participants had been living in North

[1] The perceptual assimilation results obtained differed somewhat from the results obtained by Baker et al. (2002) and Trofimovich, Baker, and Mack (2001), probably due to a difference in the phonetic contexts in which the vowel stimuli occurred, the participants, or both. Unlike the results obtained here, the NK participants tested previously (Trofimovich et al., 2001 ; Baker et al., 2002) identified vowels making up the /ɑ /-/ʌ / contrast in terms of two different Korean vowels, suggesting little discrimination difficulty.

Table 3
Characteristics (*mean*, *SD*, *range*) of the six groups of participants in Experiments 2 and 3

Group	*N*	Age	LOR	AOA	% English
Child-NE	7 m, 11f	12.7 (2.5) 8–17	—	—	—
Child-3	6 m, 12f	12.3 (2.4) 9–17	2.9 (0.4) 2–4	9.4 (2.4) 6–14	58 (15) 30–80
Child-5	7 m, 11f	13.7 (2.0) 10–17	4.9 (0.6) 4–6	8.9 (2.1) 6–13	58 (14) 40–80
Adult-NE	7 m, 11f	32.3 (4.4) 26–41	—	—	—
Adult-3	8 m, 10f	30.4 (5.1) 23–40	3.0 (0.6) 2–4	27.4 (5.0) 21–38	41 (19) 15–80
Adult-5	5 m, 13f	33.1 (5.3) 27–41	4.8 (0.6) 4–6	28.3 (5.2) 22–36	40 (20) 10–70

Note: NE = native English; m = male; f = female; Age = chronological age, in years; LOR = length of residence in North America at the first time of testing, in years; AOA = age of arrival in North America, in years; % English = overall self-reported percentage use of English. These estimates have been averaged over two testing times (T1 and T2) because the T1 and T2 estimates were correlated; $p<0.01$.

America for 2–4 years at T1, and the remaining half for 4–6 years. The NK children with an LOR of 2–4 and 4–6 years were designated the "child-3" and "child-5" groups, respectively; the NK adult groups having these LORs were designated "adult-3" and "adult-5". The "child-NE" and "adult-NE" groups were matched in chronological age to the NK children and adults, respectively.[2]

The NK children and adults differed in education and English use as well the two selection criteria. All 36 NK adults had studied English at school in Korea before arriving in North America (*mean* = 9 years, *range* = 5–13), but just seven NK children had done so, and then only briefly (*range*: 1–3 years). All 36 NK children were attending an English-speaking school in North America when tested at T1. Few of the NK adults were attending school, but 26 of them had received at least some university education in North America (*mean* = 6 years, *range* = 1–9). In general, the NK children reported using English more often than the NK adults in various situations, in particular, with friends.

3.1.2. Procedure

The participants' perception of English vowels was assessed using a categorial discrimination test employed in previous L2 research (Flege et al., 1999a; Guion et al., 2000; Flege, 2003a; Flege & MacKay, 2004). The vowel stimuli, which were described in Experiment 1, were presented in

[2]It would have been desirable to recruit groups of NK children who were younger on average, and who represented a narrower age range, but this was not possible given the LOR requirements and the desire to recruit individuals who had arrived in North America at or after the normal age of entry into school (6 years) but prior to the age that is claimed by some investigators to mark the end of a "critical period" for L2 speech learning (12 years according to Scovel, 1988; 15 years according to Patkowski, 1990).

triads via headphones at a comfortable level using a notebook computer. Each vowel contrast of interest was tested by eight change and eight no-change trials. The results of Experiment 1 suggested that, for many NK learners of English, the two vowels making up the /i/-/ɪ/, /eɪ/-/ɛ/, /ɛ/-/æ/ and /ɑ/-/ʌ/ contrasts would not reliably map onto two distinct Korean vowel categories. A premise underlying this study was that a high level of performance for these contrasts would require the establishment of new phonetic categories for one or both English vowels.

The inclusion of both change and no-change trials in the discrimination test used here was motivated by the view that establishing a phonetic category will increase sensitivity to differences between instances of the newly formed category and other L1 and L2 categories, and will also reduce sensitivity to token-to-token variation *within* the newly formed category (see e.g., Liberman, 1957; Kuhl, 1980; Guenther, Husain, Cohen, & Shinn-Cunningham, 1999; Bosch et al., 2000). The three vowel tokens presented in all change and no-change trials were spoken by different talkers, and so were always physically, if not phonetically, different. The change trials all contained an odd item out; these trials tested participants' ability to distinguish vowels drawn from two different categories. For example, a change trial testing /ɛ/-/æ/ might consist of $/\text{ɛ}/_2$ $/\text{æ}/_1$ $/\text{ɛ}/_5$ (where the subscripts indicate different talkers). The correct response for change trials was the button ("1", "2", or "3") indicating the position of the odd item out, which occurred with near-equal frequency in all three possible serial positions. The no-change trials, which contained three different instances of a single category (e.g., $/\text{æ}/_3$ $/\text{æ}/_5$ $/\text{æ}/_2$ or $/\text{ɛ}/_1$ $/\text{ɛ}/_5$ $/\text{ɛ}/_4$), tested the participants' ability to ignore audible but phonetically irrelevant within-category variation (e.g., in voice quality). The correct response to no-change trials was a fourth button marked "NO".

To reduce the use of purely auditory information, the change and no-change trials testing all five contrasts were presented in a single randomized block. The participants were required to respond to each trial, and were told to guess if uncertain. A trial could be replayed, but responses could not be changed once given. The inter-stimulus interval between the stimuli making up each trial was 0.8 s and the interval between each response and the next trial was 1.0 s.

The 108 participants were tested individually in a sound booth or quiet room. A practice session with feedback was carried out using /bub/ and /bɛb/ stimuli prior to the experiment proper. Participants were required to respond correctly to at least nine of 10 practice trials before proceeding to the experiment; all were able to do so within three presentations of the 10-item practice block. The participants were told that feedback would not be provided during the experiment. Five additional practice items, which were not analyzed, were presented at the beginning of the experiment to acclimate the participants to the withdrawal of feedback.

In order to reduce possible effects of response bias, the dependent variable examined in all analyses were A' scores based on responses to both change and no-change trials (see Snodgrass, Levy-Berger, & Haydon, 1985). These scores were based on the proportion of "hits" obtained for each contrast (i.e., the number of correct selections of the odd item out in 8 change trials) and the proportion of "false alarms" (i.e., the number of incorrect selections of an odd item out in 8 catch trials).[3] 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 $A' = 0.5 + ((H-FA)*(1+H-FA))/((4*H)*(1-FA))$. However, if

[3]Four catch trials testing each contrast contained three physically different instances of one vowel, and the remaining four catch trials contained three different tokens of the other vowel category.

FA exceeded H, then A' = 0.5−((FA−H)*(1 + FA−H))/((4*FA)*(1−H)). A score of 1.0 indicated perfect sensitivity, whereas a score of 0.5 indicated a lack of phonetic sensitivity.

3.2. Results

3.2.1. Establishing experimental control

The results of Experiment 1 suggested that the NK participants, even those who had not established new categories for English vowels, would generate two distinct phonetic codes when listening to the vowel stimuli used to test the English /i/-/ɑ/ contrast. Thus, all 108 participants were expected to obtain high scores for this contrast provided they understood the instructions and possessed an adequate working memory.

