SSLA, **23**, 527–552. Printed in the United States of America.

THE EFFECT OF EXPERIENCE ON ADULTS' ACQUISITION OF A SECOND LANGUAGE

James Emil Flege and Serena Liu

University of Alabama at Birmingham

Previous research has suggested that child but not adult immigrants to the United States and Canada make regular progress learning English as their length of residence (LOR) increases. If children and adults received the same kind of second language (L2) input, such evidence would support the existence of a critical period for L2 acquisition. The present study compared groups of Chinese adults living in the United States who differed in LOR in order to assess the role of input in adults' naturalistic acquisition of an L2. We assessed the Chinese participants' identification of word-final English consonants (experiment 1), their scores on a 144-item grammaticality judgment test (experiment 2), and their scores on a 45-item listening comprehension test (experiment 3). The Chinese participants were assigned to one of four groups ($n = 15$ each) based on LOR in the United States and their primary occupation (students vs. nonstudents). Significantly higher scores were obtained for the students with relatively long LORs than for the students with relatively short LORs in all three experiments. However, the difference between the nonstudents differing in LOR was nonsignificant in each instance. The results suggested that the lack of an effect of LOR in some previous studies may have been due to sampling error. It appears that adults' performance in an L2 will improve measurably over time, but only if they receive a substantial amount of native speaker input.

This study was funded by a grant from the National Institute for Deafness and Other Communicative Disorders (DC-02892). The authors thank David Birdsong, Katsura Aoyama, and the anonymous *SSLA* reviewers for comments on a previous version of this article.

Address correspondence to: James Emil Flege, Division of Speech and Hearing Sciences, CH20 Room 119, 1530 Third Avenue South, Birmingham, AL 35294–2042; e-mail: jeflege@uab.edu.

© 2001 Cambridge University Press 0272-2631/01 $9.50

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

Many studies of second language (L2) acquisition have shown that "earlier is better" (see Long, 1990, for a review). For example, Flege, Yeni-Komshian, and Liu (1999) found that the later in life that native Korean speakers had arrived in the United States, the stronger their foreign accents in English sentences tended to be and the lower the scores they tended to obtain on a grammaticality judgment test. The most common explanation for so-called age effects such as these is the passing of a critical period. For example, the results that DeKeyser (2000) obtained on a grammaticality judgment test led him to conclude that:

> Somewhere between the ages of 6–7 and 16–17, everybody loses the mental equipment required for the abstract patterns underlying a human language. . . . It may be that the severe decline of the ability to induce abstract patterns implicitly is an inevitable consequence of fairly general aspects of neurological maturation (pp. 518–519).

Others have also claimed that neurologically based maturational constraints exist for the learning of L2 morphology and syntax (e.g., Johnson & Newport, 1989; Patkowski, 1980) and for the pronunciation of an L2 (e.g., Patkowski, 1990; Scovel, 1988).

It is important to note, however, that other explanations have been offered for why early bilinguals, whose first extensive exposure to the L2 occurred in childhood, perform more like monolingual native speakers of the target L2 than do late bilinguals, whose first extensive exposure to the L2 occurred in adolescence or adulthood. Early bilinguals might have a stronger personal or economic motivation to learn an L2 well than late bilinguals do (e.g., Gardner, 1991). Finally, late bilinguals' native language (L1) systems may influence the developing L2 system to a greater extent than early bilinguals' because late bilinguals' L1 systems are more fully developed when L2 learning begins. According to the Speech Learning Model (Flege, 1995), as L1 phonetic categories develop through childhood and into adolescence, they become more likely to perceptually assimilate L2 speech sounds on a long-term basis, thereby blocking the formation of new categories for L2 vowels and consonants.

Another explanation that might be offered for age effects on L2 acquisition pertains to input. Early bilinguals may generally receive more native-speaker L2 input or less nonnative L2 input than late bilinguals do. For example, Flege, Yeni-Komshian, and Liu (1999) plotted the ratio of English to Korean use as a function of the age of arrival (AOA) of 240 Koreans to the United States. The ratios decreased in a linear fashion as AOA increased from 2 to 12 years, then remained constant at about 1.0 for participants having AOAs of 13–22 years. This indicated that early bilinguals used English more often than Korean, whereas late bilinguals tended to use their L1 and L2 equally.

An examination of U.S. census data suggested to Stevens (1999) that AOA-related differences attributable to "social and demographic considerations" may affect L2 performance. For example, child immigrants are usually enrolled in a school, where they interact frequently with native speakers of the L2,

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

whereas adult immigrants often enter the workplace, where they interact frequently with fellow native speakers of their L1. Early bilinguals are also more likely to marry L2 native speakers than late bilinguals are. Jia and Aaronson (1999) provided cross-sectional and longitudinal evidence suggesting that child immigrants receive more input from L2 native speakers than do individuals who immigrate during adolescence (see also Grenier, 1984; Oyama, 1982b).

There are two important obstacles to determining the basis for AOA (or, more generally, age) effects on L2 acquisition. One obstacle is that the variables used to select research participants are either imprecise, related only indirectly to the putative causative variable, or both. For example, chronological age is sometimes used as a surrogate for participants' state of neurological development at the time L2 learning begins. However, the rate at which parameters of neurological development change as a function of increasing age may vary, and may differ across individuals (Bates & Goodman, 1998). To take another example, theories that attribute age effects to the development of structures or representations in the L1 (e.g., Flege, 1995) usually do not measure the state of development of the L1 structures (representations) at the time of first exposure to an L2. Finally, studies that have investigated the role of input on L2 performance (Flege, 1993; Flege, Munro, & Skelton, 1992) have not quantified the amount or kind of input that L2 learners received from native speakers of the L2.

A second obstacle to understanding the bases of AOA (age) effects on L2 acquisition is that AOA is often confounded with factors that might themselves influence L2 performance (see Flege, 1987, for discussion). Flege (1998) reviewed studies that examined 240 native speakers each of Italian and Korean who had immigrated to North America (Canada and the United States, respectively). For both groups of immigrants, AOA was inversely correlated with length of residence ($r = -0.44$ and $r = -0.42$, respectively) and self-reported use of English ($r = -0.47$ and $r = -0.56$). That is, the later the participants arrived in a predominantly L2-speaking environment, the shorter period of time they tended to have lived there and the less they tended to use their L2 (English). Both a relatively short length of residence (LOR) and an infrequent use of the L2 might be expected to adversely affect L2 learning by late bilinguals.

A recent study demonstrated that confounds such as these are potentially important. Flege, Yeni-Komshian, and Liu (1999) examined 240 native Korean adults whose AOAs in the United States ranged from 3 to 23 years ($M = 12$ years), whose LORs ranged from 8 to 30 years ($M = 15$ years), and who had attended English-speaking U.S. schools for 0–19 years ($M = 10$ years). A strong correlation existed between AOA and the scores obtained on a grammaticality judgment test. However, when variables confounded with AOA (L1 use, L2 use, years of U.S. education, LOR) were controlled in matched subgroup analyses, the effect of AOA on the grammaticality judgment test scores became nonsignificant.

It is difficult or impossible to adequately control for some factors that are

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

typically confounded with AOA (see Flege, 1998, and Flege, Yeni-Komshian, & Liu, 1999, for discussion). However, it may be possible, in some instances, to measure or assess more precisely those factors that are thought likely to predict L2 performance. The aim of the present article was to provide greater insight into a variable that has been used to predict L2 performance in many previous studies, LOR. LOR is used to index the amount of L2 input. It is generally assumed that LOR in a predominantly L2-speaking environment (e.g., English for immigrants to the United States or certain places in Canada) is positively correlated with the amount of L2 input an immigrant has received in the L2.

