A “Critical Period” in Second Language Acquisition Reconsidered: Production Patterns and Listening/Reading Proficiencies of Japanese Learners

Michiko Mochizuki Sudo* and Ikuyo Kaneko*

SUMMARY: The present study investigated the effect of living in the target language country and that of the age factor on the acquisition of English by Japanese learners. Three subject groups participated in this study: Native speakers of American English, Japanese returnees who had lived in the U.S., and college students who had never lived overseas. For measuring listening and reading abilities, we employed official TOEIC® scores. We also examined two factors of durational control: ISI durations and stressed vowel shortening. The results showed significant differences in all parameters examined between the returnees and general college students. We also observed that the age factor had an effect on the different degrees of attainment of the production proficiency and on listening ability of the returnees, but not on their reading ability.

Key words: second language acquisition, age factor, critical period, compensatory vowel shortening, ISI

1. Introduction

A variety of factors have been reported in the literature as having influence on second language (L2) acquisition (Altman 1980, Larsen-Freeman and Long 1991, Skehan 1989). The individual learner variables include age, aptitude, motivation, attitude, personality, cognitive factors, hemisphere specialization, learning strategies, memory, awareness, will, proficiency in the first language (L1), interest, sex, birth order, and prior experiences. Among them, age is one of the factors that has received the most attention in the field of L2 acquisition (Ellis 1994). Age of initiation of language acquisition can be regarded as a crucial variable to distinguish between L1 and L2 acquisition. In language acquisition, there is a general belief that younger learners do better than older learners. This belief is based on the concept of a critical period. It refers to a biologically constrained period in which language can be acquired more easily than at any other time, and beyond which complete mastery of a foreign language is no longer possible. This concept of the critical period was originally introduced by Penfield and Roberts (1959), positing that children have a specific capacity for language learning due to cerebral flexibility that subsides at approximately age 9. Lenneberg (1967), the most well-known researcher for claiming the Critical Period Hypothesis (CPH), proposed a critical period starting from about age 2 and continuing to puberty. He also suggested that language learning may be more difficult after puberty because the brain lacks the ability for adaptation. The theory that the learner’s age is a significant factor in restricting language acquisition has also been described as the notion of a “Sensitive Period” or “Maturational Constraints” (Bornstein 1989, Knudsen 2004, Lamendella 1977, Long 1990, Oyama 1976, Snow 1987).

Triggered by Lenneberg’s writing (1967), the idea of the critical period was extended to the field of L2 acquisition. A number of studies have claimed that the age factor affects L2 acquisition (Asher and Garcia 1969, Bornstein 1989, DeKeyser 2000, Long 1990, Oyama 1976, Scovel 1981, Singleton 2005, Snow 1987, Tahta, Wood and Loewenthal 1981), supporting the CPH. Major (2001) claimed that both the learner’s age of arrival in the L2-speaking country and the age of learning are important variables governing whether, or to what degree, a learner can attain native-like

*School of Health and Sports Science, Juntendo University (順天堂大学スポーツ健康科学部)
pronunciation. Several studies have observed the differences in L2 perception and production between early learners and late learners. McCarthy, Evans and Mahon (2013) investigated the production of stops and vowels of Sylheti (L1) and English (L2) by adult speakers who had been residents in the UK for similar lengths of time but had arrived there at different ages. Analyzing Voice Onset Time (VOT) of stops and vowel formants and duration, they reported that the late arrivals (speakers who had arrived in the UK after age 16, mean age of arrival: 21) produced native-like Sylheti stops and vowels, but that their English categories were influenced by their L1. The early arrivals (speakers who had arrived in the UK before the age of 16, mean age of arrival: 6) and the second-generation speakers (those who were born in the UK) produced native-like Sylheti vowels but not Sylheti stops, while their English production was similar to the Standard Southern British English. Darcy and Kruger (2012) examined the early L2 learners and investigated the age at which L1 vowel categories influence the acquisition of L2 vowels in ten-year-old Turkish children acquiring German as their L2. They tested 28 children (14 native speakers of German and 14 Turkish–German bilinguals) by both perception and production experiments. In the perception experiment, they tested the children’s ability to discriminate four German vowel contrasts using an oddity task while they examined their production of the same vowels with an un-cued word naming task in the production experiment. The results showed that the Turkish–German bilingual children categorized perceptually difficult German contrasts differently from German monolingual children. In contrast to the perception results, the two groups of children were globally indistinguishable in production.