A preliminary ANOVA examining the scores for all six groups revealed that the effect of Testing Time was non-significant [$F(1,102) = 3.8$, *ns*] and did not enter into significant interactions with other factors. The scores shown in Fig. 1 have therefore been averaged over T1 and T2. Here it can be seen that all six groups of participants obtained high scores for /i/-/ɑ/. The scores for this contrast range from a low of 0.96 for child-3 to 0.99 for adult-NE. Also as expected, the NK participants obtained lower scores than the NE participants for the remaining four contrasts. The adult-3 and adult-5 groups obtained the lowest scores overall. The child-3 and child-5 groups obtained scores that were intermediate to the scores obtained by the NK adults, on the one hand, and the two NE groups (child-NE, adult-NE), on the other hand.

In order to determine if experimental control was established, scores averaged over the two times of testing were submitted to an ANOVA in which Group (6 levels) was a between-subjects factor and Contrast (5 levels) served as a within-subjects factor. Both main effects as well as the two-way interaction reached significance [Group $F(5,102) = 40.1$; Contrast $F(4,408) = 94.5$; G × C $F(20,408) = 8.2$; all p-values <0.01]. The simple effect of Group was non-significant for

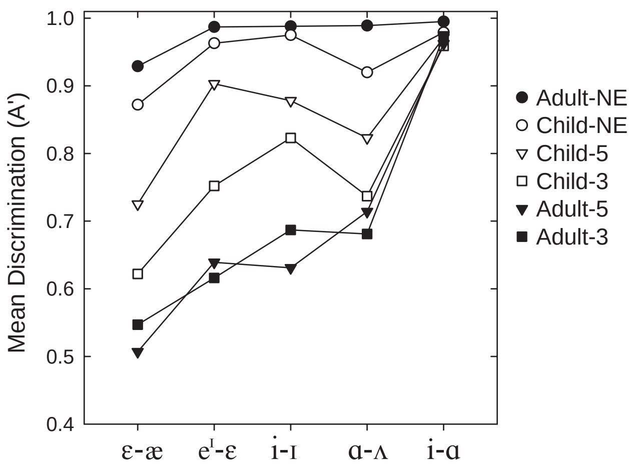

Fig. 1. Mean discrimination scores obtained for five vowel contrasts in Experiment 2. The NE and NK participants differed in age (child vs. adult); the NK participants differed in age and length of residence (3 vs. 5 years) in North America. The scores have been averaged over two testing times (T1 and T2).

/i/-/ɑ/ [$F(5,102) = 2.4$, *ns*], but significant for the four test contrasts [/i/-/ɪ/ $F(5,102) = 25.3$; /e$^{\text{ɪ}}$/-/ɛ/ $F(5,102) = 23.8$; /ɛ/-/æ/ $F(5,102) = 25.9$; /ɑ/-/ʌ/ $F(5,102) = 14.5$, Bonferroni adjusted $p < 0.05$]. The relatively high scores obtained for /i/-/ɑ/ and the lack of a significant effect of Group for this contrast confirmed that experimental control had indeed been established for all six groups, including both groups of NK children. The results for /i/-/ɑ/ will therefore not be discussed further.

3.2.2. Effects of age and LOR

The scores obtained by the NK participants for the four English contrasts (/i/-/ɪ/, /e$^{\text{ɪ}}$/-/ɛ/, /ɛ/-/æ/, /ɑ/-/ʌ/) were submitted to an ANOVA in which age (child vs. adult) and LOR (3 vs. 5 years) served as between-subjects factors and Testing Time (T1 vs. T2) and Contrast (4 levels) served as within-subjects factors. As in the preliminary analysis, the main effect of Testing Time was non-significant [$F(1,68) = 2.7$, *ns*] and did not interact significantly with any other factor. The main effect of LOR was non-significant [$F(1,68) = 2.6$, *ns*], as were the LOR × Contrast [$F(3,204) = 1.4$, *ns*] and LOR × Age × Contrast interactions [$F(3,204) = 0.4$, *ns*].

Fig. 2(a) shows that the NK children obtained higher scores than the NK adults did, but that child–adult differences were greater for participants with an LOR of 5 than 3 years, resulting in a significant Age × LOR interaction [$F(1,68) = 3.9$, $p = 0.05$]. Simple effects tests revealed that the NK children with an LOR of 5 years obtained higher scores than those with an LOR of 3 years [$F(1,34) = 4.6$, $p < 0.05$] whereas the effect of LOR was non-significant for the NK adults [$F(1,34) = 0.1$, *ns*]. Scores were higher for the NK children than adults, both those having an LOR of 3 years [$F(1,34) = 7.4$, $p < 0.05$] and 5 years [$F(1,34) = 26.6$, $p < 0.01$].

Fig. 2(b) shows that the NK children obtained higher scores than the NK adults did for all four English vowel contrasts. The ANOVA yielded a significant main effect of Age [$F(1,68) = 31.8$, $p < 0.01$] as well as a significant Age × Contrast interaction [$F(3,204) = 3.0$, $p < 0.05$]. Simple

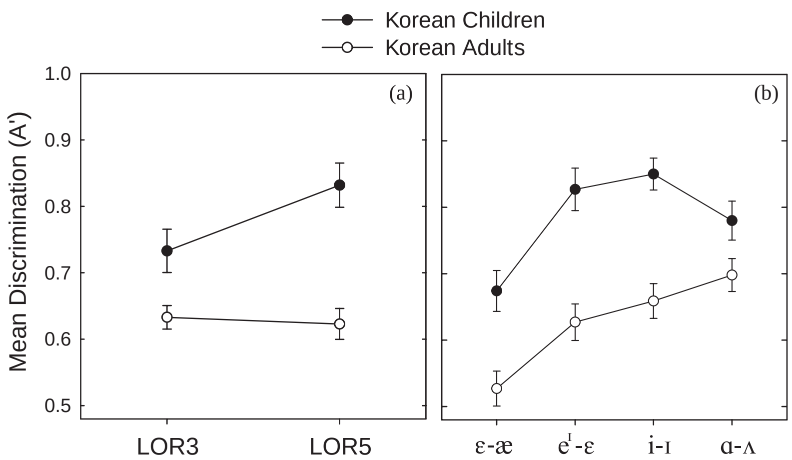

Fig. 2. Interactions obtained in the analyses of scores for native Korean children and adults in Experiment 2. The scores have been averaged over two testing times (T1 and T2). The error bars enclose ± one standard error.

effects tests revealed that the interaction arose because the child–adult difference was significant for only three of the four contrasts [/i/-/ɪ/ $F(1,70) = 29.1$; /eᴵ/-/ɛ/ $F(1,70) = 22.5$; /ɛ/-/æ/ $F(1,70) = 13.0$, Bonferroni $p < 0.05$, but not /ɑ/-/ʌ/ $F(1,70) = 4.6$, *ns*]. The simple effect of Contrast was significant for both age groups [children $F(3,105) = 18.0$, $p < 0.01$; adults $F(3,105) = 8.4$, $p < 0.01$]. Tukey's tests revealed that both age groups obtained lower scores for /ɛ/-/æ/ than for the other three contrasts ($p < 0.05$), but only the NK children obtained lower scores for /ɑ/-/ʌ/ than /i/-/ɪ/ ($p < 0.05$).