Studies examining the overall strength of foreign accent in an L2 and the accuracy with which L2 phonetic segments have been produced support the conclusion that the effect of LOR on L2 performance is "inconsequential" (Moyer, 1999, p. 83). In a study by Oyama (1982b), the correlation between native Italian speakers' foreign accents in English and LOR became nonsignificant when the effect of AOA in the United States was statistically controlled. Flege (1988) did not observe a significant difference in foreign accents between groups of Chinese adults who had lived in the United States for averages of 1.1 and 5.5 years (see also Flege, Yeni-Komshian, & Liu, 1999; Piske, MacKay, & Flege, 2001). Flege (1993) did not observe significant differences in the production of English phonetic segments by groups of Chinese adults with average LORs in the United States of 1.2 and 5.1 years; nor did Flege, Munro, and Skelton (1992) observe any significant differences between groups of native Mandarin adults with LORs averaging 0.9 and 5.5 years, or between groups of native Spanish adults with average LORs of 0.4 and 9.0 years.

Studies examining L2 morphosyntax and sentence comprehension have also failed to show that L2 performance improves as LOR increases. Johnson and Newport (1989) administered a grammaticality judgment test to native Korean and Chinese speakers who were professors, research associates, and graduate students at an American university. A significant correlation was found to exist between the test scores and AOA, $r = -0.77$, $p < .01$, but not between the test scores and LOR, $r = .16$, *ns*. Oyama (1982a) assessed the ability of native Italian speakers to repeat English sentences presented in noise. A significant correlation was obtained between the sentence repetition scores and AOA when variation in the participants' LOR was partialled out, $r = -0.57$, $p < .01$. The later the participants had arrived in the United States, the fewer words they were able to recognize and repeat. A significant negative correlation was also obtained between LOR and repetition scores when AOA was partialled out, $r = -0.39$, $p < .01$. This finding was paradoxical, inasmuch as a relatively long residence in the United States was associated with relatively poor performance in English.

One possible explanation for why LOR has not been found to predict L2 performance in previous research is that amount of L2 input is not an important determinant of L2 performance. Another possible explanation, inasmuch as most participants in the studies cited earlier were late bilinguals, is that

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

L2 learning is constrained by a critical or sensitive period (DeKeyser, 2000; Patkowski, 1980, 1990; Scovel, 1988). Still another possibility is that the amount or the nature of L2 input, or both, does influence L2 learning, but that LOR provides a good index of L2 input only for certain individuals.

In support of this, a number of studies have shown an effect of LOR on L2 performance. Flege (1988) observed a significantly better pronunciation of English sentences by Spanish adults with an average LOR in the United States of 14.3 years than by adults with an average LOR of 0.7 years. Purcell and Suter (1980) observed a significant positive correlation between LOR and strength of foreign accents for 61 nonnative adults, and Riney and Flege (1998) observed a decrease in the strength of two Japanese students' foreign accents following a stay in the United States. Finally, Flege, Bohn, and Jang (1997) observed a significantly more accurate production of one English vowel (of the four vowels studied) by nonnative adults having an average LOR in the United States of 7.3 than by nonnative adults having an average LOR of 0.7 years; a modest effect of LOR on the perception of English vowels was also observed.

Another reason to think that LOR provides a good index of L2 input for some individuals but not others is the observation that immigrants to an L2-speaking country vary greatly in terms of how much they use the L2, and also in terms of the kind of L2 input they receive. Oyama (1982b) observed that social and economic factors might conspire to "prohibit or inhibit" L2 learning by individuals who would otherwise be capable of learning an L2. As a result, some adult immigrants who have lived for decades in an L2-speaking country never learn to speak the L2.

Still another reason to think that LOR provides a good index of L2 input for some individuals but not others is evidence of L2 learning differences between children and adults. The L2 proficiency of most immigrant children increases rapidly over time (Cummins, 1981, 1991). LOR may be a better predictor of L2 acquisition by children than adults because children are often enrolled in schools where most students are native speakers of the L2. On the other hand, most adult immigrants who work in the home may have little contact with L2 native speakers, and adults who work outside the home may or may not come into regular contact with L2 native speakers.

The results of two studies suggest that there may be a closer relation between LOR and amount of L2 input for child than adult immigrants. Bialystok and Miller (1999) examined the relation between grammaticality judgment test scores and LOR. The correlation for late Chinese bilinguals with a mean LOR of 4 years (range: 1–6) was nonsignificant, whereas the correlation for early bilinguals with a mean LOR of 8 years (range: 5–18) was significant.[1] Snow and Hoefnagel-Höhle (1982a, 1982b) administered a battery of Dutch language tests to 51 native English speakers who were learning Dutch naturalistically in the Netherlands. Much the same results were obtained in tasks examining Dutch pronunciation, morphology, vocabulary, and sentence structure. Adults performed better than 8- to 10-year-old children did at the first time of testing, whereas a trend in the opposite direction was observed one year later. The

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

observed crossover might be attributed to the fact that the children received more Dutch input than the adults did because they were attending Dutch public schools.

It is important to consider the quality as well as the quantity of L2 input that is received. An immigrant to North America may be especially likely to use the L2 (usually English) as a lingua franca with nonnatives from diverse L1 backgrounds and also with native speakers of their own L1 if others are present who do not speak the L1. The use of English with other nonnative speakers is likely to result in inauthentic L2 input. If the input comes from speakers of a learner's own L1, it may reinforce the kind of errors the learners themselves might be apt to make.

Differences in formal education must also be considered when investigating the role of LOR on L2 acquisition. In addition to receiving input from L2 native speakers, L2 learners are also likely to receive formal instruction in the L2 while attending school in a predominantly L2-speaking environment. As mentioned earlier, Flege, Yeni-Komshian, and Liu (1999) obtained grammaticality judgment test scores for 240 Korean adults. Significant correlations were found to exist between the scores and AOA, $r(238) = –0.75$, $p < .01$, between the scores and LOR, $r(238) = .39$, $p < .01$, and between the scores and years of education in the United States, $r(238) = .79$, $p < .01$. When the effects of AOA and LOR were partialled out, the correlation between the scores and years of education remained significant, $r(236) = .36$, $p < .01$. That is, the longer the Koreans attended school in the United States, the higher their grammaticality judgment test scores tended to be. However, the correlation between the scores and AOA became nonsignificant when the effects of years of education and LOR were partialled out, $r(238) = –0.11$, $p = .09$. Additionally, the correlation between LOR and the scores became nonsignificant when the effects of years of education and AOA were partialled out, $r(238) = .01$, $p > .10$.

THE PRESENT STUDY

Some of the research cited earlier showed no relation between LOR and late bilinguals' performance in an L2, and one study revealed a negative correlation (increasingly poor performance as LOR increased). However, other studies showed a modest positive effect of LOR on performance in an L2. We inferred from several studies that a closer relation exists between LOR and L2 input for more child than adult immigrants. That is, LOR may provide a more accurate index of how much L2 input that children enrolled in an L2-speaking school have received than it provides for adult immigrants who are working outside the home (and, a fortiori, for adults working in the home). The ambiguity of the results obtained in previous studies examining LOR, when taken together with the theoretical importance of understanding the role of input for L2 acquisition, led us to reexamine the role of LOR in this study.

The participants examined here were adults from China who were learning English as an L2. All of the Chinese participants were living in Birmingham,

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

Alabama, at the time of testing and were associated in some way with the University of Alabama at Birmingham. The 60 Chinese participants and five native English controls participated in three experiments designed to assess different aspects of performance in English.

As in previous research, the Chinese participants' experience in English was indexed by their LOR in the United States. Participants in the short-LOR group had LORs ranging from 0.5 to 3.8 years, whereas those in the long-LOR group had LORs ranging from 3.9 to 15.5 years. To test the hypothesis that LOR provides a good index of L2 input for some L2 learners but not others, the participants in the two groups were assigned to one of two subgroups based on years of education in the United States. Half of the short-LOR participants ($n = 15$), as well as half of the long-LOR participants ($n = 15$), had been students during most or all of their stay in the United States. These participants were designated the *students*. The remaining half of the participants in both the short-LOR and the long-LOR groups, who were designated the *nonstudents*, had worked full-time during most or all of their stay in the United States.

