The first years in an L2-speaking environment: A comparison of Japanese children and adults learning American English*

KATSURA AOYAMA, SUSAN G. GUION, JAMES EMIL FLEGE, TSUNEO YAMADA AND REIKO AKAHANE-YAMADA

Abstract

This study examined Japanese speakers' learning of American English during their first years of immersion in the United States (U.S.). Native Japanese-speaking (NJ) children (n=16) and adults (n=16) were tested on two occasions, averaging 0.5 (T1) and 1.6 years (T2) after arrival in the U.S. Age-matched groups of native English-speaking children (n=16) and adults (n=16) also participated. The NJ adults' scores for segmental perception and production were higher than the NJ children's at T1. The NJ children's foreign accent scores and pronunciation of English fricatives improved significantly between T1 and T2, whereas the NJ adults' scores did not. These findings suggest that adults may have an advantage over children initially but that children may improve in production more quickly than adults when immersed in an L2-speaking environment.

1. Introduction

1.1. *Previous research*

It has been reported that age of acquisition has strong effects on the degree of accent in a second language (L2) (e.g., Flege 1995; Ioup, 2005; Moyer 1999, 2004; Oyama 1976; Piske et al. 2001; Snow 1987; Tahta et al. 1981). Asher and Garcia (1969), for example, studied 71 Cuban children between the ages of 7 and 19 who arrived in the United States (U.S.) at different ages. Their oral productions were judged by native speakers of English as "definitely native",

* This research was supported by NIH Grant DC00027 and in part by Japan Society for the Promotion of Science (JSPS) Grant 17202012. The authors thank L. Legault for her help in data collection, M. Burnett for editorial assistance, and the reviewers for comments that were provided on previous versions of this article.

IRAL 46 (2008), 61–90 0019042X/2008/046-061
DOI 10.1515/IRAL.2008.003

"near native", "slight foreign accent" or "definite foreign accent". It was found that those who arrived in the U.S. before age 6 were often judged as "near native" in terms of production in English whereas those who arrived after the age of 13 years were often judged to be "definitely" foreign-accented with 7–12 year-olds falling in between the two groups. From these findings, Asher and Garcia (1969) concluded the younger the child is at the onset of L2 acquisition, the more likely his or her accent will be considered "near native". Moyer (2004) studied the effects of age on the degree of foreign accent in German as an L2. The participants were 25 adult learners of German as an L2 who spoke various native languages (L1s), such as Russian, Polish and English. The German speech samples were judged by 3 native German speakers for the degrees of foreign accent. The results indicated that the earlier the adult participants' exposure to the L2 (German), the less foreign accented the speech samples were perceived to be, regardless of the participants' L1 background.

In addition to showing that "earlier is better", the results of previous research has also suggested that learners who are exposed to their L2 early in life ("early learners") may not speak their L2 in a completely "native-like" manner, even though they may have milder foreign accent compared to those who learn it later in life ("late learners"). For instance, none of the 71 children in Asher and Garcia (1969) were judged as "definitely native", including the ones who arrived in the U.S. between the ages of 1 and 6. Yeni-Komshian et al. (2000) evaluated degree of foreign accent in English spoken by 240 Korean adults whose age of arrival (AOA) in the U.S. ranged from 1 to 23 years. Twenty-four English sentences were elicited from the participants, and the speech samples were rated by 10 native English speakers. The results indicated that all of the Korean-English bilingual groups, including the group with the youngest AOA (1 to 5 years), differed from the monolingual English group in terms of their oral production scores (i.e., overall degree of perceived foreign accent in English sentences). In addition, the results indicated a negative linear relationship between AOA and production scores in English. That is, the mean production scores decreased as the AOA of the group increased.

In sum, all of the above studies confirm age effects on L2 learning and suggest that the younger the learners are, the more advantage they would have in L2 learning at least in oral production skills. Based on a review of studies on age and L2 acquisition, Ioup (2005) suggested that child L2 acquisition is fundamentally different from adult L2 acquisition. She also stated that direct comparisons of adults and children are necessary in order to understand how L2 acquisition processes in children differ from those in adults.

During the first years in an L2-speaking environment, however, younger learners may not have an advantage over older learners (Ervin-Tripp 1974; Snow and Hoefnagel-Höhle 1977, 1978; Winitz 1981). Snow and Hoefnagel-Höhle (1977) studied production of Dutch phonemes in 47 native English-speaking (NE) adults and children learning Dutch in the Netherlands for the first two years after their arrival. Participants were assigned to subgroups based on age: 3–5 years, 6–7 years, 8–10 years, 12–15 years, and adults (21 and older). The participants' productions of 24 Dutch sounds were examined on 3 occasions with 4–5 month intervals. At the first testing (within 6 months of arrival in the Netherlands), there was a positive linear relationship between the production scores and age. That is, the 3–5 year-olds' scores were the lowest and the adults' were the highest. However, the children's oral production scores improved after one year of stay in the Netherlands whereas the adults' production scores in Dutch did not (Snow and Hoefnagel-Höhle 1977). Snow (1987) attributed the adult-child difference to differences in input and other factors that favored the younger learners. The children studied by Snow and Hoefnagel-Höhle (1977, 1978) were all enrolled in Dutch-speaking schools, whereas the adults were in the environment where their L1 (English) was commonly spoken.

Snow and Hoefnagel-Höhle (1978) examined multiple aspects of L2 learning in the same group of participants: pronunciation (imitation and spontaneous), auditory discrimination, morphology, sentence repetition, sentence translation, sentence judgment, vocabulary, and story comprehension. It was found that, within their first 6 months of residence in the Netherlands, the adults obtained higher scores than children younger than 10 on all tasks except for the imitation task. However, the NE children either outperformed, or performed equivalently to the NE adults, for all measures when tested about one year later. The 12–15 year-old group scored the highest of all groups in all tasks except for imitation. The 12–15 year-old also showed most rapid improvement, whereas adults showed a smaller improvement than the younger participants did. Similar results were obtained by Ervin-Tripp (1974) for English-speaking children learning French as an L2 in Switzerland. Thirty-one children, from 4 to 9 years of age, were studied. All of these children had lived in the L2 environment for less than 9 months. Their skills in French were tested through imitation, sentence comprehension, and translation. The results indicated that the older children (7–9 year-olds) learned French more quickly than the younger children (4–7 year-olds) in every aspect examined in this study.

More recently, Flege et al. (2006) and Tsukada et al. (2004, 2005) studied native Korean-speaking (NK) adults and children (mean age = 12–13 years) with differing length of residence (LOR) in the U.S. There were a total of 108 participants with 18 participants in each of the following 6 groups: NE adults, NE children, NK adults with LOR of 5 years, NK adults with LOR of 3 years, NK children with LOR of 5 years, and NK children with LOR of 3 years. A series of experiments examined these participants' production of oral stop consonants (Tsukada et al. 2004), vowel perception and production (Tsukada et al. 2005) and the degree of foreign accent (Flege, et al. 2006). Overall, the results indicate that the NK children perceived and produced English vowels more accurately than the NK adults (Tsukada et al. 2005), and that their production of stop consonants was closer to that of NE speakers (Tsukada et al. 2004). Flege et al. (2006) showed that the NK children obtained higher ratings on the degree of foreign accent than the NK adults, indicating a milder foreign accent, although both the NK children and adults received significantly lower ratings than the NE children and adults. Taken together, the results of these studies suggested that the Korean children had already surpassed the Korean adults in terms of segmental perception and production as well as overall degree of perceived foreign accent after 3 to 5 years of residence in an L2-speaking environment. The results suggested, further, that L2 learning in children may take place quite rapidly during the first years of immersion.

1.2. *The current study*

The primary aim of this study was to investigate phonetic measures related to aspects of L2 phonological acquisition[1] in Japanese adults and children learning English in the first two years of immersion in an L2-speaking environment. We will present fricative perception and production data as one instance of segmental proficiency and foreign accent rating as an index of global production proficiency. The current study is similar to Flege et al. (2006) and Tsukada et al. (2004, 2005) in design, except that the native Japanese-speaking (NJ) participants studied here had been living for a shorter period of time in the U.S. than the NK participants (mean LOR values of 0.6 and 1.6 years vs. 3 and 5 years in the research with Koreans). Our main question is whether the NJ adults tested here would demonstrate an initial advantage over the NJ children, and if so, whether the adults would maintain the advantage throughout the first two years that were examined in this study. The NJ adults and children differed greatly in terms of their previous experience in English. All of the NJ adults had studied written English for at least 6 years in Japan whereas only one NJ child had studied English before arriving in the U.S. (this child had studied English for two years). Thus, it is expected that the NJ adults would have at least some advantage over the NJ children in English generally, but the question investigated here was whether they would have an advantage over children in phonological aspects of the L2, as well as whether they would maintain their advantage over children for the period of investigation.

---

1. We used "phonetic measures" of "L2 phonological acquisition" throughout this article. We refer to "L2 phonology" as the sound system of the language that the learners are acquiring (in this case, English phonology). Our method is "phonetic" in the sense that our focus is on examining L2 speakers' speech perception and production. We then use these phonetic measures to draw inferences about the development of phonological systems.