Given the assumption that between-contrast differences for the NK participants arose from differences in how various English vowel stimuli mapped onto long-term memory representations for Korean vowels (e.g., Best, 1993, 1995), one might have expected the same pattern of between-contrast differences for the NK adults and children. To help understand the apparent discrepancy involving /ɑ/-/ʌ/, the NE participants' scores were submitted to an Age (child vs. adult) × Contrast (/i/-/ɪ/, /eᴵ/-/ɛ/ /ɛ/-/æ/, /ɑ/-/ʌ/) ANOVA. Both main effects reached significance [Age $F(1,34) = 10.3$, $p < 0.01$; Contrast $F(3,102) = 29.8$, $p < 0.01$] as well as the two-way interaction [$F(3,102) = 3.8$, $p < 0.05$]. Simple effects tests revealed that the NE children obtained lower scores than the NE adults for /ɑ/-/ʌ/ [$F(1,34) = 10.9$, Bonferroni $p < 0.05$] but not for the remaining three contrasts [$F(1,34) = 1.2$–6.4, *ns*][4] (see also Baker, Trofimovich, Mack, & Flege, 2002). This suggests that the ability of English monolinguals to discriminate /ɑ/-/ʌ/ develops slowly, either because the auditory difference between /ɑ/ and /ʌ/ is less robust than the difference between other pairs of English vowels, because input for /ɑ/ and/or /ʌ/ is more variable than for other vowels, or because relatively few words are distinguished by /ɑ/-/ʌ/. Whatever the explanation, the NE participants' results suggest that some factor(s) other than cross-language phonetic interference might have influenced the NK children's performance for /ɑ/-/ʌ/.

3.2.3. Native-nonnative differences

To determine which NK groups differed from age-matched NE speakers, one-way ANOVAs were carried out to test for differences between the three groups of children (child-3, child-5, child-NE) and the three groups of adults (adult-3, adult-5, adult-NE). Separate analyses were carried out for each contrast. As summarized in Table 4, the effect of Group was significant in all eight (2 age groups × 4 contrasts) ANOVAs (Bonferroni $p < 0.05$). Tukey's tests indicated that, for all four contrasts, higher scores were obtained for the NE adults than for both groups of NK adults (adult-3, adult-5), who did not differ from each other. The child-NE group obtained significantly higher scores than the child-3 group did for all four test contrasts, but higher scores than the child-5 group for just two contrasts (/i/-/ɪ/, /ɛ/-/æ/). Also, the child-5 group obtained significantly higher scores than the child-3 group did for /eᴵ/-/ɛ/.

[4] Although the Time effect never reached significance in preceding analyses, given the unexpected pattern of results just for the /ɑ /-/ʌ / contrast, we reexamined the discrimination scores for the NE participants. The NE children's scores showed greater improvement from T1 (*mean* = 0.90) to T2 (*mean* = 0.94) than the NE adults' (*means* = 0.98 at T1 and 0.99 at T2). However, an Age (child vs. adult) × Time (T1 vs. T2) ANOVA yielded significant effects only for the main effects [Age $F(1,34) = 10.9$, $p < 0.01$; Time $F(1,34) = 4.2$, $p < 0.05$]. A lack of significant two-way interaction [$F(1,34) = 1.9$, *ns*] suggests that the effect of Time was not unique to the NE children.

Table 4
Summary of one-way ANOVAs testing for native-nonnative differences for four vowel contrasts in Experiment 2

	Children		Adults	
	$F(2, 51)$	Tukey's test	$F(2, 51)$	Tukey's test
/i/-/ɪ/	7.4*	NE > LOR-5, LOR-3	40.5*	NE > LOR-5, LOR-3
/eɪ/-/ɛ/	9.8*	NE, LOR-5 > LOR-3	41.6*	NE > LOR-5, LOR-3
/ɛ/-/æ/	12.0*	NE > LOR-5, LOR-3	55.1*	NE > LOR-5, LOR-3
/ɑ/-/ʌ/	6.8*	NE > LOR-3	34.0*	NE > LOR-5, LOR-3

Note: The discrimination scores have been averaged over two testing times (T1 and T2).
*Indicates significance at a Bonferroni-adjusted level of 0.05.

3.3. Discussion

The results obtained here showed that the NK children were better able to discriminate English vowels than the NK adults. This finding, which agrees with the results of studies examining the discrimination of L2 vowels by groups of adults differing in age of exposure to the L2 (Flege et al., 1999a; Flege & MacKay, 2004), is consistent with the traditional view of children's superior ability to learn L2 speech. However, the NK children with an LOR of 3 years discriminated English vowels less accurately than the age-matched NE children did. This finding does not agree with the traditional view of adult–child differences, but it does agree with the results of studies showing differences in the perception and processing of Catalan vowels by native speakers of Catalan and native Spanish adults who were first exposed to Catalan as young children (Pallier, Bosch, & Sebastián-Gallés, 1997; Sebastián-Gallés & Soto-Faraco, 1999; Bosch et al., 2000; see also Flege & MacKay, 2004).

The most important finding of this experiment was that the NK children with an LOR of 3 years obtained significantly lower scores than the NE children did for all four English vowel contrasts, whereas the NK children with an LOR of 5 years differed significantly from the NE children for just two of these contrasts. This finding, when taken together with the significant effect of LOR for the NK children, suggested that the NK children were in the process of learning to perceive English vowels in a more native-like way. This evidence does not support the view that children's phonological representations become attuned to the language-specific characteristics of their L1 by about the age of 2 years and will thereafter be "relatively unaffected by later acquisitions in either the same or a different language" (Peperkamp & Dupoux, 2002, page 204; see also Sebastián-Gallés & Soto-Faraco, 1999, page 120).

4. Experiment 3

This experiment examined the production of the English vowels /i ɪ eɪ ɛ æ ɑ ʌ/ by the same individuals who participated in Experiment 2. English words containing the vowels of interest were elicited prior to Experiment 2 in order to prevent exposure to the perceptual stimuli from

Table 5
The words produced by the 108 participants in Experiment 3

Vowel	Words
/i/	Eat, feet, key
/ɪ/	Fish, pig, six
/eɪ/	Cage, date, eight
/ɛ/	Bed, head, neck
/æ/	Bat, cat, hat
/ɑ/	Dog, pot, sock
/ʌ/	Duck, cup, bug

affecting vowel production.[5] The participants' vowel production was assessed via phonetic classification and acoustic analysis.

4.1. Speech elicitation

Twenty-one words were elicited three times each using a picture-naming task. Care was taken to select picturable words that were likely to be familiar to all of the participants, including the NK children. Achieving these criteria resulted in variation in consonantal context for the seven target vowels (see Table 5).

A digitized token of each of the target words that had been spoken by an adult male NE speaker was used as an auditory prompt during word elicitation. The auditory prompts were presented via a loudspeaker as the picture representing each word appeared, in a random order, on the screen of a notebook computer. The first elicitation of the 21 target words will therefore be referred to as "cued" productions. The target words were auditorily cued during the second and third elicitations of the test words only when the participants could not say a target word after seeing its picture. The words were recorded using a Sony (model TCD-D8) DAT recorder, digitized (22.05 kHz, 16-bit resolution) and then normalized for peak amplitude.

The decision was made to examine just the first elicitation of the target words because variation existed as to whether the test words were or were not cued during the second and third elicitations. We cannot rule out the possibility that the presence of a native-produced model augmented the accuracy with which the NK participants produced the vowels examined here. It is possible, also, that the NK children benefited more from the auditory prompts than the NK adults did. This might hold true, for example, if children exhibit a greater tendency (or ability) to imitate than adults do.

4.2. Phonetic classification

Vowels in the 4536 words (108 participants $\times$ 7 target vowels $\times$ 3 words $\times$ 2 testing times) were classified as one of the 11 simple vowels of American English (/i ɪ eɪ ɛ æ ɑ oʊ ʌ u ʊ ɚ/) by an NE

[5]Sentences evaluated for foreign accent by Flege et al. (in press) were elicited after the words, but before the discrimination experiment.

adult with training in phonetics. The use of a single listener necessarily limits the generalizability of the results obtained here. However, to reduce the possibility of bias in the classifications, the vowels to be judged were presented in a random order to the listener, who was thus unaware of the identity of the speakers.