Based on the research cited earlier as well as our own observations of Chinese individuals living in Birmingham (Flege, 1988, 1989, 1993; Flege, Munro, & Skelton, 1992; Flege & Wang, 1990), we thought it likely that the students had received more input from native English speakers (their professors and fellow students) than had the nonstudents, most of whom were full-time research assistants or scientists engaged in biomedical research. From this, we hypothesized that an LOR effect would be greater for the students than for the nonstudents in each of the three experiments that were carried out to assess performance in English.

In experiment 1, the participants identified naturally produced English stop consonants. Previous research has shown that speakers of Chinese languages identify word-final English stop consonants less accurately than native speakers of English do, especially if the stops do not have a release burst (Flege, 1989; Flege & Wang, 1990). Experiment 2 examined the Chinese participants' knowledge of basic aspects of English sentence structure. The grammaticality judgment test used here comprised 144 sentences, of which half were grammatical and half were ungrammatical. Finally, experiment 3 examined the Chinese participants' comprehension of English through the use of The University of Michigan English Language Institute Listening Comprehension Test (LCT).

GENERAL METHOD

Participants

A total of 62 participants who were born in China were recruited through personal contact and newspaper ads for this study. All participants were required to have self-reported normal hearing, to have lived in the United States for at least 0.5 years, to be between 21 and 45 years of age, and to have arrived in

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

Table 1. Characteristics of the four groups of Chinese participants

	Short LOR		Long LOR		
Characteristics	Nonstudents	Students	Nonstudents	Students	*M*
Gender	6m, 9f	8m, 7f	5m, 10f	7m, 8f	26m, 34f
Chronological age	30.7 (4.6)	29.5 (5.6)	34.9 (5.0)	32.5 (3.5)	31.9 (5.1)
Age of arrival in the U.S.	29.1 (4.0)	27.0 (5.6)	28.2 (3.2)	25.1 (2.9)	27.4 (4.2)
Length of residence in the U.S.	1.7 (1.2)	2.5 (0.7)	6.6 (2.5)	7.3 (2.6)	4.5 (3.1)
University study in the U.S.	0.2 (0.2)	2.1 (0.7)	0.9 (1.0)	5.0 (1.2)	2.0 (2.0)
Full-time work in the U.S.	1.5 (1.3)	0.2 (0.4)	5.2 (3.1)	1.8 (2.1)	2.2 (2.7)
English study in China	8.9 (2.7)	10.8 (3.5)	8.2 (2.9)	9.1 (2.8)	9.2 (3.1)
% self-reported use of English	45% (16)	47% (25)	47% (17)	54% (16)	48% (19)

Note. All values except gender and self-reported use of English are expressed in years; standard deviations are in parentheses.

the United States between the ages of 16 and 40 years. One Chinese participant was excluded from the study because he failed to meet the chronological age requirement; another was excluded because he had returned to China for 3.5 years after first arriving in the United States.

As summarized in Table 1, the 60 Chinese participants included in the study had a mean age of 32 years, had arrived in the United States at a mean age of 27 years, and had lived here for an average of 5 years. These participants reported having studied English at school in China for 1–18 years (M = 9 years) before coming to the United States and reported using English between 10% and 90% of the time (M = 48%). The Chinese participants spoke one of three native languages: Mandarin (28 participants), Shanghainese (21), or Cantonese (11). Additionally, five native English (NE) speakers participated as controls.

The following procedure was used to assign the Chinese participants to one of four groups of 15 each. To begin, the Chinese participants were assigned to one of two nonoverlapping groups based on LOR in the United States (short LOR = 0.5–3.8 years, long LOR = 3.9–15.5 years). The participants in the short- and long-LOR groups were then subdivided according to their primary occupation in the United States. Of the 60 participants, 18 had been enrolled for more than 3 academic years in an American university, whereas 18 others had never been enrolled. The participants in the two LOR-defined groups were assigned to nonoverlapping subgroups based on years of U.S. education.

The participants assigned to the nonstudent groups were likely to have received less English-language input than those assigned to the student groups,

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

especially if they worked in a laboratory with other speakers of a Chinese language. Many participants in the nonstudent groups arrived from China with an M.D. or a Ph.D. degree or with the training needed for full-time employment in a research laboratory. Others studied in the United States briefly before taking on full-time employment, usually in a laboratory. A few others were not employed outside the home but were married to a full-time student or university employee.

A two-way ANOVA examining the LOR values for the four groups of Chinese participants revealed that the nonstudents and students did not differ significantly in terms of LOR, $F(1, 56) = 2.6$, $p > .10$, and that LOR did not interact significantly with occupational status, $F(1, 56) = 0.0$, $p > .10$. The lack of a significant interaction indicated that the LOR differences between the two groups of nonstudents (1.7 vs. 6.6 years) and the two groups of students (2.5 vs. 7.3 years) were comparable.

Other ANOVAs revealed that the four groups of Chinese participants were balanced in terms of two other variables shown in Table 1. An analysis of the Chinese participants' self-reports of percentage use of English yielded nonsignificant main effects of LOR, $F(1, 56) = 0.7$, $p > .10$, and occupational status, $F(1, 56) = 0.8$, $p > .10$, and a nonsignificant two-way interaction, $F(1, 56) = 0.3$, $p > .10$. The analysis of years of English study in China yielded nonsignificant main effects of LOR, $F(1, 56) = 2.4$, $p > .10$, and occupational status, $F(1, 56) = 3.4$, $p > .05$, and a nonsignificant two-way interaction, $F(1, 56) = 0.4$, $p > .10$.

As summarized in Table 1, the nonstudents with relatively short LORs had received an average of 0.2 years of education in the United States (range = 0–0.5 years), whereas the students with relatively short LORs had received an average of 2.1 years of education in the United States (range = 0.8–3.5 years). The nonstudents with relatively long LORs had received an average of 0.9 years of U.S. education (range = 0.2–2.3) whereas the students with relatively long LORs had received an average of 5.0 years of U.S. education (range = 2.5–8.0). A two-way ANOVA examining years of U.S. education yielded an interaction between LOR (short vs. long) and occupational status (student vs. nonstudent), $F(1, 56) = 22.6$, $p < .01$. The interaction arose because the difference in years of U.S. education between the two groups with relatively long LORs (0.9 vs. 5.0 years) was larger than the difference between the two groups with relatively short LORs (0.2 vs. 2.1 years).

Two other ANOVAs revealed confounds. The difference in AOA between participants having relatively short LORs (AOA = 28.0 years) and long LORs (AOA = 26.7 years) was nonsignificant, $F(1, 56) = 1.6$, $p > .10$. LOR did not interact significantly with occupational status in the analysis of AOA values, $F(1, 56) = 0.2$, $p > .10$. However, the difference in AOA between the nonstudents and students (28.7 vs. 26.0 years) was significant, $F(1, 56) = 6.2$, $p < .05$. When chronological age was examined, the difference between the nonstudents (mean age = 32.8 years) and students (mean age = 31.0 years) was found to be nonsignificant, $F(1, 56) = 2.2$, $p > .10$. LOR and occupational status did not interact significantly, $F(1, 56) = 0.2$, $p > .10$. However, the participants with short

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

LORs were significantly younger (mean age = 30.1 years) than those with long LORs (mean age = 33.7 years), $F(1, 56) = 8.4$, $p < .01$.

Procedures

All of the participants were required to pass a pure-tone hearing screening at octave frequencies between 500 and 8000 Hz at 20 dB HL before taking part in the study. The participants were tested individually in a sound booth in a single 75-minute session. They responded to a language background questionnaire, then participated in the three experiments presented below in the same fixed order. The stimuli in all three experiments were presented via headphones at a self-selected comfortable level using a personal computer. The participants responded using a mouse.