Differences in the English and Japanese phonemic inventories motivated the decision to focus on the perception and production of word-initial English /s/ and /θ/ in this study (as in the words *six* and *think*). The English inventory includes nine fricatives, /h f v s z θ ð ʃ ʒ/. The Japanese inventory includes /s/, /z/ and /h/, which also occur in English, and a fricative that does not occur in English (the alveolo-palatal /ɕ/) (Vance 1987). Most importantly for the present study, Japanese lacks a /θ/. Japanese /ɕ/ contrasts with /s/ before /a o u/ (Vance 1987). Japanese /s/ and /ɕ/ do not contrast before the vowels /i/ and /e/, because only [ɕ] occurs before /i/, and [s] occurs only before /e/ (Vance 1987: 21). The Japanese phoneme /h/ is realized as a palatal fricative [ç] before /i/, and as a bilabial fricative [ɸ] before /u/ [ɯ]. [ɸ] also occurs before /a i e o/ in the loanwords that begin with an /f/ (e.g., [ɸirɯmɯ] for "film", Vance 1987).[2]

Previous research has shown that NJ speakers have difficulty producing and perceiving a contrast between English /s/ and /θ/ (Brannen 2002; Guion et al., 2000; Lambacher 1994; Singh and Black 1966; Yoshida and Hirasaka 1983). Yoshida and Hirasaka (1983) tested 96 Japanese participants for the discrimination of English contrasts that were considered difficult for Japanese learners of English. The /s/–/θ/ contrast was one of the contrasts tested and found to have an average error rate of 16 %. Similarly in Brannen (2002), the error rate for an /s/-/θ/ discrimination task was 26 % among 5 Japanese college students learning English in Canada. In an identification experiment by Lambacher et al. (2001), college students in Japan with little or no prior exposure to spoken English misidentified both /θ/ (as /s/, mean = 28 % of instances) and /s/ (as /θ/, mean = 25 % of instances).

Guion et al. (2000) examined the perception of English consonants, including /θ/ and /s/, by 30 NJ speakers living in the U.S. The NJ speakers' discrimination scores for the /s/–/θ/ contrast were substantially lower than the NE speakers' scores, suggesting little or no perceptual sensitivity to this English fricative contrast. In order to explain why some English contrasts may be more difficult than others for NJ speakers, Guion et al. (2000) also examined NJ speakers' identification of English consonants in terms of Japanese categories. Nine Japanese college students participated in this "perceptual assimilation" experiment; they identified /θ/ as a poor example of Japanese /s/ or a poor example of Japanese [ɸ]. The results of Guion et al. (2000) demonstrate that English /θ/ does not fit well with any of the Japanese phonemic categories, and that NJ speakers had great difficulty discriminating /s/ from /θ/.

---

2. Vance also discussed whether or not [ʑ] (voiced counterpart of [ɕ]) contrasts with other sounds (Vance 1987: 24–25). It will not be discussed further here because voiced fricatives are not examined in this study.

Weinberger (1997) used the term "differential substitution" for cases in which different L1 sounds are substituted for the same target L2 sound depending on the learners' L1 background. For instance, Russian speakers typically substitute their Russian /t/ for /θ/ in English, whereas Japanese speakers typically substitute /s/ for /θ/, even though both Russian and Japanese possess /s/ and /t/ (Weinberger 1997). Weinberger suggested that /s/ be considered the "default" obstruent in Japanese (1997: 29), while /t/ be considered the default in Russian. Brannen (2002) also argued that /s/ is the best match for /θ/ for Japanese speakers, particularly because Japanese /s/ is laminal and "mellow" (2002: 18).

It has also been reported that English fricatives are mastered relatively late by children learning English as their L1 (Ingram 1978; Ingram et al. 1980; Nissen and Fox 2005; Velleman 1988). It seems to be especially difficult for NE children to learn to distinguish /f/ from /θ/ perceptually (Abbs and Minifie 1969; Phatate and Umano 1981; Velleman 1983, 1988). Although most children are able to produce word-initial fricatives intelligibly by the age of 6 years (Ingram 1978), substitutions and moderate distortions may persist in some children beyond that age (Snow 1963). For example, developmental production studies have identified a variety of segmental substitutions in NE children's speech, including [f]-for-/θ/ (i.e., /θ/ tokens heard as [f]), [f]-for-/s/, [s]-for-/θ/ and [θ]-for-/s/ (Ingram et al. 1980; Moskowitz 1975; Smith 1973; Velleman 1983, 1988).

Our primary research question is whether the NJ adults examined in this study, with their previous experience in English, would have an advantage over the NJ children in terms of mastering phonological aspects of English, and whether they would maintain their advantage over the children in the first two years of immersion. We examined segmental perception and production as well as global foreign accent rating as measures of phonological aspects of L2 learning. Based on the above studies, we selected the discrimination of /s/-/θ/ contrast for the segmental perception, and examined the production of /s/, /θ/, and /f/ for the segmental production.

The study is organized as follows. First, the participants will be described in Section 2. Sections 3 to 5 describe the results of 3 experiments: segmental perception (Experiment 1), segmental production (Experiment 2), and overall foreign accent (Experiment 3). The questions of interest were whether the NJ adults would outperform the NJ children, either at the first time of testing or the second time of testing one year later, and whether either group (or both) would show improvements over one year of residence in the U.S.

2. Participants

Sixteen NJ adults and 16 NJ children who were living in the U.S. participated. The NJ participants planned to return to Japan after several years in the U.S., and therefore were not immigrants. Thirty of the NJ participants were living in Houston or Dallas, Texas, two in Birmingham, Alabama at the time of testing. Sixteen adults and 16 children who spoke American English as their native language (NE) also participated as a control group. The NE adults and children lived in Birmingham, Alabama, and spoke no language other than American English. All 64 participants[3] were tested on two occasions separated by about one year.

The mean characteristics of the participants are summarized in Table 1. The NJ participants had been living in the U.S. for an average of 0.5 years when tested the first time (T1), and for an average of 1.6 years when tested the second time (T2). The interval between T1 and T2 was at least one year for all participants. The NJ and NE children had mean ages of about 10 years at T1; the NJ and NE adults had mean ages of about 40 years. In both NJ and NE groups, the adult participants were the parents of at least one of the child participants. There were approximately equal numbers of female and males in each of the 4 groups.

All participants responded to a language background questionnaire. No participant reported speech, language or hearing problems. All of the NJ adults had begun to take English in Japan between the ages of 11 and 13 years (mean 12;2). The NJ children, on the other hand, had no experience in English except for one child who took English for two years in Japan.[4] The NJ children began attending English-speaking schools upon arriving in the U.S. All participants were tested in a quiet room at their homes or at the University of Alabama at Birmingham.

3. Experiment 1: Segmental perception

In Experiment 1, the perception of English fricatives was examined as a measure of L2 learning at the segmental level. All 64 participants' perception of the contrast between /s/ and /θ/ was examined in this experiment.

---

3. Eighty-three NE and NJ adults and children (approximately 20 participants in each group) participated at T1. However, only 16 each of Japanese adults and children were available for testing at T2 because of relocations. Sixteen each of NE adults and children were retained in the study in order to have an equal cell size in all groups.
4. In Japan, students typically start taking English in junior high school (at around age 12). The only NJ child who had taken English was the oldest among the group (NJ16 in Table 5), and therefore he had taken English in junior high school for two years prior to coming to the U.S.

Table 1. *Characteristics (mean, standard deviation, range) of the NE and NJ participants. LOR = Length of residence in the U.S. in years. Age = Chronological age in years;months. Standard deviations are in parentheses.*

	N (male/ female)	Mean age at T1 range	Mean LOR range		Mean time between T1 and T2 range
			T1	T2	
NE adults	7m, 9f	40;4 (4;8) 33;11–49;11	–	–	1.3 (0.08) 1.0–1.4
NE children	10m, 6f	10;7 (2;1) 7;0–13;10	–	–	1.3 (0.08) 1.0–1.4
NJ adults	8m, 8f	39;11 (3;9) 35;1–48;2	0.5 (0.2) 0.1–1.0	1.6 (0.3) 1.1–2.3	1.2 (0.15) 1.0–1.4
NJ children	9m, 7f	9;11 (2;5) 6;6–14;0	0.4 (0.2) 0.1–1.0	1.6 (0.3) 1.1–2.3	1.1 (0.14) 1.0–1.3

3.1. *Stimuli*

Naturally produced English consonant-vowel (CV) syllables served as stimuli. An adult male native speaker of American English produced seven CV syllables (/lɑ/, /rɑ/, /wɑ/, /bɑ/, /sɑ/, /θɑ/, /vɑ/) multiple times in a carrier phrase "*Then I saw* /Cɑ/ *there*".[5] The recording was made using a Sony DAT recorder, and then digitized at 22.05 kHz with 16-bit amplitude resolution. Five clearly produced tokens of each syllable were edited out of the carrier phrase and normalized for peak intensity (50 % of the full scale).