The 216 tokens of each word (108 participants × 2 testing times) were randomly presented twice in separate blocks. The listener's two replicate classifications of each vowel token agreed in 87% of instances, ranging from 81% for /ɪ/ to 99.5% for /eɪ/. Vowel tokens that did not receive the same classification were randomly presented three times each to the same listener in a subsequent session. The classification receiving the majority of responses in the second session was the one used in all analyses.

Confusion matrixes were created for the vowels spoken by each group at T1 and T2. The vowels spoken by each group could be heard as intended a maximum of 54 times (18 participants × 3 words). Of primary interest was how often the target vowels were heard as intended ("intelligibility"). As shown in Table 6, much the same average intelligibility scores were obtained for the NE children (*mean* = 84% correct averaged over the two times of testing) and the two groups of NK children (*mean* = 83% correct, averaged over the two times of testing and LOR). The NK children's /ɪ/s were misheard as /ɛ/ in 19% of instances, their /ɛ/s as /æ/ in 39% of instances, and their /ʌ/s as /ɑ/ in 33% of instances, on the average. The NE children's /ɪ/s were also heard as /ɛ/ (*mean* = 27%) and their /ɛ/s were heard as /æ/ (*mean* = 45%), but their /ʌ/s were seldom heard as /ɑ/ (*mean* = 6%).

As shown in Table 7, intelligibility scores were higher for the NE adults (*mean* = 94% correct) than for the NK adults with an LOR of 5 years (*mean* = 66%) and those with an LOR of 3 years (*mean* = 69%). The NK adults' /ɛ/s and /ʌ/s were misclassified in much the same way as vowels spoken by the NK children. That is, the NK adults' /ɛ/s were heard as /æ/ (*mean* = 66%) and their /ʌ/s were heard as /ɑ/ (*mean* = 82%). However, the NK adults differed from the NK children for high vowels. The NK adults' /i/s were heard as /ɪ/ in 37% of instances and their /ɪ/s as /i/ in 17% of instances whereas the NK children's /ɪ/ tokens were rarely misperceived as /i/. Flege et al. (1997) reported a bi-directional confusion between /ɛ/ and /æ/ by 20 NK adults living in the US. The error rate for adults in that study was lower than that reported here for /ɛ/ (*mean* = 19%), but higher than that reported here for /æ/ (*mean* = 70%; see also Park, 1997). Note that the NE participants' /ɛ/s (especially those of the NE children) were often misheard as /æ/. The /ɛ/-/æ/ confusion may arise from a partial merger of these vowels in the speech of certain native speakers of American English (see acoustic measures at vowel midpoint reported by Hillenbrand, Getty, Clark, & Wheeler, 1995). We cannot exclude the possibility, however, that these confusions were the result of the perceptual standards of the single NE-speaking listener who classified all tokens, or a difference in the degree of articulatory difficulty in producing a contrast between /ɛ/-/æ/ as compared to contrasts between other pairs of English vowels.

Two analyses of the classification data were undertaken. The aim of the first analysis was to determine if native-nonnative differences existed in the intelligibility scores obtained for the seven target vowels. The number of times that each target vowel was heard as intended was determined. The 14 intelligibility scores obtained for each participant (7 vowels × 2 testing times) ranged from 0 when no vowel was heard as intended to 3 when the vowels in all three words were heard as intended.

The adults' scores were examined in 14 one-way ANOVAs, one for each of the seven vowels at both T1 and T2. The effect of Group (LOR-3, LOR-5, NE) reached significance for four vowels at

Table 6
Classifications of vowels produced by native English and native Korean children in Experiment 3. Each percentage is based on 54 tokens (18 speakers × 3 words). The percentages listed to the left and right of the slash are the values obtained at two testing times (T1 and T2). Vowels heard as intended are shown in boldface

Vowel said		Vowel heard						
		/i/	/ɪ/	/eɪ/	/ɛ/	/æ/	/ɑ/	/ʌ/
/i/	NE	**78/81**	9/6	13/13				
	LOR-5	**93/91**	6/0	2/9				
	LOR-3	**72/87**	17/4	11/9				
/ɪ/	NE	0/0	**63/67**	11/6	26/28			
	LOR-5	2/0	**74/69**	4/11	20/20			
	LOR-3	0/4	**82/74**	2/4	17/19			
/eɪ/	NE			**96/98**	4/2			
	LOR-5			**100/100**	0/0			
	LOR-3			**100/100**	0/0			
/ɛ/	NE				**59/52**	41/48		
	LOR-5				**61/54**	39/46		
	LOR-3				**56/72**	44/28		
/æ/	NE				0/0	**100/100**		
	LOR-5				2/2	**98/98**		
	LOR-3				0/4	**100/96**		
/ɑ/	NE						**94/94**	6/6
	LOR-5						**89/98**	11/2
	LOR-3						**93/96**	7/4
/ʌ/	NE						6/4	**94/96**
	LOR-5						28/35	**72/65**
	LOR-3						37/30	**63/70**

Note: NE = native English; LOR-5 = native Korean children with a length of residence in North America for 5 years at T1; LOR-3 = native Korean children with a length of residence in North America for 3 years at T1.

T1 [/i/ $F(2,51) = 13.8$; /ɪ/ $F(2,51) = 9.1$; /ɛ/ $F(2,51) = 14.6$; /ʌ/ $F(2,51) = 67.3$, Bonferroni $p<0.05$] and three vowels at T2 [/i/ $F(2,51) = 11.6$; /ɛ/ $F(2,51) = 21.0$; /ʌ/ $F(2,51) = 95.7$, Bonferroni $p<0.05$]. Tukey's tests indicated that the NE adults' scores were higher than those of both groups of NK adults ($p<0.05$), who did not differ significantly from each other. The only exception was for /ɪ/ at T1, where adult-3 but not adult-5 significantly differed from adult-NE. The scores obtained for the three groups of children were evaluated in a similar fashion. In no instance did the effect of Group reach significance (Bonferroni $p<0.05$).

The second analysis focused on adult–child differences, comparing participants who differed in age (child-NE vs. adult-NE, child-3 vs. adult-3, child-5 vs. adult-5). A total of 42 analyses (7 vowels × 2 testing times × 3 comparison groups) were performed. The Age factor reached significance for just one vowel produced by the NE participants (the NE adults' /ɪ/ scores were

Table 7
Classifications of vowels produced by native English and native Korean adults in Experiment 3. Each percentage is based on 54 tokens (18 speakers × 3 words). The percentages listed to the left and right of the slash are the values obtained at two testing times (T1 and T2). Vowels heard as intended are in shown in boldface

Vowel said		Vowel heard						
		/i/	/ɪ/	/eɪ/	/ɛ/	/æ/	/ɑ/	/ʌ/
/i/	NE	**91/85**	0/0	9/15				
	LOR-5	**48/43**	43/41	9/17				
	LOR-3	**61/61**	30/33	9/6				
/ɪ/	NE	0/0	**96/96**	2/2	2/2			
	LOR-5	17/6	**80/89**	0/2	4/4			
	LOR-3	26/19	**67/81**	0/0	7/0			
/eɪ/	NE			**100/100**	0/0			
	LOR-5			**98/100**	2/0			
	LOR-3			**100/98**	0/2			
/ɛ/	NE				**81/80**	19/20		
	LOR-5				**35/26**	65/74		
	LOR-3				**43/33**	57/67		
/æ/	NE				0/0	**100/100**		
	LOR-5				15/11	**85/89**		
	LOR-3				7/6	**93/94**		
/ɑ/	NE					0/0	**94/94**	6/6
	LOR-5					0/2	**91/94**	9/4
	LOR-3					0/0	**100/100**	0/0
/ʌ/	NE						0/0	**100/100**
	LOR-5						76/85	**24/15**
	LOR-3						83/85	**17/15**