EXPERIMENT 1

This experiment examined the identification of stops (/b d g p t k/) that occurred in the final position of naturally produced English words. Two techniques were used to ensure that the identification scores obtained from the Chinese participants would be below ceiling. The first technique involved editing. English word-final stops may or may not be produced with an audible release burst. Previous research has shown that native English speakers correctly identify word-final English stops that do not have release bursts at about the same rate as stops having release bursts. Chinese adults, on the other hand, identify final stops lacking a burst less often than final stops that have release bursts because Chinese languages permit few or no obstruents to occur in word-final position (Flege, 1989; Flege & Wang, 1990). We therefore examined English words ending in released final stops, and also copies of the words from which the final release bursts had been edited out. Second, the words ending with released stops, and those from which the release bursts had been edited out, were presented at two levels of masking noise as well as without noise.

Method

The following procedures were used to create 72 stimuli. Two adult male native speakers of English produced /amaC/ disyllables with stress on the second syllable. All six English stop consonants (/b d g p t k/) occurred in final position and were produced with a release burst. The 12 disyllables (2 talkers × 6) were digitized at 22.05 kHz, then normalized for peak intensity (50% of the full scale). The word-final release bursts were removed from copies of each stimulus. Copies were then made of the 24 stimuli. One set of copies was embedded in the center of a 1500-ms pink-noise segment that had been ramped on and off over the first and last 30 ms. This yielded stimuli having a signal to noise

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

(S/N) ratio of 16 dB. Another set of copies was embedded in a more intense copy of the pink-noise segment, yielding stimuli with an S/N ratio of 10 dB.[2]

The participants were tested one at a time in a sound booth. A practice session using appropriate nontest stimuli preceded the experiment. During the experiment, the 72 stimuli were randomly presented two times each without feedback in each of three blocks. These blocks occurred in a fixed order: first the stimuli without noise, then the stimuli with an S/N level of 16 dB, and finally the stimuli with an S/N level of 10 dB. The participants were instructed to identify the word-final stop by clicking one of six buttons marked "p," "t," "k," "b," "d," and "g" that were shown on the computer monitor. They were told to make their best guess if uncertain.

The percentage of times that each participant correctly identified the unedited and edited tokens of /b d g p t k/ in each of the three blocks was calculated. The six resulting scores (two editing conditions × 3 blocks) were each based on 24 forced-choice judgments (6 stops × 2 talkers × 2 repetitions). A preliminary (2) LOR × (2) Occupational Status × (2) Editing × (3) Noise Level ANOVA revealed that neither LOR nor occupational status interacted with noise level. The decision was therefore made to average over the three S/N levels (no noise, 16 dB, 10 dB). This yielded average percent correct scores for edited and unedited stops that were each based on 72 judgments.

Results and Discussion

The mean percentage of times that stops were identified correctly by the native English speakers and the four groups of Chinese participants is shown in Figure 1. As expected based on previous research (Flege, 1989; Flege & Wang, 1990), the native English participants obtained higher consonant identification scores ($M = 97\%$, range = 94–99%) than the Chinese participants did ($M = 78\%$, range = 43–94%). Also as expected, the Chinese participants obtained higher scores for the stimuli containing a release burst than for the edited stimuli without release bursts (unedited = 90%, edited = 67%). Averaged over occupational status (student vs. nonstudent), higher scores were obtained for the participants with relatively long LORs than for those with relatively short LORs (81% vs. 76%). Averaged over the two levels of LOR (short vs. long), higher scores were obtained for the nonstudents than for the students (80% vs. 77%).

The consonant identification scores were submitted to a (2) LOR × (2) Occupational Status × (2) Editing Condition ANOVA, which yielded a significant main effect of editing, $F(1, 56) = 752.0$, $p < 0.01$. The main effects of LOR, $F(1, 56) = 2.8$, $p > .10$, and Occupational Status, $F(1, 56) = 0.7$, $p > .10$, were nonsignificant, as were the interactions between Occupational Status and Editing, $F(1, 56) = 0.0$, $p > .10$, and the three-way interaction, $F(1, 56) = 1.6$, $p > .10$. However, the interactions between LOR and Occupational Status, $F(1, 56) = 9.3$, $p < .01$, and between LOR and Editing, $F(1, 56) = 6.0$, $p < .05$, were significant.[3]

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

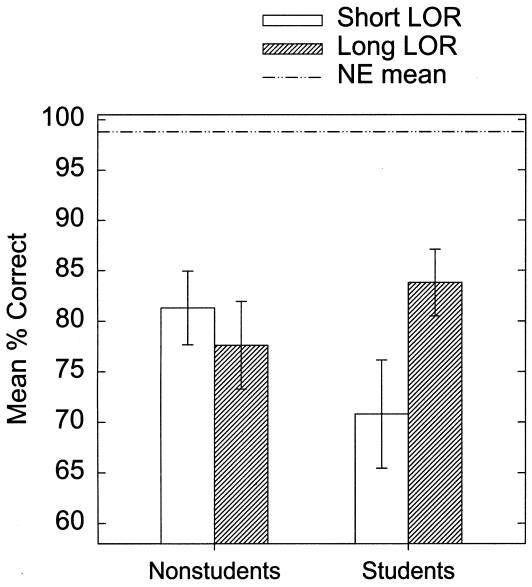


Figure 1. The mean percentage of correct identifications of word-final English stops in experiment 1 by groups of Chinese participants who differed according to their primary occupation in the United States (students vs. nonstudent workers) and length of residence (LOR) in the United States, and by five native English (NE) speakers, indicated by the dashed line. The brackets enclose ±1.0 *SE*.

The interaction between LOR and Editing was explored through simple effects tests. The difference in scores for the unedited and the edited stops was significant for the participants having short LORs (89% vs. 81%), $F(1, 29) = 487.3$, $p < .01$, as was the difference for the participants having long LORs (92% vs. 70%), $F(1, 29) = 290.9$, $p < .01$. The interaction seems to have arisen because the effect of LOR was greater for edited than unedited stops. The effect of LOR for unedited stops (long LOR = 89%, short LOR = 92%) was nonsignificant, $F(1, 58) = 1.0$, $p > .10$, whereas the difference for edited stops (63% vs. 70%) was marginally significant, $F(1, 58) = 3.6$, $p = .06$.

The interaction between LOR and occupational status was also explored through tests of simple effects. The effect of LOR for the students (short LOR = 71%, long LOR = 84%) was significant, $F(1, 28) = 8.9$, $p < .01$, whereas the effect of LOR for the nonstudents (short LOR = 81%, long LOR = 78%) was nonsignificant, $F(1, 28) = 1.3$, $p > .10$. The students with long LORs showed a nonsignificant tendency to identify stops at higher rates than the nonstudents

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

with long LORs did, $F(1, 28) = 3.1$, $p > .05$. However, the students with short LORs identified stops at significantly lower rates than the nonstudents with short LORs did, $F(1, 28) = 6.3$, $p < .01$.

These findings suggest two conclusions. The first is that simply living in the United States for an additional 5 years is not sufficient to enable adults from China to perceive word-final English stop consonants more accurately. The second is that the students may have received more English speech input than the nonstudents did. One might also infer either that some time was needed for the advantage associated with a student status to manifest itself or that the nonstudents received more English input during their initial period in the United States than the students did.

EXPERIMENT 2

This experiment examined the Chinese participants' knowledge of surface characteristics of English morphosyntax. The grammaticality judgment test used here was the one used in a recent study of Korean–English bilinguals by Flege, Yeni-Komshian, and Liu (1999). The test items were presented both aurally and orthographically at the same time. This meant that errors were unlikely to occur as the result of the misperception of particular vowels or consonants (e.g., the failure to hear the morpheme marking past tense in a word like *died*).