3.2. *Procedure*

A categorial discrimination test, which has been used in previous L2 research (Flege 2003; Flege and MacKay 2004; Flege et al. 1999; Guion et al. 2000; Wayland and Guion 2003), was employed because of its suitability as a test of category formation. Five consonant contrasts (/b/–/s/, /b/–/v/, /r/–/l/, /r/–/w/, /s/–/θ/) were examined. The NJ adults and children were expected to have little difficulty discriminating English /b/ and /s/, the control contrast, because the Japanese phonemic inventory includes both /b/ and /s/ (Vance 1987) and because these two phonemes differ in a wide array of acoustic and phonetic

5. Sounds other than /s/ and /θ/ were also recorded to examine other contrasts. They were reported in Aoyama et al. (2004).

features. The /b/–/v/, /r/–/l/, and /s/–/θ/ contrasts were examined because it has been reported that these contrasts are difficult for NJ speakers (Guion et al. 2000; Lambacher et al. 2001; Singh and Black 1966; Yoshida and Hirasaka 1983). This study will focus on the results for /s/-/θ/.

All trials contained 3 stimuli separated by an inter-stimulus interval of 0.5 s. Each contrast was tested by 8 change and 8 no-change trials (4 trials for each consonant in the pair). The correct response to a change trial was a button ("1", "2", or "3") indicating the position of the odd item out, which occurred with near equal frequency in all 3 possible serial positions. The correct response to no-change trials, which consisted of 3 physically different tokens of a single category, was a fourth button marked "no" (indicating that no one item differed from the other two in terms of categorical identity). The change trials tested the participants' ability to distinguish consonants drawn from different categories. The no-change trials, on the other hand, tested the participants' ability to ignore audible but phonemically irrelevant within-category variation. Trials testing all 5 contrasts were presented together in one randomized block. The same randomized order was used for all the participants. The interval between each response and presentation of the next triadic trial was 1 s.

All 64 participants described in Section 2 were tested individually in a quiet room at their homes or at the University of Alabama at Birmingham. The stimuli were presented via headphones using a laptop computer. Before the experiment began, a comfortable stimulus presentation level was established. The participants adjusted the level as stimuli similar to those used in the test were presented via headphones. Once a presentation level was selected, it was fixed for the remainder of the experiment.

To ensure that that the participants understood the task, they next took part in a practice session using /wɑ/ and /sɑ/ stimuli. The participants were required to respond correctly to at least 9 out of 10 practice items before proceeding to the actual experiment. The practice block was repeated up to 4 times, as needed. Feedback was provided during the practice. Feedback was not presented during the experiment, which began with 10 trials that were presented for familiarization and were not analyzed.

3.3. *Analyses*

To control for response bias, A' values were calculated for each participant based on the proportion of "hits" (correct selections of the odd item out in change trials) and "false alarms" (incorrect selections of an odd item out in no-change trial) obtained for each contrast using the formula provided by Snodgrass et al. (1985). There were a maximum number of 8 hits and 8 false alarms for each contrast. 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

$$A' = 0.5 + \frac{(H - FA) \times (1 + H - FA)}{(4 \times H) \times (1 - FA)}.$$

If *FA* exceeded H, then

$$A' = 0.5 - \frac{(FA - H) \times (1 + FA - H)}{(4 \times FA) \times (1 - H)}.$$

A score of 1.0 indicated correct responses to all 16 trials testing a contrast, and a score of 0.5 indicated a theoretically defined chance level of response (see Snodgrass et al. 1985).

3.4. *Results*

As shown in Figure 1, the NJ adults' and children's scores for the control contrast, /b/–/s/, were high ($> .90$) at both T1 and T2, indicating that they understood the task. The *A'* scores for the /b/–/s/ contrast were submitted to an ANOVA in which Language (English vs. Japanese) and Age (adult vs. child) served as between-subject variables, and Time (T1 vs. T2) served as a within-subjects variable. The three-way interaction was non-significant ($F(1,60) = 2.55$, $p > .1$) as well as all of the two-way interactions ($F(1,60) = .13$ to 2.36, $p > .1$), and the main effects of Age ($F(1,60) = .03$, $p > .1$) and Time ($F(1,60) = .39$, $p > .1$). However, the difference in scores obtained from the NJ and NE participants resulted in a significant main effect of Language (NE mean = .97, NJ mean = .92, $F(1,60) = 8.23$, $p < .01$, $\eta_p^2 = .11$). This small but significant difference was unexpected because, as mentioned, the Japanese inventory includes both /b/ and /s/. We speculate that it arose from small differences in the phonetic implementation of /b/ and /s/ in Japanese and English.

Figure 1 also shows the results for /s/–/θ/. The NE adults obtained high (mean $> .94$) scores for /s/–/θ/ at both T1 and T2. The NE children's A' score averaged .86 at T1, and .96 at T2. The NJ adults' scores were higher than the NJ children's both at T1 (adult mean = .79, child mean = .57) and T2 (adult mean = .86, child mean = .66). The NJ adults' and children's scores were lower than the NE adults' and children's both at T1 and T2 (NJ adults mean =.79 and.86; NJ children mean = .57 and .66; NE adults mean = .94 and .94; NE children mean = .86 and .96).

An ANOVA examining the A' scores for the /s/–/θ/ contrast yielded significant main effects of Language (NE mean =.92 vs. NJ mean = .72, $F(1,60) = 46.59$, $p < .01$, $\eta_p^2 = .43$) and Age (adult mean = .88 vs. child mean = .76, $F(1,60) = 15.71$, $p < .01$, $\eta_p^2 = .20$). The main effect of Time also reached

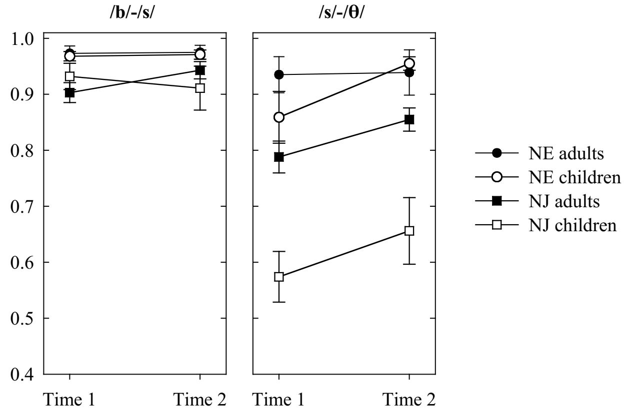

Figure 1. *Mean A′ scores obtained for the /b/–/s/ and /s/–/θ/ contrasts in Experiment 1. A score of 1.0 indicated correct responses to all 16 trials testing a contrast, and a score of 0.5 indicated a theoretically defined chance level of response.*

significance, even though mean difference between the T1 and T2 scores was small (T1 mean = .79 vs. T2 mean = .85, $F(1,60) = 6.63$, $p < .05$, $\eta_p^2 = .09$). The Time factor did not interact significantly with other factors ($p > .1$).

The three-way interaction was non-significant ($F(1,60) = .65$, $p > .1$). Just one of the two-way interactions, that between Language and Age, reached significance ($F(1,60) = 8.85$, $p < .01$, $\eta_p^2 = .12$). Simple effects tests revealed that the interaction arose because age exerted a stronger effect on the NJ than on the NE participants' discrimination scores. *F*-tests examining scores averaged over T1 and T2 revealed that the NJ adults obtained significantly higher scores than the NJ children (adults mean = .82, children mean = .62, $F(1,30) = 25.38$, $p < .01$, $\eta_p^2 = .46$), whereas the NE adults' and children's scores (mean = .94 vs. .91) did not differ significantly ($F(1,30) = .47$, $p > 0.1$). Tests of the simple effect of Language revealed that the NE adults obtained significantly higher scores than the NJ adults (NE mean = .94, NJ mean = .82) ($F(1,30) = 7.77$, $p < .01$, $\eta_p 2 = .20$) and the NE children obtained higher scores than the NJ children (mean = .91 vs. .62) ($F(1,30) = 45.92$, $p < .01$, $\eta_p^2 = .60$).

In summary, Experiment 1 confirmed that NJ learners of English have difficulty discriminating English /s/–/θ/ (Brannen 2002; Guion et al. 2000; Lambacher et al. 2001; Singh and Black 1966; Yoshida and Hirasaka 1983). In addition, it was found that the NJ children's /s/–/θ/ discrimination scores were significantly lower than the NJ adults', both 0.5 and 1.6 years after their arrival in the U.S.

4. Experiment 2: Segmental production

The purpose of Experiment 2 was to examine the production of English fricatives as another measure of L2 learning at the segmental level. All 64 participants' production of fricatives (/s/, /θ/, and /f/) was examined in this experiment.

4.1. *Elicitation procedures*

The following 26 English words were elicited 3 times each from each participant: *book, bug, cage, dog, eat, egg, eight, feet, fish, food, foot, hug, leaf, light, neck, read, six, shoe, sock, thousand, think, vase, voice, watch, wing, write*. These words contained a variety of English vowels and consonants, and included no consonant clusters in syllable-initial position. They are all frequently occurring words and are likely to have been familiar to the NJ participants.