Note: NE = native English; LOR-5 = native Korean adults with a length of residence in North America for 5 years at T1; LOR-3 = native Korean adults with a length of residence in North America for 3 years at T1.

higher than the NE children's at both T1 [$F(1,34) = 14.2$] and T2 [$F(1,34) = 10.2$, Bonferroni $p < 0.05$]). The NK adults' scores were lower than the NK children's in several instances. Age effects were significant for the LOR-3 groups for /ɛ/ at T2 [$F(1,34) = 15.7$] and for /ʌ/ at both T1 [$F(1,34) = 16.2$] and T2 [$F(1,34) = 30.1$, Bonferroni $p < 0.05$]. The effect of Age was significant for the LOR-5 groups for /i/ at both T1 [$F(1,34) = 31.1$] and T2 [$F(1,34) = 33.2$], and for /ʌ/ at both T1 [$F(1,34) = 17.7$] and T2 [$F(1,34) = 21.9$, Bonferroni $p < 0.05$].

4.3. Acoustic analysis

Age- and gender-related differences in vocal tract size affect the formant frequencies of vowels (e.g., Peterson & Barney, 1952; Eguchi & Hirsh, 1969). Researchers have attempted to reduce such physiologically based acoustic difference through normalization (e.g., Disner, 1980; Yang, 1996), but no technique appears to be completely satisfactory. The present study included participants differing greatly in vocal tract size. We reasoned, therefore, that direct comparisons of the formant frequencies in vowels produced by the six groups would be inappropriate. We decided, instead, to compare groups in terms of the magnitude of between-vowel formant frequency differences.

The classification results obtained in the preceding section, as well as data reported by others (Flege et al., 1997; Ingram & Park, 1997; Park, 1997; Tsukada et al., 2003), suggested that the NK adults have difficulty in producing a perceptually effective contrast between English /ɛ/ and /æ/. Specifically, the classification results suggested that when the NK participants erred in producing English /ɛ/, they tended to produce an [æ]-quality vowel; and when they erred for /æ/, they tended to produce an [ɛ]-quality vowel. The NK participants' /ɑ/ and /ʌ/ productions also tended to show a pattern of bi-directional confusion. Their difficulty in producing effective distinctions between /ɛ/-/æ/ and /ɑ/-/ʌ/ might, therefore, be characterized as one of vowel contrast *reduction*.

The pattern of errors in vowel production led us to examine the magnitude of formant frequency differences between the two pairs of vowels which exhibited a pattern of bi-directional confusion, /ɛ/-/æ/ and /ɑ/-/ʌ/. We decided against examining the magnitude of differences between other pairs of vowels. The errors observed for /i/-/ɪ/ and /eɪ/-/ɛ/ were not bi-directional, and thus did not afford the opportunity for a between-group comparison of vowel contrast magnitude. For instance, if /ɛ/ were misproduced as /æ/, it would result in an *increase* of acoustic contrast between /eɪ/ and /ɛ/.

4.3.1. Measurement techniques

The Kay Elemetrics MultiSpeech program was used to measure formant frequencies in 2592 vowel tokens (108 participants $\times$ 4 target vowels $\times$ 3 words $\times$ 2 testing times). As mentioned earlier (see Table 5), the target vowels were produced in diverse phonetic contexts. We therefore examined vowels at their acoustic midpoint in order to minimize the effect of the differing consonantal contexts.

The beginning and end of each vowel token was identified by inspection of wide-band spectrograms and time domain waveforms. A 20-ms Hamming window was centered at the acoustic midpoint of each token. The frequency of the first three vowel formants (F1–F3) at this location was estimated using the auto-correlation method of linear predictive coding (LPC) analysis. Twenty-two LPC coefficients were calculated for most tokens. Where they failed to produce reliable formant frequency values, the number of LPC coefficients was adjusted accordingly. Usually, this involved using a higher number of coefficients (e.g., 26 or 28) for vowels produced by children. The same software was used to obtain estimates of fundamental frequency (f_0). When spurious values were returned, f_0 was calculated manually on the basis of the inverse of the average duration (in ms) of three glottal pulses.

The following procedure was adopted to quantify the magnitude of spectral differences between each participant's productions of /ɛ/-/æ/ and /ɑ/-/ʌ/. First, the formant frequency values were converted from Hertz to Bark units using the procedure described by Syrdal and Gopal (1986). Bark difference values were then calculated for each token. One set of values ($B_1 - B_0$) indexed vowel quality in a high-low dimension of the acoustic vowel space. The other set of values ($B_2 - B_1$) indexed vowel quality in a front-back dimension. Next, the mean Bark difference values obtained for the three tokens of each vowel were calculated. Finally, the Euclidean distance between the mean Bark difference values representing the two vowels in each contrast was calculated. The larger the Euclidean distance, the greater the magnitude of the frequency contrast that was produced.

Given the pattern of bi-directional confusions noted for /ɛ/-/æ/ and /ɑ/-/ʌ/ and the larger differences in intelligibility scores between the NE and NK adults than between the NE and NK children, we expected to obtain smaller distance scores for vowels produced by the NK adults than the NK children. The primary questions of interest, therefore, were whether significant differences would exist between the NE and NK children, and if the NK children's distance scores would differ as a function of LOR or time of testing.

4.3.2. Results

Fig. 3 shows the mean Bark distance scores obtained for the six groups for /ɛ/-/æ/ (left panel) and /ɑ/-/ʌ/ (right panel). The NK children produced larger frequency contrasts between both pairs of vowels than the NK adults did, but there was little difference between the NE children and NE adults. Also, differences between the NE and NK children were smaller than those between the NE and NK adults. Statistical analyses confirmed that the NK adults produced significantly smaller between-vowel contrasts than the NE adults did. The analyses also showed that the NK children produced larger contrasts than did the NK adults who were matched for LORs, and that they did not differ significantly from the NE children.

The Bark distance scores obtained for the two vowel contrasts were examined in separate Age (child vs. adult) × English-language Experience (LOR-3, LOR-5, NE) × Testing Time (T1 vs. T2) ANOVAs. The analyses of both vowel contrasts yielded significant main effects of age [/ɛ/-/æ/ $F(1,102) = 38.5$, $p < 0.01$; /ɑ/-/ʌ/ $F(1,102) = 21.3$, $p < 0.01$] and Experience [/ɛ/-/æ/ $F(2,102) = 13.4$, $p < 0.01$; /ɑ/-/ʌ/ $F(2,102) = 19.5$, $p < 0.01$]. The main effect of Testing Time was