Method

The grammaticality judgment test used here comprised nine sets of sentences, each intended to evaluate different morphosyntactic structures. As summarized in Table 2, the test assessed the participants' knowledge of past tense, plurals, third-person singular, determiners, pronouns, particle movement, subcategorization, lexically specified subject-object raising, and question formation. Half of the 144 sentences were grammatical and half were ungrammatical sentences created by eliminating a required morpheme or word from a grammatical sentence, by changing a word, or by moving some word(s) to an ungrammatical position. For example, the sentence *The girl cooked dinner for her family last night* was rendered ungrammatical by changing the tense of the verb from past to present (**The girl cooks dinner for her family last night*).

The 144 sentences were spoken by an adult male native speaker of English, then digitized at 22.05 kHz. The sentences were presented in a pseudo random order to the participants in two modalities. A written version of each sentence was presented on the screen of the PC used for testing as the sentence was presented via headphones. The participants were told to judge the grammaticality of each sentence by clicking a button marked "Yes—grammatical" or "No—ungrammatical." The dependent variable calculated for each participant was the overall percentage of correct responses to the 144 sentences.

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

Table 2. Number and examples of the nine grammaticality judgment test sentence types

n	Sentence type	Examples
8	Past tense	A policeman gave Alan a ticket for speeding yesterday. *A policeman gived Alan a ticket for speeding yesterday.
8	Plural	Todd has many coats in his closet. *Todd has many coat in his closet.
8	Third-person singular	Every Friday our neighbor washes her car. *Every Friday our neighbor wash her car.
8	Determiners	The boy is helping the man build a house. *The boy is helping the man build house.
8	Pronouns	Susan is making some cookies for us. *Susan is making some cookies for we.
6	Particle movement	Kevin called up Nancy for a date. *Kevin called Nancy for a date up.
14	Subcategorization	The little boys laughed at the clown. *The little boys laughed the clown.
4	Lexically specified subject/object raising	Larry believed himself to be brave. *Larry believed that himself to be brave.
4	Y/N questions	Should Timothy have gone to the party? *Should have Timothy have gone to the party?
4	*Wh*-questions	Where did she put the book? *Why did she put the book?

Note. n is the number of sentence pairs; the ungrammatical member of each sentence pair is marked by an asterisk.

Results and Discussion

As expected, the native English participants obtained higher grammaticality judgment test scores ($M = 98\%$, range = 96–99%) than the 60 native Chinese participants did ($M = 80\%$, range = 57–92%). Averaged across the two LOR groups, the scores were higher for the students than for the nonstudents (83% vs. 77%). Averaged across occupational status, the scores were higher for the long-LOR participants than for the short-LOR participants (82% vs. 80%). However, as shown in Figure 2, the students with long LORs obtained higher scores than the students with short LORs did (87% vs. 79%). Conversely, the nonstudents with long LORs obtained somewhat *lower* scores than the nonstudents with short LORs did (76% vs. 79%).

The grammaticality judgment test scores were submitted to a (2) LOR × (2) Occupational Status ANOVA. This analysis yielded a nonsignificant main effect of LOR, $F(1, 56) = 1.7$, $p > .10$, a significant main effect of Occupational Status, $F(1, 56) = 7.7$, $p < .01$, and a significant two-way interaction, $F(1, 56) = 7.4$, $p < .01$. The two-way interaction was explored through tests of simple main effects.

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

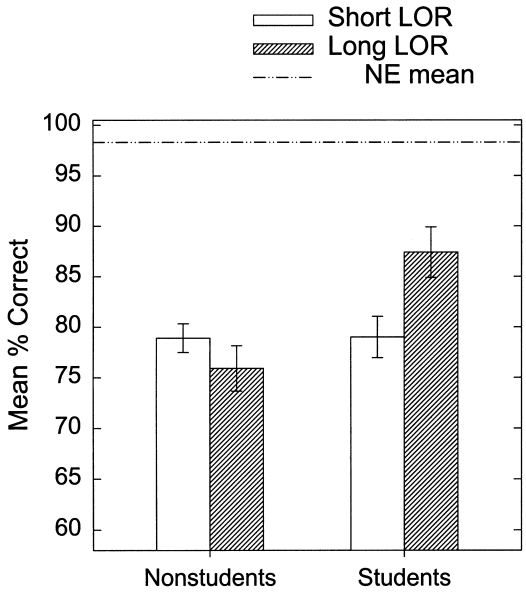


Figure 2. The mean percentage of correct responses in experiment 2 to 144 English sentences (half grammatical, half ungrammatical) by groups of Chinese participants who differed according to their primary occupation in the United States (students vs. nonstudent workers) and length of residence (LOR) in the United States, and by five native English (NE) speakers, whose mean score is indicated by the dashed line. The brackets enclose ±1.0 *SE*.

The LOR effect was found to be significant for the students, $F(1, 28) = 6.8$, $p < .05$, but not for the nonstudents, $F(1, 28) = 1.3$, $p > .10$. The students with long LORs obtained significantly higher scores than the nonstudents with long LORs, $F(1, 28) = 11.6$, $p < .01$, whereas the students and nonstudents with relatively short LORs obtained the same average score, $F(1, 28) = 0.0$, *ns*.

These findings suggest that simply living in the United States for an additional 5 years did not in itself increase the Chinese participants' knowledge of English morphosyntax. An effect of LOR on the grammaticality judgment test scores was obtained, but only for the Chinese participants who had been enrolled as students during most of their stay in the United States. One might conclude that some aspect of student life contributed to an increase in the students' knowledge of English morphosyntax, but only after several years of study in the United States.

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

EXPERIMENT 3

This experiment assessed the Chinese participants' comprehension of English using the University of Michigan English Language Institute Listening Comprehension Test (LCT). The LCT has been described as an "aural grammar test" because all of its aurally presented stimuli and written responses are "grammar based" (English Language Institute Listening Comprehension Test, 1986).

Method

Each of the 45 LCT items consisted of a short question or statement followed by three short written answers, of which one was correct. The participants were required to select one of three answers within a specified time. For example, the first item on the LCT was the question: *Was that a good movie you saw?* Three possible written answers were provided for this question: "Yes, it is," "Yes, it was," and "Yes, I have." Choosing the correct answer (*Yes, it was*) required knowledge of verb tense.

The LCT was modified slightly so that the participants' responses could be collected using a PC. The 45 LCT items on an audio tape were digitized at 22.05 kHz, then normalized for peak intensity. In a standard administration of the LCT, the stimuli are presented aurally one at a time, and the three written responses for each item are presented in a booklet. However, rather than checking the correct answer in a test booklet, the participants in this experiment clicked one of three buttons shown on a computer screen (alongside the number of the item being tested). They were told to respond as accurately and rapidly as possible. The next item was presented 1 second after a response was given, or 10 seconds later if no response was given (rather than after 12 seconds, as in the standard administration). The participants were not permitted to hear a test item more than a single time nor to change an answer once given. An item was scored as incorrect if a response was not given within 10 seconds.

Results and Discussion

As expected, the five native English participants obtained a higher percentage of correct responses on the LCT ($M = 96\%$, range = 93–98%) than did the 60 native Chinese participants ($M = 80\%$, range = 56–96%). Averaged across the participants differing in occupational status, the scores were higher for the long-LOR participants than for the short-LOR participants (83% vs. 78%). Averaged across the LOR groups, higher scores were obtained for the students than for the nonstudents (84% vs. 77%). However, as shown in Figure 3, the students with long LORs obtained higher scores than the students with short LORs did (89% vs. 79%). Conversely, the nonstudents with relatively long and short LORs obtained similar average scores (77% vs. 76%).