The participants wore a head-mounted Shure microphone (Model SM 10A) connected to a Sony DAT tape recorder. The 26 words were elicited 3 times in different random orders for each participant. At the first elicitation, the participants saw a picture on the screen of a laptop computer and heard the corresponding word via a loudspeaker. An auditory model was not usually provided to elicit the second and third tokens of the test words. The experimenter played out the auditory model of the word only when the participant was unable to say the word in response to the corresponding picture. An equivalent word in Japanese was displayed in Japanese orthography in addition to the picture in order to reduce uncertainty as to what word was to be produced. Each word was elicited 3 times to examine possible differences between imitated and nearly spontaneous productions.

4.2. *Stimuli*

The words used to examine the participants' production of fricatives were *six*, *think*, and *fish*. The participants' first and third productions of these words were selected for analysis. The first productions always followed an auditory model whereas the third productions were usually produced without the need for an auditory model and, thus, are likely to be representative of spontaneous or uncued productions.

The selected tokens of *six* /sɪks/, *think* /θɪŋk/, and *fish* /fɪʃ/ were digitized (22.05 kHz, 16-bit amplitude resolution) and then normalized for peak intensity (50 % of the full scale). Words beginning with /f/ were examined, in addition to those beginning with /s/ and /θ/, because it has been reported that children learning English as their L1 may have difficulty with /f/ (Ingram et al. 1980;

Moskowitz 1975; Smith 1973; Velleman 1983, 1988). These 3 words were selected because all of the participants, including the NJ children, were familiar with them, and also all included the same vowel.

A total of 768 CV stimuli (4 groups x 16 participants x 3 words x 2 tokens x 2 times of testing) were derived from productions of the test words. After editing, the tokens derived from *six*, *think*, and *fish* sounded like /sɪ/, /θɪ/ and /fɪ/, respectively.[6] The editing consisted of removing everything following complete constriction of the consonant following the [ɪ] vowel, then reducing the intensity in the final 30 milliseconds of each stimulus from 100 % to 0 %. This procedure was adopted to reduce the effect of perceived lexical identity on the identification of the word-initial consonants of interest (see Flege et al. 1996).

4.3. *Procedure*

The CV stimuli were presented to 12 monolingual native speakers of American English (6 males, 6 females) with a mean age of 26 years. All listeners passed a pure-tone hearing screening at 20 dB from 250 to 8000 Hz in both ears.

Listener judgments were used to evaluate how well the segments of interest had been produced. To begin, an NE speaker with phonetic training listened to all the stimuli and devised a list of orthographic response alternatives, which, in his opinion, would allow listeners without training in phonetics to report the consonants they heard when evaluating the stimuli. These alternatives were ⟨*f*⟩, ⟨*th*⟩, ⟨*s*⟩, ⟨*sh*⟩, ⟨*h*⟩, ⟨*fw*⟩, ⟨*hw*⟩, ⟨*d*⟩ and ⟨*t*⟩. Subsequently, 12 native English adults used these response categories to classify each production stimulus by clicking one of nine buttons shown on a computer screen. Three keywords that began with /f θ s ʃ h d t/ were given above each key on the screen to help orient the listeners (e.g., *thank, thought, theme* for /θ/). The experimenter demonstrated how [fw] and [hw] might sound.

Once the classification experiment began, the listeners were permitted to replay stimuli but they could not change a response once given. The 32 adults' and the 32 children's productions were presented in separate counterbalanced blocks, each with 384 tokens (32 participants × 3 consonants × 2 repetitions × 2 times of testing). The tokens presented in each block were randomly presented; each listener received a unique randomized order.

6. Three tokens had an overlapping noise on the initial consonant and were considered as missing data (one token of *fish* by an NE child at T1, one token of *fish* by an NE adult at T2, and one token of *six* by an NJ child at T1). These tokens were presented among the other stimuli, but responses to these stimuli were excluded when the analyses were conducted.

4.4. *Analyses*

A total of 4,608 identifications (16 participants × 4 groups × 3 consonants × 2 repetitions × 12 listeners) were obtained for analyses. The dependent variable for the analysis was the "intelligibility score". This score represented the percentage of times that the listeners (maximum = 12) heard the consonant in each CV syllable as intended. (For example, if an /s/ token was heard as /s/ by 6 out of 12 listeners, the intelligibility score was 50 %.) A total of 12 intelligibility scores were calculated for each speaker (3 consonants × 2 repetitions × 2 times of testing). Only 11 intelligibility scores were calculated for three participants due to missing data.

A preliminary analysis revealed that the two repetitions (first vs. third) did not differ in the intelligibility scores they received. An ANOVA on intelligibility scores indicated that scores for these two sets of tokens did not differ significantly ($F(1,60) = .08$, $p > 0.1$) and the effect of Token (first vs. third) did not interact significantly with any other factor ($F(1,60) = .00$ to 2.41, $p > .1$). Thus, the scores for the first and third repetitions were averaged in the analyses presented here.

4.5. *Results*

The boldfaced values in Tables 2 and 3 indicate the percentage of times that productions of /s/, /θ/ and /f/ were identified as intended by the listeners. The NE adults obtained high (> 95 %) intelligibility scores for /s/ and /θ/. However, the NE adults' productions of /f/ were heard as intended just 78 % and 80 % of the time at T1 and T2, respectively. The NE children also obtained higher intelligibility scores for /s/ and /θ/ (> 84 % at both T1 and T2) than for /f/ (60 % at T1, 59 % at T2). The NJ children showed improvement from T1 to T2 for all 3 consonants: 15 % improvement for /f/, 11 % for /s/, and 30 % for /θ/. For the NJ adults, there was a small improvement for /θ/ (6 %). The NJ adults' intelligibility scores for /s/ were lower at T2 than at T1 (89 % at T1, 75 % at T2), and about the same for /f/ at both T1 and T2 (71 % at T1, 70 % at T2).

The intelligibility scores were submitted to an ANOVA in which Language (English vs. Japanese) and Age (adult vs. child) served as between-subjects factors, and Time (T1 vs. T2) and Consonants (/s/, /θ/, /f/) served as within-subjects factors. It yielded significant main effects of Language (NE mean = 85 %, NJ mean = 68 %, $F(1,60) = 31.25$, $p < .01$, $\eta_p^2 = .34$), Age (adults mean = 81 %, children mean = 72 %, $F(1,60) = 7.91$, $p < .01$, $\eta_p^2 = .12$), Time (T1 mean = 74 %, T2 mean = 79 %, $F(1,60) = 8.33$, $p < .01$, $\eta_p^2 = .12$) and Consonant (/f/ mean = 68 %, /s/ mean = 88 %, /θ/, mean = 74 %, $F(2,120) = 26.26$, $p < .01$, $\eta_p^2 = .30$). Just 3 interactions reached significance; Age × Time

Table 2. *Confusion matrices for the NE participants' productions in Experiment 2. Percentages smaller than 1 % are not shown. Each percentage is based on 384 identifications (16 participants × 2 repetitions × 12 listeners), except for two cases that were based on 372 identifications due to missing data.*

	Intended target	Listeners' identifications (%)								
		f	s	th	sh	h	fw	hw	d	t
NE adults, Time 1	/f/	**78**		20						
	/s/		**98**							
	/θ/	4		**95**						
NE adults, Time 2	/f/	**80**		19			2			
	/s/		**97**	3						
	/θ/	4		**95**						
NE children, Time 1	/f/	**60**	4	33			2			
	/s/		**89**	10						
	/θ/		10	**84**					5	
NE children, Time 2	/f/	**59**		37			3			
	/s/		**97**	1						
	/θ/	4	5	**87**			1			1

($F(1,60) = 13.85$, $p < .01$, $\eta_p^2 = .19$), Language × Consonant ($F(2,120) = 13.76$, $p < .01$, $\eta_p^2 = .19$), and Language × Age × Time ($F(1,60) = 7.16$, $p < .01$, $\eta_p^2 = .11$). All other two-way interactions and three-way interactions were non-significant ($F(1,60) = 0.43$ to 3.01, $p > .05$). The four-way interaction was also non-significant ($F(1,60) = 0.06$, $p > .05$).

The Language × Age × Time interaction was evaluated by simple effects tests. One source of the interaction was the fact that, of the 4 groups tested, only the NJ children's scores increased significantly from T1 to T2. Scores at the two times of testing were computed by averaging over the results for /s/, /θ/ and /f/. The simple effect of Time was non-significant for the NE adults, the NE children, or the NJ adults ($F(1,30) = .01$ to 1.09, $p > .1$), whereas the T1-T2 difference for the NJ children (T1 mean = 56 %, T2 mean = 74 %) did reach significance ($F(1,30) = 20.44$, $p < .01$, $\eta_p^2 = .58$).