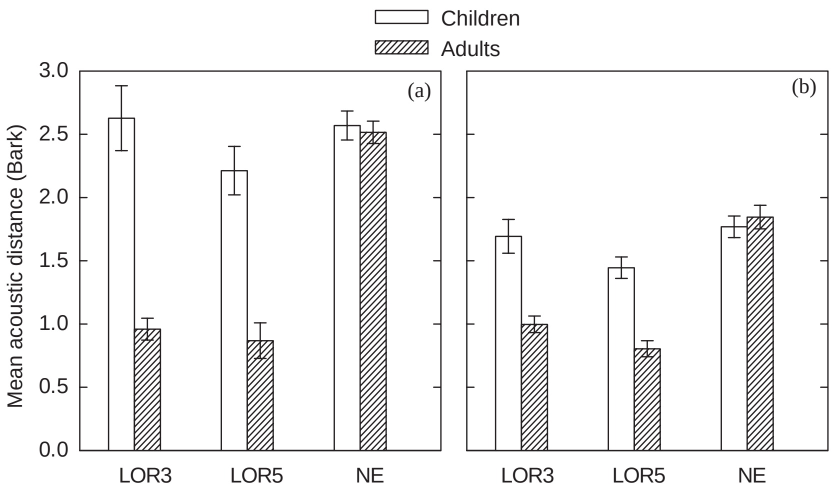

Fig. 3. Mean acoustic distance, in Bark, for (a) /ɛ/-/æ/ and (b) /ɑ/-/ʌ/. The NE and NK participants differed in age (child vs. adult); the NK participants differed in age and length of residence in North America (LOR-3 vs. LOR-5). The scores have been averaged over two testing times (T1 and T2). The error bars enclose ± one standard error.

non-significant in both analyses [/ɛ/-/æ/ $F(1,102) = 2.0$, *ns*; /ɑ/-/ʌ/ $F(1,102) = 0.05$, *ns*] and did not enter into any significant interactions. However, both analyses yielded significant Age × Experience interactions [/ɛ/-/æ/ $F(2,102) = 9.0$, $p < 0.01$; /ɑ/-/ʌ/ $F(2,102) = 7.5$, $p < 0.01$].

Simple effects tests revealed that the Age × Experience interactions arose from adult–child differences for the NK but not NE participants, and native-nonnative differences for adults but not children. Significant differences were found to exist between the NK children and adults with an LOR of 3 years [/ɛ/-/æ/ $F(1,34) = 23.4$, $p < 0.01$; /ɑ/-/ʌ/ $F(1,34) = 14.8$, $p < 0.01$], and 5 years [/ɛ/-/æ/ $F(1,34) = 20.2$, $p < 0.01$; /ɑ/-/ʌ/ $F(1,34) = 23.5$, $p < 0.01$], but not between the NE children and adults [/ɛ/-/æ/ $F(1,34) = 0.08$, *ns*; /ɑ/-/ʌ/ $F(1,34) = 0.24$, *ns*]. The effect of Experience was significant for adults [/ɛ/-/æ/ $F(2,51) = 44.9$, $p < 0.01$; /ɑ/-/ʌ/ $F(2,51) = 36.8$, $p < 0.01$] but not children [/ɛ/-/æ/ $F(2,51) = 0.8$, *ns*; /ɑ/-/ʌ/ $F(2,51) = 1.7$, *ns*]. Tukey's tests revealed that the NE adults produced significantly larger contrasts for both /ɛ/-/æ/ and /ɑ/-/ʌ/ than both NK adult groups ($p < 0.05$), who did not differ significantly from each other.

4.4. Discussion

As with the discrimination of English vowels (see Experiment 2), the NK children more closely resembled the NE children than the NK adults resembled the NE adults. The effect of Age that was obtained here for the NK but not NE participants probably cannot be attributed to physiological differences in vocal tract size because adult–child vocal tract differences were likely to have been comparable for the NK and NE participants.

Although this experiment did not reveal significant differences in vowels produced by the NE and NK children, a recent study (Flege et al., in press) showed that English sentences spoken by the NK children received significantly lower ratings (indicating the presence of detectable foreign accents) than sentences spoken by the NE children did. Perhaps between-group differences in vowel production would have been observed here if other parameters (e.g., extent or direction of formant movement) had been examined in addition to the magnitude of acoustic distance at the vowel midpoint. Although it would have been desirable to provide analyses of vowel duration (see Kim, 1994), we did not undertake such analyses because the target vowels appeared in a variety of contexts (see Table 5).

5. General discussion

As discussed in the Introduction, the traditional view of adult–child differences in L2 speech learning suggests that children are rapid and successful learners whereas adults are slow and imperfect learners (see discussion by Snow, 1987, page 192). Most previous research showing age effects on L2 learning have focused on groups of adults who differed in their age of first exposure to the L2 (oftentimes, many years in the past). The aim of this study was to evaluate the traditional view of adult–child differences by directly comparing adults and children who were in the process of learning an L2.

Age effects on the perception of L2 phonetic segments shown in various studies (Yamada, 1995; Flege et al., 1999a; Flege & MacKay, 2004) suggested that, in the present study, the NK children would discriminate English vowels more accurately than the NK adults. This inference was confirmed in Experiment 2. However, contrary to the traditional view of children as rapid and perfect learners of L2 speech, the NK children with an LOR of 3 years were found to have discriminated English vowels less accurately than the age-matched NE children did.

The production of English vowels was examined in Experiment 3. One analysis was based on the classifications by a single phonetically trained observer. The NK adults' vowels were found to be significantly less intelligible than the NE adults' vowels, whereas the intelligibility scores obtained for vowels spoken by the NK and NE children did not differ significantly. The NK children's vowels were more intelligible than the NK adults' in several instances. The magnitude of formant frequency differences between English /ɛ/-/æ/ and /ɑ/-/ʌ/ were determined through acoustic measurement. Analyses of these values yielded results that were similar to those obtained in analyses of the classification data. The NK adults produced smaller between-vowel contrasts than the NE adults did whereas differences between the NE and NK children were non-significant. The vowel production results agreed with the results of previous studies employing a retrospective developmental design (e.g., Oyama, 1976; Munro et al., 1996; Piske et al., 2002).

Unlike the vowel discrimination results obtained in Experiment 2, the NK children were not found to differ significantly from the NE children in producing English vowels (either in terms of intelligibility for the seven target vowels, or the magnitude of between-vowel formant frequency differences for /ɛ/-/æ/ and /ɑ/-/ʌ/). Baker et al. (2002) reported that NK children with an LOR of less than 1 year produced English vowels less accurately than NE children did. This earlier finding, when taken together with the present results, suggests that differences in L2 vowel production between native and nonnative children may largely disappear by the time nonnative children reside in a predominantly L2 speaking environment for about 3 years. This inference needs to be evaluated in future research comparing groups consisting of individuals with LOR ranging from roughly 6 months to 4 years.

Although the pattern of results presented here are straightforward, interpretation of the results are not entirely clear. A question of longstanding interest is why the NK children produced and perceived English vowels more accurately than the NK adults. One possible explanation was offered by Scovel (1988, 2000), who proposed that a critical period for L2 speech learning ends at about the age of 12 years. On this view, the adult–child differences obtained in this study arose from the fact that all of the NK adults began learning English after the end of a critical period, whereas this held true for only some of the NK children.

To evaluate the possible effect of passing a critical period at the age of 12 years, we compared the performance of the 12 younger NK children (*mean* age at $T1 = 10.5$ years) to that of the 12 older NK children (*mean* age $= 15.6$ years). The "pre-" and "post-critical period" groups did not differ significantly in LOR (*means* at $T1 = 3.6$ vs. 4.2 years) [$F(1,22) = 1.7$, *ns*]. The two groups were not found to differ in terms of their average discrimination of English vowels in Experiment 2 [*means* $= 0.76$ vs. 0.82, $F(1,22) = 1.1$, *ns*] nor in the average intelligibility of the vowels they produced in Experiment 3 [*means* $= 87\%$ vs. 83%, $F(1,22) = 0.7$, *ns*]. These results did not support a critical period account of the adult–child differences observed here, at least not if the critical period ends at about the age of 12 years as proposed by Scovel (1988, 2000).