An ANOVA examining the LCT scores yielded a significant main effect of

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

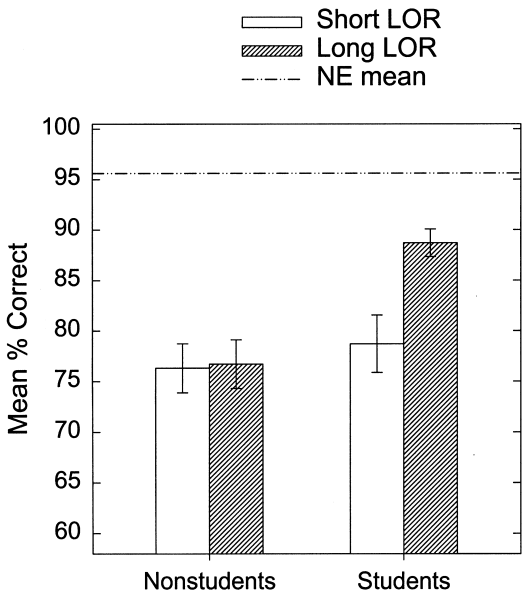


Figure 3. The mean percentage of correct responses to a 45-item listening comprehension test in experiment 3 by groups of Chinese participants who differed according to their primary occupation in the United States (students vs. nonstudent workers) and length of residence (LOR) in the United States, and by five native English (NE) speakers, whose mean value is indicated by the dashed line. The brackets enclose ±1.0 *SE*.

LOR, $F(1, 56) = 5.2$, $p < .05$, a significant main effect of Occupational Status, $F(1, 56) = 9.6$, $p < .01$, and a significant two-way interaction, $F(1, 56) = 4.3$, $p < .05$. Simple effects tests revealed that the effect of LOR was significant for the students, $F(1, 28) = 10.2$, $p < .01$, but not for the nonstudents, $F(1, 28) = 0.0$, $p > .10$. The students with long LORs obtained significantly higher scores than the nonstudents with long LORs did, $F(1, 28) = 19.0$, $p < .01$. However, the difference between students and nonstudents with short LORs was nonsignificant, $F(1, 28) = 0.4$, $p > .10$. These findings suggested that simply living in the United States for an additional 5 years did not, in itself, result in an improved comprehension of English. The fact that an LOR effect was obtained only for the students suggested that something associated with student life led to an improved comprehension of English. Perhaps the students received more input from native speakers of English than the nonstudents did.

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

ADDITIONAL ANALYSES

The 60 Chinese participants obtained lower average scores in all three experiments than the five NE speakers. In experiment 1, they correctly identified word-final English stop consonants less often (range = 43–94%, *M* = 78%) than the NE speakers (range = 94–99%, *M* = 97%). In experiment 2, they obtained lower scores on a 144-item grammaticality judgment test (range = 57–99%, *M* = 81%) than the NE speakers (range = 96–99%, *M* = 98%). Finally, in experiment 3, they obtained lower scores on a 45-item listening comprehension test (range = 56–96%, *M* = 80%) than the NE speakers (range = 93–98%, *M* = 96%).

Both student status and a relatively long LOR seem to have contributed to a relatively high level of performance in English. However, most of the students with a relatively long LOR obtained lower scores than the NE speakers. Just 2 of the 15 long-LOR students obtained a consonant identification score in experiment 1 that fell within 2.0 standard deviations (*SD*s) of the mean value obtained for the NE speakers. Two long-LOR students obtained a grammaticality judgment test score in experiment 2 that fell within 2.0 *SD*s of the NE mean, and six obtained a comprehension test score in experiment 3 that fell within 2.0 *SD*s of the NE mean. (Just 1 of the remaining 45 Chinese subjects met the 2.0-*SD* criterion for experiment 1, none for experiment 2, and 3 for experiment 3.)

Much the same effect of LOR was obtained in experiments 1–3. Students having long LORs obtained higher scores than the students having short LORs, whereas the effect of LOR was nonsignificant for the nonstudents. An interpretation of these findings is straightforward because the LOR difference between the two groups of students (2.5 vs. 7.3 years) was comparable to the LOR difference between the two groups of nonstudents (1.7 vs. 6.6 years). The LOR effect observed for the students can be attributed to the kind of English-language input they received, to their continuing education in English, or to some combination of both.

A more complex pattern of results was obtained for occupational status. For both the grammaticality judgment test scores (experiment 2) and the comprehension (LCT) scores (experiment 3), long-LOR students obtained higher scores than the long-LOR nonstudents. The consonant identification scores (experiment 1) obtained by the long-LOR students and nonstudents did not differ significantly, however, and the short-LOR nonstudents obtained higher scores than the short-LOR students. Interpreting these results is difficult because a 2.9-year difference in years of U.S. education existed between the two groups of students (short-LOR = 2.1 years of study, long-LOR = 5.0 years) as compared to just a 0.7-year difference between the two groups of nonstudents (short-LOR = 0.2 years of study, long-LOR = 0.9 years of study). Also, LOR was confounded with the Chinese participants' age of arrival (AOA) in the United States.

A series of correlation analyses was therefore undertaken to provide a better understanding of the effects of LOR and years of U.S. education. Simple

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

Table 3. Simple and partial correlations between two variables and the percent correct consonant identification scores obtained for the Chinese participants in experiment 1

	Nonstudents		Students	
Variables	Simple	Partial	Simple	Partial
Years of education in the U.S.[a]	.28	.24	.41*	.19
Length of residence in the U.S. (years)[b]	–.34	.26	.42*	.42*

Note. An asterisk indicates significance at the .05 level.
[a]In the partial correlation analyses examining years of U.S. education, age of arrival, chronological age, and LOR served as control variables.
[b]In the partial correlation analyses examining length of residence, age of arrival, chronological age, and years of education served as control variables.

Table 4. Simple and partial correlations between two variables and the grammaticality judgment test scores obtained for the Chinese participants in experiment 2

	Nonstudents		Students	
Variables	Simple	Partial	Simple	Partial
Years of education in the U.S.[a]	.43*	.27	.37*	.22
Length of residence in the U.S. (years)[b]	–.40*	–.03	.30	.10

Note. An asterisk indicates significance at the .05 level.
[a]In the partial correlation analyses examining years of U.S. education, age of arrival, chronological age, and LOR served as control variables.
[b]In the partial correlation analyses examining length of residence, age of arrival, chronological age, and years of education served as control variables.

correlations were computed between the outcome measures of experiments 1–3 and both LOR and years of U.S. education. In partial correlation analyses examining the relation between the outcome measures and the two variables of interest (LOR, years of U.S. education), the participants' chronological age, AOA in the United States, and either U.S. education or LOR served as control variables. Separate analyses were carried out for the 30 students and the 30 nonstudents.

The results of the simple and partial correlation analyses examining LOR are presented in Tables 3–5 for experiments 1–3, respectively. Negative correlations were obtained between the nonstudents' LORs and the outcome measures of all three experiments. The negative correlation involving the grammaticality judgment test scores was significant. This means that the longer the nonstudents had resided in the United States, the lower their grammaticality judgment test scores tended to be. Perhaps the longer the nonstudents lived in the United States, the more they forgot what they had learned about English grammar in school in China (see also Oyama, 1982a).

The simple correlations between the nonstudents' LORs and the outcome

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

Table 5. Simple and partial correlations between two variables and scores obtained on the listening comprehension test for the Chinese participants in experiment 3

	Nonstudents		Students	
Variables	Simple	Partial	Simple	Partial
Years of education in the U.S.[a]	.38*	.10	.36*	.22
Length of residence in the U.S. (years)[b]	–.04	–.07	.30	.14

Note. An asterisk indicates significance at the .05 level.
[a]In the partial correlation analyses examining years of U.S. education, age of arrival, chronological age, and LOR served as control variables.
[b]In the partial correlation analyses examining length of residence, age of arrival, chronological age, and years of education served as control variables.

measures of all three experiments were positive. The positive correlation for the word-final consonant identification scores was significant. This means that the longer the students had resided in the United States, the more often they tended to correctly identify word-final English stop consonants.

In the partial correlation analyses, just one correlation reached significance: that between the students' LORs and the consonant identification scores. This suggests that, of the three aspects of L2 acquisition examined here, segmental phonetic perception may be influenced most importantly by native-speaker input.