Another source of the three-way interaction was the fact that the NE adults and children differed at both T1 and T2 whereas the NJ adults and children differed only at T1. The simple effect of Age was tested for all 4 Language × Time combinations. The NE adults obtained significantly higher scores than the NE children did at both T1 (adults mean = 91 %, children mean = 78 %, $F(1,30) = 8.47$, $p < .01$, $\eta_p^2 = .22$) and at T2 (adults mean = 91 %, children mean = 82 %, $F(1,30) = 9.21$, $p < .01$, $\eta_p^2 = .23$). The NJ adults obtained

Table 3. *Confusion matrices for the NJ participants' productions in Experiment 2. Percentages smaller than 1 % are not shown. Each percentage is based on 384 identifications (16 participants x 2 repetitions x 12 listeners), except for one case that was based on 372 identifications due to missing data.*

	Intended target	Listeners' identifications (%)								
		f	s	th	sh	h	fw	hw	d	t
NJ adults, Time 1	/f/	**71**		2			23	3		
	/s/		**89**	10						
	/θ/	11	29	**58**			2			
NJ adults, Time 2	/f/	**70**		4			22	3		
	/s/	5	**75**	20						
	/θ/	12	23	**64**						
NJ children, Time 1	/f/	**55**	1	10		1	25	7		
	/s/	7	**74**	17	1		1			
	/θ/	13	25	**39**	1		8	4	4	6
NJ children, Time 2	/f/	**70**		12			16	2		
	/s/	4	**85**	11						
	/θ/	13	16	**69**						2

significantly higher scores than the NJ children at T1 (adults mean = 72 %, children mean = 56 %, $F(1,30) = 8.66$, $p < .01$, $\eta_p^2 = .22$), but did not differ significantly from the NJ children at T2 (adults mean = 69 %, children mean = 74 %, $F(1,30) = 0.66$, $p > .1$).

The most likely source of the three-way interaction, however, was the fact that the NE and NJ adults differed at both T1 and T2, whereas the NE and NJ children differed only at T1. Tests of the simple effect of Language revealed that the NE adults obtained significantly higher scores than the NJ adults at T1 (NE mean = 91 %, NJ mean = 72 %, $F(1,30) = 24.50$, $p < .01$, $\eta_p^2 = .45$) and T2 (NE mean = 91 %, NJ mean = 69 %, $F(1,30) = 21.48$, $p < .01$, $\eta_p^2 = .42$). The NJ children's scores were significantly lower than the NE children's at T1 (NE mean = 78 %, NJ mean = 56 %, $F(1,30) = 12.35$, $p < .01$, $\eta_p^2 = .29$) but not at T2 (NE mean = 81 %, NJ mean = 74 %, $F(1,30) = 2.04$, $p > .1$).

The Language × Consonant interaction was also explored through simple effects tests. Scores for the NE and NJ participants were computed by averaging over the two times of testing (T1, T2) and the two age groups (adult, child) for each of the 3 consonants. The simple effect of Language was significant for /s/ and /θ/ (for /s/, $F(1,62) = 12.29$, $p < .01$, $\eta_p^2 = .17$, for /θ/, $F(1,62) = 56.34$, $p < .01$, $\eta_p^2 = .48$) but not for /f/ ($F(1,62) = .37$, $p > .1$). This indicates that the NE speakers' scores for /s/ and /θ/ ($M = 95$ % and 90 %, respectively) were significantly higher than the NJ speakers' scores for /s/ and /θ/ ($M = 81\,\%$ and $57\,\%$), but that the differences between the NE and NJ speakers' scores for /f/ did not differ statistically (mean = 69 % and 66 %, respectively).

Scores for each of the three consonants (/s/, /θ/, /f/) were computed by averaging over the scores obtained for the two age groups (adult, child) and the two times of testing. These tests revealed that the simple effect of Consonant was significant for both the NE and NJ groups (for NE, $F(1,62) = 36.28$, $p < .01$, $\eta_p^2 = .54$; for NJ, $F(1,62) = 12.30$, $p < .01$, $\eta_p^2 = .28$). Tukey post-hoc tests revealed that the scores for /f/ were significantly lower than the scores for /s/ and /θ/ for the NE participants. For the NJ participants, the scores for /f/ and /θ/ were significantly lower than the scores for /s/.

4.6. *Discussion of Experiments 1 and 2*

The results of Experiments 1 and 2 provided further evidence that Japanese speakers have difficulty distinguishing English /s/ and /θ/ in both production and perception (Cairns 1988; Guion et al. 2000; Lambacher 1994; Lambacher et al. 2001; Michaels 1974; Singh and Black 1966; Yoshida and Hirasaka 1983). In perception, the NJ participants' discrimination scores for /s/–/θ/ were significantly lower than those of age-matched NE speakers at both T1 and T2 (Experiment 1). Further, their productions of /θ/ were misidentified as /s/, and their /s/ productions were misidentified as /θ/ (Experiment 2). The NJ groups appeared to have little difficulty producing /f/ even though /f/ does not occur in Japanese. As mentioned in the introduction, a bilabial fricative [ɸ] appears in Japanese lexical items and in loanwords (Vance 1987). It is possible that the NJ speakers substituted Japanese [ɸ] for /f/ in the English word *fish*, and the NE listeners judged [ɸ] as ⟨*f*⟩, because [f] and [ɸ] share similar acoustic characteristics (Strevens 1960).

The results of Experiment 2 showed that the NJ adults' intelligibility scores for English /s/, /θ/ and /f/ were higher than the NJ children's scores at T1, but that the NJ children showed improvement in production of these consonants between T1 and T2, whereas the NJ adults showed no such improvement between T1 and T2. As a consequence, there was no difference between the NJ adults and children at T2. Additionally, the NJ children did not differ from the NE children at T2 on this segmental production measure. The NE children's production scores for /s/, /θ/ and /f/ were lower than the NE adults', indicating that the NE children's production of /s/, /θ/ and /f/ in later childhood differed from the NE adults'.

Additionally, it is worth noting that there was more variation among the NJ children's intelligibility scores at T1 compared to the NJ adults'. For instance, the NJ children's production /θ/ was identified as ⟨*sh*⟩, ⟨*fw*⟩, ⟨*hw*⟩, ⟨*d*⟩, and

⟨*t*⟩, as well as the 3 common categories ⟨*f*⟩, ⟨*s*⟩, and ⟨*th*⟩ (see Table 3). Interestingly, this variation disappeared at T2, and the NJ children's pattern seems to be similar to the NJ adults'. This variation indicates that there were more individual differences among the NJ children at T1, and the differences disappeared as they gained more knowledge of English between T1 and T2. Wode (1989) hypothesized that young bilingual children (between ages 2 and 3) first need to establish "segment-based coding" (1989: 179) in their phonological memory before L1-to-L2 transfer would occur in their production. According to his observations, children's production might exhibit a wider variety before they establish segment-based coding of both L1 and L2 sounds. The variations in the NJ children's production at T1 and disappearance of variation at T2 might mean that the NJ children were in the process of coding segments in L2.

It was somewhat surprising that the intelligibility scores for /f/ were so low ($< 80\,\%$) for the NE adults and children (see Table 2). When /f/ tokens produced by the NE adults were misidentified, they were usually heard as /θ/. It is unlikely that the NE participants mistakenly produced [θ] for /f/ in *fish*. It is also unlikely that their misidentified segments were the result of articulatory problems, given that the participants did not report having any speech problems, and that their /s/ and /θ/ productions were almost always ($> 95\,\%$) heard as intended. It has been reported that [f] and [θ] have similar acoustic properties (Abbs and Minifie 1969; Strevens 1960) and that they are among the most difficult pairs to discriminate (Abbs and Minifie 1969; Miller and Nicely 1955; Singh and Black 1966). It is likely that the low intelligibility scores for /f/ among the NE speakers were due to the non-robust auditory difference between [f] and [θ] in the absence of semantic context.

The NE children's intelligibility scores for /f/ were even lower than the NE adults' (60 % vs. 79 %); their averaged intelligibility scores across T1 and T2 for /f/, /s/ and /θ/ were significantly lower than the NE adults'. These results indicate that the NE children's production in relatively late childhood (range 7 to 13 years) differed from the NE adults'. This finding is consistent with previous studies that have shown small but measurable adult-child differences in the perception and production of speech sounds. Lee et al. (1999) reported that children's production of native language vowels progressively approximate adult norms between the ages of 5 and 18 years. Other studies have documented differences in categorization and identification of contrasts between adults and children up to the ages of 12 years (Hazan and Barrett 2000; Johnson 2000; Pursell et al. 2002). Thus, the present findings agree with previous research in showing that L1 speech development continues into late childhood.

5. Experiment 3: Global foreign accent

The purpose of Experiment 3 was to examine changes in global foreign accent in NJ adults and children from T1 to T2 as one of the measures for phonological learning of the L2. English words produced by all 64 participants (16 participants $\times$ 4 groups) were examined in this experiment.

5.1. *Stimuli*

Thirteen of the 26 words that were originally recorded (*eight, neck, read, six, book, dog, light, wing, egg, fish, leaf, watch, write*) were selected for analysis here because all the participants were able to produce them without an auditory cue for the second and third elicitations. The participants' third production of each word were digitized (22.05 kHz, 16-bit resolution), then normalized for peak intensity (50 % of the full scale). The third token of each word was examined because these tokens were considered to be closest to spontaneous production.

After normalization, the words were concatenated into the following three strings:

(1) *eight* *neck* *read* *six*
(2) *book* *dog* *light* *wing*
(3) *egg* *fish* *leaf* *watch* *write*

The strings provided a relatively good indication of each participant's pronunciation of English in that the words in each string contained a variety of initial consonants, final consonants and vowels. No string had any sequence of words that could be interpreted as a meaningful phrase or sentence (e.g., *eight fish* or *read book*). There were approximately 200 milliseconds of silence between the words. Because the words were originally produced in isolation, each word in the concatenated word-strings was produced under its own intonational contour, giving the overall impression of a list-like intonation.