If adult–child differences are not the result of the passing of a critical period, then what does account for age effects on L2 vowel learning? We speculate that the age effects seen here and in previous research (e.g., Baker et al., 2002) arise from age-related differences in how the L1 and L2 phonetic systems interact, to age-related differences in input, or some combination of both factors. It is likely that the NK adults' L1 categories were more fully developed than the NK children's when L2 learning began (see e.g., Hazan & Barrett, 2000). Flege (1995, 1999a, 2003b) hypothesized that the formation of new categories for L2 vowels and consonants becomes less likely with increasing age. By hypothesis, as L1 categories develop, they become more likely to block the formation of new L2 vowel and consonant categories. The presence versus absence of new categories for L2 speech sounds may, in turn, account for the magnitude of native-nonnative differences (see e.g., Flege & Eefting, 1987, 1988).

The NK children were likely to have received more input from NE speakers than the NK adults had. Work by Jia and Aaronson (1999) suggests that immigrant children are more likely than immigrant adults to become dominant in the L2 because they, unlike adults, are typically immersed in an L2-rich environment. Flege, Yeni-Komshian, and Liu (1999b) found that NK adults who arrived in the US prior to the age of 12 years used English more (and Korean less) than NK adults who arrived in the US between the ages of 13–22 years did. Not surprisingly, the NK children examined in this study reported using English more than the NK adults (see Table 3).

Children are traditionally seen as being able to learn an L2 more rapidly than adults. Although this may be true, it is important to note that the present study has not provided valid measures of rate of L2 learning. Rate of learning by adults and children can only be compared if the groups received identical input in a unit of time. Many factors are responsible for the slope of the curve relating age of exposure to the L2 and performance in the L2. These factors may include cognitive changes, amount and kind of education received during the study interval, and social and motivational factors. As noted by Hakuta, Bialystok, and Wiley (2003), joining two disparate data points in a time-performance space (e.g., for young children and middle-aged adults) tells us little about the path between those points.

We did not observe significant differences between groups of NK adults with LORs of 3 versus 5 years, nor significant changes between T1 and T2. The lack of significant differences for the NK adults cannot be attributed to test insensitivity because differences were observed for the NK children. In our view, the present results should not be taken as evidence that adults are incapable of L2 speech learning. Learning effects might have been obtained, for example, if we had examined NK adults who did not study English in Korea before coming to North America, if the LOR-3 group were replaced with a group consisting of individuals with less than one year of residence in North America, or if the LOR-5 group were replaced by individuals who were more experienced in English. Baker et al. (2002) observed that NK adults with an LOR of 9 years perceived and produced /ɑ/ and /ʌ/ as accurately as NE adults. How much English-language experience is needed to show significant improvements in L2 performance, and the question of whether the NK adults can ever match the performance of NE speakers needs to be further evaluated in studies comparing groups having both more and less English-language experience than the groups considered here (see e.g., Flege & Fletcher, 1992).

It is generally agreed that the relationship between production and perception is complex and is affected by factors such as amount of L2 experience (e.g., Yamada et al., 1994; Llisterri, 1995). A number of investigators have suggested that L2 acquisition recapitulates L1 acquisition in the sense that perceptual development tends to "lead" development in segmental production (Flege, 1995; Rochet, 1995). The pattern of results obtained in Experiments 2 and 3 might be taken as a disconfirmation of this hypothesis, inasmuch as the NK children seem to have been more native-like in their production than discrimination of English vowels. The present findings might be taken as support for an alternative hypothesis, namely that L2 segmental production leads L2 segmental perception (e.g., Goto, 1971; Sheldon & Strange, 1982; Yamada et al., 1994). However, there are several reasons for caution in concluding that production learning precedes perceptual learning in an L2.

First, it is difficult to directly compare the extent of native-like attainment in L2 production and perception data. This is because the tests used to assess production and perception are inherently incommensurable (e.g., Gass, 1984; Flege, 1999b). For example, the tasks used in Experiments 2 and 3 might not have evaluated productive and perceptual abilities with the same degree of thoroughness.

Second, the tasks used in Experiments 2 and 3 to assess vowel perception and production might not have been of equivalent difficulty for the NK participants. It is conceivable that the categorial discrimination test in Experiment 2 was cognitively more demanding than the picture-naming task used to elicit vowel production. Moreover, the use of auditory models to elicit production might have allowed the NK participants to produce the English vowels more accurately than they typically do.

Two other points are also worth making with respect to how one might interpret the relation between L2 production and perception learning. The English vowels /ɛ æ ɑ ʌ/ did not occur in identical phonetic contexts when examined in perception (Experiment 2) and production (Experiment 3). It is not possible, therefore, to rule out a differing effect of phonetic context in the two experiments (see e.g., Strange, Akahane-Yamada, Kubo, Trent, & Nishi, 2001; Bohn & Steinlen, 2003). Secondly, previous research has suggested that the social consequences of errors in production and perception might differ (e.g., Sheldon, 1985; Mack, 1989; Bohn & Flege, 1997). For example, adults might feel more pressure to demonstrate ‘‘language loyalty’’ by speaking with a foreign accent than children do. A similar social pressure might not exist for L2 vowel perception, however.

In summary, this study revealed that the NK children resembled the age-matched NE speakers to a greater extent than the NK adults did in both discriminating and producing English vowels. The NK adults failed to show native-like performance for either vowel discrimination or production. The NK children's vowel production abilities largely matched those of the NE children. However, the NK children with an LOR of 3 years differed significantly from the NE children in discriminating four pairs of English vowels and the NK children with an LOR of 5 years differed for two vowel contrasts. These results are consistent with the view that children are more successful than adults in L2 speech learning, but may not be indistinguishable from native children.

Acknowledgements

This research was supported by NIH grant DC02892. The authors thank Lance Gunnin for help collecting data, and Terry Gottfried, Anders Højen, Satomi Imai, D. H. Whalen and two anonymous reviewers for their comments on previous versions of this article.

References

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2003). Foreign accent in English words produced by Japanese children and adults. In *Proceedings of the fifteenth international congress of phonetic sciences* (pp. 3201–3204), Barcelona.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /ɾ/ and English /l/ and /ɹ/. *Journal of Phonetics*, *32*, 233–250.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. Developmental effects in the acquisition of English fricatives (under review).

Asher, J. J., & García, R. (1969). The optimal age to learn a foreign language. *The Modern Language Journal*, *53*, 334–341.

Baker, W., Trofimovich, P., Mack, M., & Flege, J. E. (2002). The effect of perceived phonetic similarity on nonnative sound learning by children and adults. *Paper presented at the The Twenty-sixth Annual Boston University Conference on Language Development*, Somerville, MA.

Best, C. T. (1993). Emergence of language-specific constraints in perception of nonnative speech: a window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: speech and face processing in the first year of life* (pp. 289–304). Dordrecht: Kluwer.

Best, C. T. (1995). A direct realist view of cross-language speech. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

Bohn, O.-S., & Flege, J. E. (1997). Perception and production of a new vowel category by adult second language learners. In J. Leather, & A. James (Eds.), *Second language speech: structure and process* (pp. 53–73). Berlin: Mouton de Gruyter.

Bohn, O. -S. & Steinlen, A. K. (2003). Consonantal context affects cross-language perception of vowels. In *Proceedings of the fifteenth international congress of phonetic sciences* (pp. 2289–2292), Barcelona.

Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, *12*, 189–221.

Disner, S. F. (1980). Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America*, *67*, 253–261.

Eguchi, S., & Hirsh, I. J. (1969). Development of speech sounds in children. *Acta Oto-Laryngologica, Supplementum*, *257*, 1–51.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, *84*, 70–79.

Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (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* (pp. 213–232). Chicago: Chicago Linguistic Sociey.

Flege, J. E. (1999a). Age of learning and second language speech. In D. Birdsong (Ed.), *New perspectives on the critical period hypothesis for second language acquisition* (pp. 101–131). Hillsdale, NJ: Lawrence Erlbaum.

Flege, J. E. (1999b). The relation between L2 production and perception. In *Proceedings of the fourteenth international congress of phonetic sciences* (pp. 1273–1276). San Francisco.

Flege, J. E. (2003a). Methods for assessing the perception of vowels in a second language. In E. Fava, & A. Mioni (Eds.), *Issues in clinical linguistics* (pp. 3–28). Padova: UniPress.

Flege, J. E. (2003b). Assessing constraints on second-language segmental production and perception. In A. Meyer, & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production, differences and similarities* (pp. 319–355). Berlin: Mouton de Gruyter.

Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. The production of English sentences by native English and native Korean adults and children. *Journal of Phonetics*, in press.

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

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

Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*, 67–83.

Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: evidence for phonetic category formation. *Journal of the Acoustical Society of America*, *83*, 729–740.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America*, *91*, 370–389.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*, 1–34.

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

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

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*, 467–491.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999b). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*, 78–104.

Gass, S. (1984). Development of speech perception and speech production abilities in adult second language learners. *Applied Psycholinguistics*, *5*, 51–74.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia*, *9*, 317–323.

Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America*, *106*, 2900–2912.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: the case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, *107*, 2711–2724.

Guion, S. G., Flege, J. E., & Loftin, J. D. (2000). The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics*, *28*, 27–42.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: a test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*, 31–38.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *25*, 377–396.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*, 3099–3111.

Hong, Y. (1988). *A sociolinguistic study of Seoul Korean.* Unpublished Ph.D. thesis, University of Pennsylvania.

Ingram, J. C. L., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, *25*, 343–370.

Jia, G., & Aaronson, D. (1999). Age difference in second language acquisition: the dominant language switch and maintenance hypothesis. *Paper presented at the twenty-third Annual Boston University Conference on Language Development*.

Kang, S.-K. (1996). Acoustic analysis of South and North Korean vowels. *Language Research*, *32*, 1–18.

Kim, R. (1994). *The sensitive period hypothesis in L2 production: /i/ versus /ɪ/ and voice-dependent vowel duration in the English speech produced by Korean speakers of English*. Unpublished Ph.D. thesis, University of Illinois at Urbana-Champaign.

Kuhl, P. K. (1980). Perceptual constancy for speech sound categories in early infancy. In G. H. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology: perception*, Vol. 2 (pp. 41–66). New York: Academic Press.

Lee, H. B. (1993). Korean. *Journal of the International Phonetic Association*, *23*, 25–28.

Lee, H. B., & Zhi, M.-J. (1987). A spectrographic study of Korean vowels. *Korea Journal*, *27*, 37–41.

Lee, I., & Ramsey, S. R. (2000). *Korean language*. Albany: State University of New York Press.

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, *29*, 117–123.

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In *Proceedings of the thirteenth international congress of phonetics sciences* (pp. 92–99), Stockholm.

Mack, M. (1989). Consonant and vowel perception and production: early English–French bilinguals and English monolinguals. *Perception & Psychophysics*, *46*, 187–200.

MacKay, I. R. A., Flege, J. E., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *Journal of the Acoustical Society of America*, *110*, 516–528.

Magen, H. S., & Blumstein, S. E. (1993). Effects of speaking rate on the vowel length distinction in Korean. *Journal of Phonetics*, *21*, 387–409.

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.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*, 261–283.

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

Park, S.-G. (1997). *Australian English pronunciation acquisition by Korean and Japanese learners of English.* Unpublished Ph.D. thesis, The University of Queensland, St. Lucia, Queensland.

Patkowski, M. (1990). Age and accent in a second language: a reply to James Emil Flege. *Applied Linguistics*, *11*, 73–89.

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress 'deafness'. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology 7* (pp. 203–240). Berlin: Mouton de Gruyter.

Peterson, G. E., & Barney, H. L. (1952). Control method used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175–184.

Piske, T., Flege, J. E., & MacKay, I. R. A. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, *29*, 191–216.

Piske, T., Flege, J. E., MacKay, I. R. A., & Meador, D. (2002). The production of English vowels by fluent early and late Italian–English bilinguals. *Phonetica*, *59*, 49–71.

Polka, L. (1995). Linguistic influences in adult perception of nonnative vowel contrasts. *Journal of the Acoustical Society of America*, *97*, 1286–1296.

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 379–410). Timonium, MD: York Press.

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

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, *20*, 213–223.

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and nonnative phonemic contrasts in early bilinguals. *Cognition*, *72*, 111–123.

Seliger, H. W., Krashen, S. D., & Ladefoged, P. (1975). Maturational constraints in the acquisition of second language accent. *Language Sciences*, *36*, 20–22.

Sheldon, A. (1985). The relationship between production and perception of the /r/-/l/ contrast in Korean adults learning English: a reply to Borden, Gerber, and Milsark. *Language Learning*, *35*, 107–113.

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

Singleton, D. (1989). *Language acquisition: the age factor*. Clevedon, England: Multilingual Matters.

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

Snow, C. E. (1987). Relevance of the notion of a critical period to language acquisition. In M. H. Bornstein (Ed.), *Sensitive periods in development: interdisciplinary perspective* (pp. 183–209). Hillsdale, NJ: Lawrence Erlbaum.

Snow, C. E., & Hoefnagel-Höhle, M. (1977). Age differences in the pronunciation of foreign sounds. *Language and Speech*, *20*, 357–365.

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: evidence from second language learning. *Child Development*, *49*, 1114–1128.

Sohn, H.-M. (1999). *The Korean language*. Cambridge: Cambridge University Press.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America*, *109*, 1691–1704.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America*, *60*, 213–224.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, *79*, 1086–1100.

Trofimovich, P., Baker, W., & Mack, M. (2001). Context- and experience-based effects of the learning of vowels in a second language. *Studies in the Linguistic Sciences*, *31*, 167–186.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2003). The perception and production of English /ɛ/ and /æ/ by Korean children and adults living in North America. In *Proceedings of the fifteenth international congress of phonetic sciences* (pp. 1589–1592), Barcelona.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception and Psychophysics*, *26*, 95–104.

Yamada, R. A. (1995). Age and acquisition of second language speech sounds perception of American English /r/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 305–320). Timonium, MD: York Press.

Yamada, R. A., Strange, W., Magnuson, J. S., Pruitt, J. S., & Clark, W. D. (1994). The intelligibility of Japanese speakers' production of American English /r/, /l/, and /w/, as evaluated by native speakers of American English, In *Proceedings of the third international conference on spoken language processing* (pp. 2023–2026), Yokohama.

Yang, B. (1992). An acoustical study of Korean monophthongs produced by male and female speakers. *Journal of the Acoustical Society of America*, *91*, 2280–2283.

Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*, *24*, 245–261.

Yeni-Komshian, G. H., Flege, J. E., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean–English bilinguals. *Bilingualism: Language and Cognition*, *3*, 131–149.