The results of the simple and partial correlation analyses for years of U.S. education are also presented in Tables 3–5. In all six instances (2 groups $\times$ 3 outcome measures), the simple correlations were positive. Five of these six correlations were significant. However, in no instance did any correlation remain significant when the effects of other variables were statistically controlled.

GENERAL DISCUSSION

As reviewed in the introduction, the results obtained in previous research led Moyer (1999, p. 83) to conclude that LOR in an L2-speaking environment has an "inconsequential" effect on L2 learning. Were this conclusion to hold true across the board, it would mean that adults' ability to learn an L2 is severely limited, possibly by a critical period. However, common experience tells us that many adults eventually do succeed in learning an L2. The purpose of the present study, therefore, was to determine if Chinese adults living in the United States would resemble NE speakers more closely as their LOR in the United States increased. L2 performance was assessed in experiments examining the identification of word-final English stops (experiment 1), knowledge of basic morphosyntactic characteristics of English (experiment 2), and the ability to aurally comprehend English (experiment 3). The hypothesis tested in all three

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

experiments was that adult learners' L2 performance will improve over time, but only if they occupy the kind of input-rich L2 environment that is typical for children who immigrate to North America (Jia & Aaronson, 1999; see also Cummins, 1981, 1991).

Chinese participants were recruited to test the LOR hypothesis. The 30 students had been enrolled at an American university during most or all of their stay in the United States, whereas the 30 nonstudents had worked full-time during most of their stay. (In fact, most of the nonstudents worked in settings that often afford little opportunity for native-speaker input, biomedical research laboratories.) Half of the students and half of the nonstudents had relatively short LORs ($M = 2.1$ years). The remaining half of the participants in the student and nonstudent groups had relatively long LORs ($M = 6.9$ years). A significant LOR × Occupational Status (student vs. nonstudent) interaction was obtained in analyses of the scores from all three experiments. As predicted, the students with relatively long LORs obtained significantly higher scores than the students with relatively short LORs, whereas no significant differences were observed for the groups of nonstudents who differed in LOR.

Much the same LOR differences existed between the two groups of students (means = 2.5 vs. 7.3 years) and between the two groups of nonstudents (means = 1.7 vs. 6.6 years). It therefore seems reasonable to conclude that the LOR effect observed for the students was due to the kind of English-language input or formal education that they, but not the nonstudents, had received in the United States. It is important to note, however, that we did not directly assess either the quantity or the quality of English-language input that any of the Chinese participants had received in the United States. This limitation of the present study will need to be remedied in future research. Still, the results obtained for the nonstudents showed clearly that simply living for an additional 5 years in the United States did not, in itself, result in a more nativelike level of L2 performance. Moreover, the results obtained for the students indicated that Moyer's (1999) conclusion regarding LOR is not correct for all samples of L2 learners. In some instances, adults' L2 performance does improve as LOR increases.

Most participants in the Chinese group obtaining the highest scores—the students with relatively long LORs—obtained scores that were lower than the NE speakers' scores. An important question for future research is whether, and under what circumstances, adults can learn an L2 sufficiently well in order to perform in a nativelike manner. The long-LOR students examined in this study had lived in the United States for an average of 7.3 years and reported using English 54% of the time, on average. If their self-reported use of English was accurate, the long-LOR students had probably received less "full-time" English input (roughly 3.9 years) than children learning English as their first and only language would need in order to obtain scores in experiments 1–3 equaling those obtained for the adult NE speakers. There have been reports of adult learners showing a nativelike level of proficiency in their L2 (e.g., Birdsong,

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

1993; Bongaerts, Summeren, Planken, & Schils, 1997). It would be valuable to examine adult learners who inhabit the kind of input-rich L2 environment that is typical for child L2 learners (Jia & Aaronson, 1999) and who have also had more than ten years of "full-time" L2 input. Support would be provided for the critical period hypothesis (e.g., DeKeyser, 2000; Scovel, 1988) if such advantaged adult L2 learners could be shown to differ significantly from monolingual native speakers of the target L2.

We can only speculate at present as to what specific aspects of English-language experience led to a better performance by the students as their LORs in the United States increased. We suspect that the students received a large amount of English-language input from teachers and fellow students who were native speakers of English. They may also have received instruction from their teachers regarding grammatical aspects of English (e.g., marginalia on written assignments) or from fellow students. We infer that conversational English input contributed more to the LOR effects observed here than any formal instruction in English that the Chinese students may have received. This is because much the same LOR effects were obtained in the experiment examining the perception of word-final English stops, for which the students were unlikely to have received formal instruction, as in two other experiments testing aspects of English that may have benefited from formal instruction (i.e., the experiments examining aural comprehension and grammatical sensitivity).

All four groups of native Chinese participants reported using English about 50% of the time (see Table 1). Although not reported earlier, their estimates of percentage use of English were not correlated significantly with the scores obtained for any of the experiments reported here. This suggests indirectly that it was not a greater quantity of English-language input that enabled the long-LOR students to outperform the short-LOR students. Rather, it may have been the amount of English-language input they received from *native speakers* of English that enabled the long-LOR students to perform better in English than the short-LOR nonstudents.

One incidental, but nonetheless interesting, finding of the present study was that a significant negative correlation was found to exist between the nonstudents' LORs in the United States and their grammaticality judgment test scores. This negative correlation, which is reminiscent of a finding obtained by Oyama (1982a), suggested that the longer the nonstudents had lived in the United States, the more they forgot what they had learned about English grammar in China. However, the correlation did not remain significant when variations in confounded variables were partialled out. On the other hand, the positive correlation between the students' LORs and their consonant identification scores remained significant when the effects of other variables were statistically controlled. The fact that this was the only correlation to remain significant in partial correlation analyses suggested that, of the three aspects of L2 acquisition examined here, segmental phonetic perception may be influenced most by native-speaker input.

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

IMPLICATIONS

Previous research has shown that the later L2 learners are first extensively exposed to an L2, the stronger their foreign accents in the L2 tend to be (e.g., Oyama, 1982b; Patkowski, 1990) and the lower scores they tend to receive on a grammaticality judgment test (e.g., Johnson & Newport, 1989; Shim, 1994). Such age effects, which have been frequently replicated, are thought by some to arise from a critical period for L2 learning. Patkowski (p. 74) suggested that a "fundamental" difference exists between individuals who begin learning an L2 before versus after a critical period (see also DeKeyser, 2000). Johnson and Newport (p. 96) concluded that the capacity to learn an L2 "disappears or declines with maturation" up to about puberty, perhaps as the result of a maturational change in a "language acquisition device" or some "more general" cognitive mechanism. Scovel (1988) suggested that a critical period for speech learning may result from a loss of cortical plasticity near the onset of puberty. Mack (1998) suggested that a sensitive period for language acquisition may be associated with the pruning of overproduced synapses (see, e.g., Greenough, Black, & Wallace, 1987).

One problem for a critical period account of age-related effects on L2 learning is that some age-related effects that are thought to support the critical period hypothesis can be questioned on methodological grounds (see Flege, 1987, for discussion). For example, AOA typically shows a negative correlation with LOR and L2 use (the later the AOA, the shorter the LOR in an L2-speaking country and the less immigrants typically use the L2; see Flege, 1998). As mentioned in the introduction, Flege, Yeni-Komshian, and Liu (1999) used a subgroup matching technique to remove the effect of such confounds. These authors evaluated the critical period hypothesis by testing 240 native speakers of Korean who differed in AOA to the United States (range = 1–23 years). As the participants' AOA increased, their foreign accents in English tended to grow stronger, and the scores they obtained on a grammaticality judgment test tended to decrease. When factors confounded with AOA were controlled, the effects of AOA on foreign accent remained significant; however, the strong effect of AOA on the grammaticality judgment test scores disappeared.