5.2. *Listener judgments*

The three word-strings produced by the 64 participants were rated by adult native speakers of English for overall degree of perceived foreign accent. The 16 listeners (9 males, 7 females) had a mean age of 33 years, and spoke no language other than American English. None of the listeners was among the 64 participants (i.e., speakers) described in section 2. Although the listeners were recruited separately at different times for listener judgments in Experiments 2 and 3, there were two people who participated as listeners in both experiments.

It is unlikely that the previous experience in participating as a listener affected their judgments, because the two kinds of judgments are very different and the listeners returned almost one year after they first participated. All listeners passed a pure-tone hearing screening at octave frequencies between 250 and 4000 Hz at 20 dB in both ears and did not have any special training in speech or language. Seven of the 16 listeners were from Alabama, and 9 were from Texas. Native English speakers from Alabama and Texas were recruited as listeners because the 64 NE and NJ speakers were living either in Alabama or Texas at the time they were recorded.

Each listener participated individually in a sound booth at the University of Alabama at Birmingham. They participated in two sessions held on separate days, each lasting about 75 minutes. The three word-strings were presented in three separate blocks, and adult and child productions were presented in separate blocks. This yielded 6 blocks of stimuli in all (3 adult blocks, 3 child blocks). Each block consisted of 64 stimuli (2 groups [NE/NJ adults or NE/NJ children] $\times$ 16 participants $\times$ 2 times of testing). Half of the listeners rated blocks of stimuli produced by children followed by blocks of stimuli produced by adults; the remaining listeners rated the adult block first. The order of the word-strings was counterbalanced within the adult or child blocks.

The word-strings were presented at a self-selected comfortable level via loudspeakers to each listener. The 64 stimuli in each block were randomly presented three times to each listener with an inter-stimulus-interval of 1 s. Different random orders were used in each of the six blocks for all 16 listeners. The listeners rated each word-string using a scale that ranged from 1 ("strongest foreign accent") to 9 ("least foreign accent").

Fifteen extra trials were given as practice in the beginning of each block. Judgments of these trials were not analyzed. The listeners could re-play a stimulus, but ratings could not be changed once given. The listeners were required to rate each word-string before moving on to the next word-string. The median rating of the three presentations was obtained for each word-string and for each listener, and these median ratings were used in analyses.

5.3. *Analyses*

The procedure described above yielded a total of 6,144 median ratings (4 groups $\times$ 16 participants $\times$ 2 times of testing $\times$ 3 word-strings x 16 listeners). Intraclass correlations were high in all 6 blocks ($R = .76$ to .90, $F(63, 945) =$ 78.50 to 197.95, $p < .001$), suggesting that there was a high level of inter-rater agreement (Glass and Hopkins 1996). To control for variation in both the stimuli (i.e., the words that had been produced by the participants) and the listeners who auditorily evaluated those stimuli, two types of scores were derived from the ratings. The "talker-based" scores were computed for each participant by averaging over the median ratings given by the 16 listeners. This yielded 384 talker-based scores (4 groups x 16 talkers x 3 word-strings x 2 times). There was no guarantee that the results obtained in analyses of the talker-based scores would generalize to other native English-speaking listeners, so "listener-based" scores were also computed. These scores were computed by averaging over the ratings each listener had given to the 16 participants in each group. This procedure yielded 384 listener-based scores (4 groups × 16 listeners × 3 word-strings × 2 times). A difference was not considered statistically meaningful unless it was significant in both the talker-based and listener-based analyses.

5.4. *Results*

The talker-based scores for each participant at T1 and T2 were submitted to a four-way ANOVA. Age (adults vs. children) and Language (English vs. Japanese) served as between-group variables, and Time of Testing (T1 vs. T2) and Word-string (3 word-strings) served as within-group variables. This analysis yielded main effects of Language (NE mean = 8.3, NJ mean = 3.4, $F(1,60) = 683.41$, $p < .01$, $\eta_p^2 = .91$), Time (T1 mean = 5.7, T2 mean = 6.1, $F(1,60) = 19.46$, $p < .01$, $\eta_p^2 = .24$) and Word-string (string 1 mean = 6.9, string 2 mean = 5.8, string 3 mean = 5.8, $F(2,120) = 11.94$, $p < .01$, $\eta_p^2 = .16$). It also yielded a significant two-way interaction, Language × Word-string ($F(2,120) = 8.46$, $p < .01$, $\eta_p^2 = .12$) and a significant three-way interaction, Language × Age × Time ($F(1,60) = 8.43$, $p < .01$, $\eta_p^2 = .12$). The four-way interaction was non-significant ($F(2,120) = 0.30$, $p > .1$).

The listener-based scores were also examined in (2) Age × (2) Language × (2) Time of Testing × (3) Word-string ANOVA. This analysis yielded main effects of Language ($F(1,60) = 855.99$, $p < .01$, $\eta_p^2 = .93$), Time ($F(1,60) = 161.01$, $p < .01$, $\eta_p^2 = .73$) and Word-string ($F(2,120) = 14.02$, $p < .01$, $\eta_p^2 = .19$).[7] It also yielded a significant two-way interaction, Language × Word-string ($F(2,120) = 9.93$, $p < .01$, $\eta_p^2 = .14$) and a three-way interaction, Language × Age × Time ($F(1,60) = 69.77$, $p < .01$, $\eta_p^2 = .54$). The four-way interaction was non-significant ($F(2,120) = .98$, $p > .1$). In the listener analysis, a two-way interaction Language × Age ($F(1,60) = 9.61$, $p < .01$, $\eta_p^2 = .13$), and a three-way interaction Language × Time × Word-string ($F(2,120) = 5.18$, $p < .01$, $\eta_p^2 = .08$) were also significant. These interactions will not be

7. The mean values obtained for each group in the listener analysis are the same as the talker analysis.

Table 4. *Mean foreign accent scores given by the NE listeners to English word-strings produced by the NE participants or by the NJ participants in Experiment 3. Standard deviations are in parentheses.*

	T1	T2	statistical significance between T1 and T2
NE adults	8.6 (0.5)	8.4 (0.4)	ns
NE children	8.2 (0.5)	8.2 (0.5)	ns
NJ adults	2.9 (1.0)	3.2 (0.8)	ns
NJ children	3.0 (1.0)	4.6 (1.9)	$p < .01$

examined further, because they were not significant in the talker-based analysis.

Table 4 summarizes the overall results of Experiment 3. The mean values for foreign accent scores obtained for each group in the talker- and listener-analyses were the same. Although only effects that were significant for both the talker- and listener-based analyses will be presented below, only the F- and p-values generated in analyses of the talker-based scores will be presented, in order to avoid redundancy.

The Language × Age × Time interaction was explored by averaging the scores for three different word-strings for each participant group at T1 and T2 (see Table 4). The simple effect of Time was significant for the NJ children (T1 mean = 3.0, T2 mean = 4.6, $F(1,15) = 17.90$, $p < .01$, $\eta_p^2 = .54$), but not for the NJ adults, the NE adults or the NE children ($F(1,15) = 3.54$ to 4.68, $p > .05$). This suggested that the NJ children's scores improved from T1 to T2, but not the NJ adults', NE adults' or NE children's scores.

The simple effect of Age was significant for the NE groups at T1 (adults mean = 8.6, children mean = 8.2, $F(1,30) = 7.33$, $p < .05$, $\eta_p^2 = .19$), suggesting that the NE children's scores were significantly lower than the NE adults' at T1. At T2, the difference between the NE adults' and children's scores was non-significant (adults mean = 8.4, children mean = 8.2, $F(1,30) = 2.44$, $p > .1$). The simple effect of Age was non-significant for the NJ participants at T1 (adults mean = 2.9, children mean = 3.0, $F(1,30) = .072$, $p > .1$). However, at T2, the NJ adults' scores were significantly lower than the NJ children's (adults mean = 3.2, children mean = 4.6, $F(1,30) = 8.57$, $p < .01$, $\eta_p^2 = .22$). This indicates that the NJ adults' and children's scores did not differ at T1, but that the NJ children's productions were judged to be less foreign-accented than the NJ adults' at T2.

The simple effect of Language was significant for all Age × Time combinations ($F(1,30) = 60.78$ to 811.10, $p < .01$, $\eta_p^2 = .67$ to .96), indicating that the NJ adults' and children's scores were significantly lower than the NE adults' and children's at both T1 and T2. This suggests that both the NJ adults' and children's production of English words remained noticeably foreign-accented both at T1 and T2.

The Language $\times$ Word-string interaction was explored by averaging scores over Age (adult, child) and Time (T1, T2) for the NE and NJ groups. For the NE participants, the differences among the scores for the three word-strings were non-significant ($F(2,62) = 2.18$, $p > .1$). However, differences among the scores for the three word-strings were significant for the NJ participants ($F(2,62) = 12.38$, $p < .01$, $\eta_p^2 = .28$). Tukey post-hoc tests indicated that the scores for word-string 1 were significantly higher than those for word-strings 2 and 3 (mean = 3.8 vs. 3.3 and 3.2, respectively). This finding suggested that foreign accent may be more evident in some words than others, and that the listeners' evaluations of non-native speech might change somewhat depending on the words that are being considered.

5.5. *Discussion*

The results of Experiment 3 showed that the NJ children's production of English words improved over one year of residence in the U.S., whereas the NJ adults' production did not change measurably. They also suggested that there was no significant difference in the degrees of foreign accent between the NJ adults and children at T1, even though the NJ adults had studied English in Japan prior to their arrival in the U.S. All NJ participants' scores, including the NJ children's at T2, remained significantly lower than the NE participants' scores, suggesting that the NJ participants' production was measurably foreign-accented. These findings suggest that children's production in L2 may improve more rapidly than adults', but that their production may still be noticeably foreign-accented after living in an L2-speaking environment for 1.6 years.

It may be noted in Table 4 that there was a greater variation among the NJ children than in other groups especially at T2. We conducted an ad-hoc analysis in order to explore individual differences among the NJ children. Table 5 shows the individual children's chronological age at T1 and foreign accent ratings at T1 and T2, rank-ordered by the NJ children's chronological age. NJ16 is the child who had two years of English education in Japan; he received the highest rating at T1.

Pearson's correlation analyses were conducted between the NJ children's chronological age at T1 and foreign accent scores at T1 and T2. The correlation between the NJ children's age and foreign accent scores was non-significant at T1 ($r(14) = -0.13$, $p > 0.1$). However, a significant negative correlation was found between the NJ children's chronological age and foreign accent scores at T2 ($r(14) = -0.65$, $p < 0.05$). This suggests that the children's age at the time of arrival in the U.S. may have influenced their learning of English speech

Table 5. *Individual NJ child's age and foreign accent ratings. Age at T1, chronological age at T1; FA T1, foreign accent rating at T1; FA T2, foreign accent rating at T2; difference, difference between T1 and T2 in foreign accent ratings*

	Age at T1	Gender	FA T1	FA T2	Difference
NJ1	6;6	F	3.9	5.7	1.8
NJ2	6;6	M	3.6	5.9	2.3
NJ3	7;0	F	2.3	5.0	2.7
NJ4	7;8	F	3.2	6.7	3.5
NJ5	8;6	M	3.2	4.8	1.6
NJ6	8;9	F	2.5	5.5	3.0
NJ7	8;11	F	2.9	6.6	3.7
NJ8	9;2	M	3.1	3.7	0.6
NJ9	9;5	M	3.0	3.9	0.9
NJ10	9;7	F	3.8	7.8	4.0
NJ11	11;8	F	4.0	5.4	1.4
NJ12	12;0	M	1.2	1.9	0.7
NJ13	12;6	M	2.8	2.4	−0.4
NJ14	12;7	M	1.8	2.1	0.3
NJ15	12;11	M	2.4	2.1	−0.3
NJ16	14;0	M	4.7	4.1	−0.6

in that the younger children in the group were judged less foreign-accented than the older children at T2. There was also a significant correlation between chronological age and the increase in foreign accent scores from T1 to T2 (the "difference" score) ($r(14) = -0.72$, $p < 0.01$).

A somewhat unexpected finding in Experiment 3 was a difference in scores among the 3 word-strings produced by the NJ participants. The higher mean scores on word-string 1 suggested that the NE listeners judged word-string 1 to be less foreign-accented than the other two word-strings. All words examined here were monosyllabic, and each word-string contained a variety of initial and final consonants. One possible reason for the difference in scores may be the initial /w/. Word-strings 2 and 3 both contained words beginning in /w/ (*wing, watch*), but word-string 1 did not. Both the Japanese and English inventories include a /w/, but Japanese /w/ has less lip-rounding than English /w/ (Vance 1987). The non-rounded /w/ may have contributed to a stronger foreign accent. It is not possible to know what caused the NE listeners to judge word-string 1 to be less foreign-accented than word-strings 2 and 3 from Experiment 3, and future research will be needed to investigate what cues listeners use to judge foreign accent.

6. Overall discussion

The primary aim of this study was to investigate phonetic measures related to aspects of L2 phonological acquisition in Japanese adults and children learning English during the first years of immersion in an L2-speaking environment. The primary question addressed by this study was whether the NJ adults, who had previous experience in English prior to arriving in the U.S., would have advantage over the Japanese children in learning phonological aspects of an L2 and whether they would maintain their advantage over children.

Overall, the findings of the current study seem to suggest that the NJ adults indeed had an advantage over the NJ children initially, although the children improved their skills to a greater extent than the adults did, with the result that adult-child differences disappeared or became smaller over a one year time frame. The results of this study indicated that, at T1 (approximately 6 months after their arrival), the NJ adults indeed had more accurate segmental perception (/s/–/θ/) (Experiment 1), and higher intelligibility in segmental production (/s θ f/) (Experiment 2) than the NJ children did. On the other hand, the NJ adults' and children's scores on the degree of foreign accent did not differ significantly at T1 (Experiment 3). The NJ adults' foreign accent scores did not improve from T1 to T2, whereas the NJ children's scores improved. A year later, the NJ adults' perception scores for /s/–/θ/ remained significantly higher than the NJ children's. For the segmental production scores (Experiment 2), the NJ children did not differ significantly from either the NJ adults or the NE children due to their improvement over one year. The NJ children's global foreign accent scores were significantly higher than the NJ adults' scores. It seems that younger learners have advantage in learning phonological aspects of an L2.

Perhaps it was to be expected that the NJ adults would have an initial advantage because the NJ adults had received instruction in English in Japan, most focused on reading and writing, prior to their arrival in the U.S., while most of the children had not. This English education influenced some aspects of L2 learning more than others, because the NJ adults' foreign accent ratings did not differ from the NJ children's at T1 while their segmental perception and production scores were higher then the NJ children's at T1. We speculate this is because the NJ adults had more opportunities to read and write English in school than to listen and speak English. However, it is interesting to note that the NJ adults did outperform children in segmental perception and production, which is also related to listening and speaking. Perhaps the NJ adults' knowledge of the English language, such as the existence of /θ/, contributed to their segmental perception and production skills.

The results of this study is similar to those in Snow and Hoefnagel-Höhle (1978) who found that NE adults had an initial advantage over NE children younger than 10 on all measures except for the imitation task 6 months after their arrival in the Netherlands. As in the current study, Snow and Hoefnagel-Höhle (1978) found that the children's skills improved to a greater extent than the adults' after the initial few months of immersion in an L2-speaking environment. This is noteworthy especially because the NJ adults in this study had at least 6 years of English education prior to arrival whereas the NE adults in Snow and Hoefnagel-Höhle (1978) had no prior knowledge in Dutch. We believe that converging patterns between this study and Snow and Hoefnagel-Höhle (1978), which examined adult learners with differing previous knowledge of the target L2, provide more evidence for the age effects in L2 learning.

Another important issue that arises from both studies is why adults seem to enjoy only an *initial* advantage over children in learning L2. One factor may have been differing amounts of exposure to English received by the NJ adults and children while living in the U.S. The NJ children all attended English-medium schools, whereas some of the NJ adults reported little opportunity or need to speak English on a daily basis.

Psychological factors, such as motivation and attitudes, may also have played an important role (see Okamura-Bichard 1985; Snow and Hoefnagel-Höhle 1977). For example, Jia and Aaronson (1999) hypothesized that younger arrivals (i.e., children) in an L2-speaking country use their L2 more than older arrivals (i.e., adults) because they switch their dominant language from their L1 to L2, whereas older arrivals learn their L2 while maintaining their L1 as the dominant language. Jia and Aaronson reported that a language preference switch from L1 to L2 by young children could take place within one year after arrival in the host country. Attitude and motivation toward learning an L2 was not a primary focus of the current study, and therefore these aspects were not explicitly evaluated. We speculate that perhaps the NJ children were more motivated to learn an L2 than the NJ adults, because they needed to speak English at school. It is also possible that at least some of the children switched their language preference from Japanese to English, or felt equally comfortable using their L1 and L2 by T2 as hypothesized by Jia and Aaronson (1999). More positive attitude and preference for L2, in turn, may have resulted in more usage of L2 and, consequently, larger improvements in L2 than adults who may not have been as motivated as children.

Although this study provided new information regarding the first two years of naturalistic learning in an L2-speaking environment, it does have limitations. First, some words that were elicited might have been structurally more difficult for the NJ participants to produce (e.g., words with syllable-final consonant clusters). Familiarity with the test words might have also affected the NJ participants' productions. Second, the editing of the stimuli in Experiment 2 could have affected listener judgments. The editing was conducted in order to test syllable-initial fricatives in a uniform CV syllable without introducing lexical bias to the listeners. This method was employed because we believed that eliciting consonants in real words was better than eliciting nonsense syllables from the participants, especially from children. However, there may be better ways to evaluate segmental productions in L2 learning in future studies. Finally, this study did not explicitly study the participants' motivation or attitude toward learning English. Investigations into the causes of adult-child differences will be needed in future research.