The findings of Flege, Yeni-Komshian, and Liu (1999) leave open the possibility that the learning of L2 speech (if not morphosyntax) is constrained by a critical period. However, four recent findings for L2 speech learning are inconsistent with a critical period account. First, many individuals who began learning their L2 as young children speak the L2 with a detectable foreign accent (e.g., Flege, Frieda, & Nozawa, 1997; Piske & MacKay, 1999). Either the foreign accents observed for such individuals did not arise from the passing of a critical period, or else the critical period for L2 speech learning ends in childhood prior to the complete acquisition of the L1 phonetic system. Second, some highly motivated individuals who began learning the L2 after the end of the putative critical period have been shown to speak the L2 without a detectable

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

foreign accent (Bongaerts et al., 1997). Third, language use has been shown to affect strength of foreign accent independently of AOA (e.g., Flege, Munro, & MacKay, 1995; Piske & MacKay). Finally, an inverse relation has been found to exist between the strength of foreign accent in the L2 and the strength of L2-inspired foreign accent in the L1 (Yeni-Komshian, Flege, & Liu, 2000). Foreign accent in an L2 might be attributed to a critical period but not the development of foreign accent in the L1.

In summary, evaluating the critical period hypothesis for L2 learning requires careful attention to the sampling of participants, as well as to factors that are confounded with the participants' age of first exposure to the L2 (which is often indexed by AOA). The results of the present study suggested that the amount of progress in English achieved by Chinese adults living in the United States depended on how much native-speaker input they received. Had we simply tested Chinese adults with relatively short and long LORs, without considering the participants' daily activities (i.e., their status as students or lab workers), we might have erroneously concluded that an additional 5 years of residence in the United States did not have a measurable effect on the participants' learning of English. Such a finding, had it been obtained, could have been used as additional support for the critical period hypothesis (see above). However, LOR effects were obtained for Chinese adults in each of three experiments when two homogenous groups of 15 participants each were compared. Evidence that age (or neurological development) constrains L2 learning should be considered persuasive only if it can be shown that the participants under examination had received the kind of rich input that is needed for successful L2 learning.

(*Received 16 March 2001*)

NOTES

1. Similar results were obtained for native speakers of Spanish. Another possible interpretation is that an effect of LOR was observed for the early but not the late bilinguals because the early bilinguals had lived in the United States for a longer time. An effect of LOR may emerge only gradually.

2. The S/N ratios were calculated according to the following formula: S/N ratio = 20 log10 (RMS value of Stimulus/RMS value of the noise).

3. Flege and Wang (1990) observed that Shanghainese and Cantonese speakers were more sensitive to the English /t/-/d/ contrast than Mandarin speakers. This was attributed to the fact that Shanghainese and Cantonese permit voiceless stops in word-final position but Mandarin does not permit any stops, either voiced or voiceless. We assessed the role of L1 background by examining the average percent correct scores for edited and unedited tokens of /p t k/ and /b d g/ in a three-way ANOVA. This analysis yielded a significant L1 (Mandarin vs. Cantonese, Shanghainese) × Voicing (/p t k/ vs. /b d g/) interaction, $F(1, 58) = 8.0$, $p < .01$. Simple effects tests revealed that the differences between the two subgroups for /p t k/ (Mandarin = 74%, Cantonese and Shanghainese = 71%) was nonsignificant, $F(1, 58) = 1.3$, *ns*, whereas the difference between the two subgroups for /b d g/ (Mandarin = 91%, Cantonese and Shanghainese = 79%) was significant, $F(1, 58) = 8.4$, $p < .001$. An explanation for this finding falls outside the scope of this study.

REFERENCES

Bates, E., & Goodman, J. (1998). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language and Cognitive Processes*, *12*, 507–586.

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

Bialystok, E., & Miller, B. (1999). The problem of age in second-language acquisition: Influences from language, structure, and task. *Bilingualism: Language and Cognition*, *2*, 127–145.
Birdsong, D. (1993). Ultimate attainment in second language acquisition. *Language*, *68*, 706–755.
Bongaerts, T., Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, *19*, 447–466.
Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics*, *2*, 131–149.
Cummins, J. (1991). Interdependence of first- and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 113–140). New York: Cambridge University Press.
DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–534.
English Language Institute Listening Comprehension Test (1986). Division of Testing and Certification, English Language Institute, University of Michigan, Ann Arbor.
Flege, J. E. (1987). A critical period for learning to pronounce foreign languages? *Applied Linguistics*, *8*, 162–177.
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. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America*, *86*, 1684–1697.
Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, *93*, 1589–1607.
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. 229–273). Timonium, MD: York Press.
Flege, J. E. (1998). The role of subject and phonetic variables in L2 speech acquisition. In M. Gruber, D. Higgins, K. Olsen, & T. Wysocki (Eds.), *Papers from the 34th annual meeting of the Chicago Linguistic Society: Vol. 2. The panels* (pp. 213–232). Chicago: Chicago Linguistic Society.
Flege, J. E., Bohn, O.-S., & Jang, S. (1997). The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.
Flege, J. E., Frieda, A. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*, 169–186.
Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*, 3125–3134.
Flege, J. E., Munro, M. J., & Skelton, L. (1992). The production of word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *Journal of the Acoustical Society of America*, *92*, 128–143.
Flege, J. E., & Wang, C. (1990). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics*, *17*, 299–315.
Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*, 78–104.
Gardner, R. C. (1991). Attitudes and motivation in second language learning. In A. Reynolds (Ed.), *Bilingualism, multiculturalism, and second language learning* (pp. 43–63). Mahwah, NJ: Erlbaum.
Greenough, W. T., Black, J. E., & Wallace, S. (1987). Experience and brain development. *Child Development*, *58*, 539–559.
Grenier, G. (1984). Shifts to English as usual language by Americans of Spanish mother tongue. *Social Science Quarterly*, *62*, 537–550.
Jia, G., & Aaronson, D. (1999). Age differences in second language acquisition: The dominant language switch and maintenance hypothesis. In A. Greenhill, H. Littlefield, & C. Tano (Eds.), *Proceedings of the 23rd annual Boston University Conference on Language Development* (pp. 301–312). Somerville, MA: Cascadilla Press.
Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99.
Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*, 251–285.
Mack, M. (1998). Consideration of the past, present, and future roles of the neurosciences in the linguistic sciences. *Studies in the Linguistic Sciences*, *28*, 41–92.
Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition*, *21*, 81–108.

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press

Oyama, S. (1982a). The sensitive period and comprehension of speech. In S. Krashen, R. Scarcella, & M. Long (Eds.), *Child–adult differences in second language acquisition* (pp. 39–51). Rowley, MA: Newbury House.

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

Patkowski, M. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, *30*, 449–472.

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

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191–215.

Piske, T., & MacKay, I. R. A. (1999). Age and L1 use effects on degree of foreign accent in English. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetics Sciences* (pp. 1433–1436). Berkeley: Dept. of Linguistics, University of California Berkeley.

Purcell, E., & Suter, R. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, *30*, 271–287.

Riney, T. J., & Flege, J. E. (1998). Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition*, *20*, 213–244.

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

Shim, R. J. (1994). *The sensitive period for second-language acquisition: An experimental study of age effects on Universal Grammar and language transfer*. Unpublished doctoral dissertation, University of Illinois, Urbana–Champaign.

Snow, C., & Hoefnagel-Höhle, M. (1982a). Age differences in the pronunciation of foreign sounds. In S. Krashen, R. Scarcella, & M. Long (Eds.), *Child–adult differences in second language acquisition* (pp. 84–92). Rowley, MA: Newbury House.

Snow, C., & Hoefnagel-Höhle, M. (1982b). The critical period for language acquisition: Evidence from second language learning. In S. Krashen, R. Scarcella, & M. Long (Eds.), *Child–adult differences in second language acquisition* (pp. 93–112). Rowley, MA: Newbury House.

Stevens, G. (1999). Age at immigration and second language proficiency among foreign-born adults. *Language in Society*, *28*, 555–578.

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*, 131–150.

https://doi.org/10.1017/S0272263101004041 Published online by Cambridge University Press