In summary, the results of this study suggest that adults have an initial advantage in L2 learning and that children improve more rapidly than adults in L2 after living in an L2-speaking environment for only one year. Adults were found to have an initial advantage for L2 learning for the segmental perception and production. The NJ children's oral production scores in English improved significantly over one year of residence in the U.S., whereas the NJ adults' scores did not. The results also indicated that the NJ children's scores for perceived foreign accent and segmental perception remained significantly lower than the NE children of their age, whereas the scores did not differ significantly for both child groups in segmental production at T2. These findings suggest that adults may have an initial advantage in phonological aspects of L2 learning, presumably due to previous education in English, and that L2 learning occurs at different rates for different skills for adults and children.

Katsura Aoyama
Texas Tech University Health Sciences Center
<katsura.aoyama@ttuhsc.edu>
Susan G. Guion
University of Oregon
<guion@uoregon.edu>
James Emil Flege
National Research Council, Rome
<jeflege@yahoo.com>
Tsuneo Yamada
National Institute of Multimedia Education, Japan
<yamada@nime.ac.jp>
Reiko Akahane-Yamada
ATR Cognitive Information Science Laboratories, Japan
<yamada@atr.jp>

References

Abbs, Mary Skeel and Fred D. Minifie (1969). Effect of acoustic cues in fricatives on perceptual confusions in preschool children. *Journal of the Acoustical Society of America* 46: 1535–1542.

Aoyama, Katsura, James Emil Flege, Susan G. Guion, Reiko Akahane-Yamada and Tsuneo Yamada (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32: 233–250.

Asher, James J. and Ramiro Garcia (1969). The optimal age to learn a foreign language. *Modern Language Journal* 53: 334–341.

Brannen, Kathleen (2002). The role of perception in differential substitution. *Canadian Journal of linguistics* 47: 1–46.

Cairns, Ronald S. (1988). Hearing and articulating /s/ in English and Japanese. *Cross Currents* 15: 47–56.

Ervin-Tripp, Susan M. (1974). Is second language learning like the first? *TESOL Quarterly* 8: 111–127.

Flege, James Emil (1995). Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, Winifred Strange (ed.), 233–277. Timonium, MD: York Press.

— (2003). Methods for assessing the perception of vowels in a second language. In *Issues in Clinical Linguistics*, Elisabetta Fava and Alberto Mioni (eds.), 19–44. Padova: Padova University Press.

Flege, James Emil, David Birdsong, Ellen Bialystok, Molly Mack, Hyekyung Sung and Kimiko Tsuakda (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics* 34: 153–175.

Flege, James Emil and Ian R. A. MacKay (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26: 1–34.

Flege, James Emil, Ian R. A. MacKay and Diane Meador (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America* 106: 2973–2987.

Flege, James Emil, Naoyuki Takagi and Virginia Mann (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *The Journal of the Acoustical Society of America* 99: 1161–1173.

Glass, Gene V. and Kenneth D. Hopkins (1996). *Statistical Methods in Education and Psychology* (third edition). Boston, MA: Allyn and Bacon.

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

Hazan, Valerie and Sarah Barrett (2000). The development of phonemic categorization in children aged 6- 12. *Journal of Phonetics* 28: 377–396.

Ingram, David (1978). The production of word-initial fricatives and affricates by normal and linguistically deviant children. In *Language Acquisition and Language Breakdown: Parallels and Divergencies*, Alfonso Caramazza and Edger B. Zurif (eds.), 63–85. Baltimore: Johns Hopkins University Press.

Ingram, David, Lynda Christensen, Sharon Veach, and Brendan Webster (1980). The acquisition of word-initial fricatives and affricates in English by children between 2 and 6 years. In *Child Phonology*, vol 1. *Production*, Grace H. Yeni-Komshian, James F. Kavanagh and Charles Ferguson (eds.), 169–192. London: Academic Press.

Ioup, Georgette. (2005). Age in second language development. In *Handbook of Research in Second Language Teaching and Learning*, Eli Hinkel (ed.), 419–435. Mahwah, NJ: Lawrence Erlbaum Associates.

Jia, Gisela and Doris Aaronson (1999). Age differences in second language acquisition: The dominant language switch and maintenance hypothesis. *Proceedings of the 23rd Annual Boston University Conference on Language Development* 23: 301–312. Somerville, MA: Cascadilla Press.

Johnson, Carole E. (2000). Children's phoneme identification in reverberation and noise. *Journal of Speech, Language, and Hearing Research* 43: 144–157.

Lambacher, Stephen G. (1994). Acoustical properties of the English voiceless fricatives [θ] and [s]: An analysis of American and Japanese production. Technical Report 94-5-001. Aizu-Wakamatsu, Japan: University of Aizu Center for Language Research.

Lambacher, Stephen, William Martens, Brian Nelson and Jonathan Berman (2001). Identification of English voiceless fricatives by Japanese listeners: The influence of vowel context on sensitivity and response bias. *Acoustic Science and Technology* 22: 334–342.

Lee, Sungbok, Alexandros Potamianos and Shrikanth Narayanan (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America* 105: 1455–1468.

Michaels, David (1974). Sound replacements and phonological systems. *Linguistics* 126: 69–81.

Miller, George A. and Patricia E. Nicely (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27: 338–352.

Moskowitz, Breyne Arlene (1975). The acquisition of fricatives: A study in phonetics and phonology. *Journal of Phonetics* 3: 141–150.

Moyer, Alene. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition, 21*: 81–108.

Moyer, Alene. (2004). *Age, Accent and Experience in Second Language Acquisition*. Clevedon, England: Multilingual Matters.

Nissen, Shawn L. and Robert Allen Fox (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *The Journal of the Acoustical Society of America* 118: 2570–2578.

Okamura-Bichard, Fumiko (1985). Mother tongue maintenance and second language learning: A case of Japanese children. *Language Learning* 35: 63–89.

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

Phatate, D. Devenny and Helen Umano (1981). Auditory discrimination of voiceless fricatives in children. *Journal of Speech and Hearing Research* 24: 162–168.

Piske, Thorsten, Ian R.A. MacKay and James E. Flege (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215.

Pursell, Susan L., Lori A. Swanson, Mark S. Hedrick and Anna K. Nabelek (2002). Categorical labeling of synthetic /ɪ/ and /ɛ/ in adults and school-age children. *Journal of Phonetics* 30: 131–137.

Singh, Sadanand and John W. Black (1966). Study of twenty-six intervocalic consonants as spoken and recognized by four language groups. *Journal of the Acoustical Society of America* 39: 372–387.

Smith, Neilson V. (1973). *The Acquisition of Phonology*. Cambridge, UK: Cambridge University Press.

Snodgrass, Joan Gay, Gail Levy-Berger and Martin Haydon (1985). *Human Experimental Psychology*. New York: Oxford University Press.

Snow, Catherine (1987). Relevance of the notion of a critical period to language acquisition. In *Sensitive Periods in Development: Interdisciplinary Perspectives*, Marc H. Bornstein (ed.), 183–209. Hillsdale, NJ: Lawrence Erlbaum.

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

— (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development* 49: 1114–1128.

Snow, Katherine (1963). A detailed analysis of articulation responses of 'normal' first grade children. *Journal of Speech and Hearing Research* 6: 277–290.

Strevens, Peter (1960). Spectra of fricative noise in human speech. *Language and Speech* 3: 32–49.

Tahta, Sonia, Margaret Wood and Kate Loewenthal (1981). Foreign accents: Factors relating to transfer of accent from the first language to a second language. *Language and Speech* 24: 265–272.

Tsukada, Kimiko, David Birdsong, Ellen Bialystok, Molly Mack, Hyekyung Sung and James Flege (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics* 33: 263–290.

Tsukada, Kimiko, David Birdsong, Molly Mack, Hyekyung Sung, Ellen Bialystok and James Flege (2004). Release bursts in English word-final voiceless stops: A cross-sectional and longitudinal study of Korean adults' and children's speech production. *Phonetica* 61: 67–83.

Vance, Timothy J. (1987). *An Introduction to Japanese Phonology*. New York, NY: State University of New York Press.

Velleman, Shelley L. (1983). *Children's Production and Perception of English Voiceless Fricatives.* Doctoral dissertation, University of Texas at Austin.

— (1988). The role of linguistic perception in later phonological development. *Applied Psycholinguistics* 9: 221–236.

Wayland, Ratree P. and Susan G. Guion (2003). Perceptual discrimination of Thai tones by naïve and experienced learners of Thai. *Applied Psycholinguistics* 24: 113–129.

Weinberger, Steven H. (1997). Minimal segments in second language phonology. In *Second-xxx Language Speech: Structure and Process,* Allan James and Jonathan Leather (eds.), 263–311. Berlin: Mouton de Gruyter.

Winitz, Harris (1981). Input considerations in the comprehension of first and second language. *Annals of the New York Academy of Sciences* 379: 296–308.

Wode, Henning (1989). Maturational changes of language acquisitional abilities. In *Variation in Second Language Acquisition, I: Discourse and Pragmatics; II: Psycholinguistic Issues*, Susan Gass, Carolyn Madden, Dennis Preston and Larry Selinker (eds.), 176–186. Clevedon, England: Multilingual Matters.

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

Yoshida, Kensaku and Fumiko Hirasaka (1983). The lexicon in speech perception. *Sophia Linguistica* 11: 105–116.