Several other studies, on the other hand, have shown that the production of consonants by late L2 learners displayed their L1 features (Flege, Munro and MacKay 1995, Sundara, Polka and Baum 2006). Likewise, the late L2 learners showed the L1 features in the L2 vowel production (Baker and Trofimovich 2005, Flege, MacKay and Meador 1999). Sudo and Kaneko (2012) reported the differences in the production of English rhythmic patterns between early L2 learners (whose age of arrival was before puberty) and late L2 learners (whose age of arrival was after puberty) by measuring durations of interstress intervals (ISI), suggesting a different durational control of stressed and unstressed syllables in the ISI by the two subject groups. It should be noted that this previous study does not assume equal ISIs in English. Quite a few studies have shown that perfect isochrony, on the level of production, does not exist (Beckman 1992, Ericksson 1991, Laver 1994, Lehiste 1977 for review). On the basis of acoustic measurements, some researchers have rejected the dichotomy between stress-timed and syllable- or moratomized languages (Dauer 1983). Others have reported a weak categorical distinction between stress-timed and syllable-timed languages, based on durational variability (Grabe and Low 2002). Evidence for the compensatory control in ISIs has also been presented in various studies from the viewpoints of both production and perception (Fowler 1977, Huggins 1975, Mochizuki-Sudo and Kiritani 1991, Rakerd, Sennett and Fowler 1987). The results of the above-mentioned studies indicate that ISI is a rhythmic unit in the production and perception of English. Lending support to this claim, Sudo and Kaneko (2012) observed that the early L2 learners showed a significant amount of approximation to the production patterns of native speakers of English in terms of the normalized ISI durations and the increments of increase in the normalized ISI durations, while no statistical differences in the ISI duration patterns were found between the late L2 learners and Japanese learners of English who had neither lived nor studied abroad. This result could be interpreted as indicating that the factor of age of arrival has an effect on different degrees of L2 attainment from the viewpoint of the production proficiency. Likewise, early L2 learners displayed production patterns that are similar to that of native speakers of both languages, L1 and L2 (Baker and Trofimovich 2005, Guion 2003). Regarding production, there are some studies which have shown that even late L2 learners are successful learners in that they have become indistinguishable from monolinguals (Bongaerts, Blanken and Schils 1995, Kinsella 2009, Neufeld 1977, 1979, 1980). In sum, significant early-late differences have been reported for L2 perception, and different degrees of early-late effects have been observed for L2 production studies.

In order to account for age effects on speech performance, various explanations have been offered in L2 research. The most common, and the biological explanation is the critical period concept, as we have mentioned above. Another type of explanation is that age effects arise from input differences between L1 and L2 (Oyama 1976, Snow and Hoefnagel-Hohle 1977, Sudo and Kaneko 2005). What this explanation indicates is that the accuracy of L2 pronunciation depends on how much native-speaker input as opposed to foreign-accented L2 input is received (Flege and Liu 2001). In other words, late L2 learners receive less adequate L2
phonetic input than early learners. The phonetic interaction between L1 and L2 can be offered as another account for age effects. The Speech Learning Model posited by Flege (Flege 1995, 1999, 2002, Flege, Schirru and MacKay 2003) suggests that the phonetic elements making up the L1 and L2 phonetic sub-systems of a bilingual exist in a common phonological space, rather than in two separate phonological systems, and the two phonetic sub-systems interact each other. Another explanation, the Perceptual Assimilation Model (Best, McRoberts and Goodell 2001) indicates that interaction between L1 and L2 perception increases with age.

Whatever the source of an age factor is, this specific factor in L2 acquisition seems to be undeniable. The differences in the age effect among various studies suggest that the degrees of attainment of L2 acquisition might vary, depending on the aspects of acquisition. In other words, the age factor has different degrees of effects on various features of L2 production and perception.

Regarding another aspect of production in English, a compensatory vowel shortening which relates to the production of English rhythmic patterns has been reported in various studies (Fant, Kruckenberg and Nord 1991, Fowler 1977, Huggins 1975, Katz 2012, Munhall, Fowler, Hawkins and Saltzman 1992, Rakerd, Sennett and Fowler 1987, Shaiman 2001). This phenomenon refers to a shortening of stressed vowels with the addition of following unstressed syllables in the ISI. Fowler (1977) reported that the average percentage of the stressed vowels from a one-syllable ISI to a two-syllable ISI was 12%, ranging from 6 to 18% for six sentence sets. Mochizuki-Sudo and Kiritani (1991) reported a foot-level shortening for the native speakers of English, observing that the average percentage of stressed vowel shortening from a one-syllable ISI to a two-syllable ISI was 19%, while Japanese non-proficient speakers of English displayed very little foot-level shortening; an average of 7% shortening. Sudo, Kaneko and Nishimura (2012, 2013), and Sudo and Kaneko (2013) analyzed the compensatory vowel shortening by Japanese learners of English and examined their English proficiency in terms of the duration of stressed vowels in ISIs. The results showed that significant differences were observed in the compensatory vowel shortening between Japanese learners of English who had studied English in Japan and Japanese returnees who had lived in the U.S.

In order to measure the proficiency of English as L2, the Test of English for International Communication (TOEIC®) is one of the most globally adopted standardized tests. It is reported to provide a valid assessment of English-language reading and listening proficiencies (In'nami and Koizumi 2012, Powers 2010). It is a multiple-choice assessment consisting of Listening and Reading Sections. Various factors related to TOEIC scores have been reported in literature. To cite a few, Sudo, Mochizuki, Itoh and Kirino (2011) reported significant relationships between the L2 reading proficiency and the L1 working memory in addition to the sex differences in L2 acquisition. They also reported that TOEIC scores and reading rate in English exhibited a positive correlation with the working memory-related parameters in the female group, but not in the male group. Mochizuki, Sudo, Itoh and Kirino (2013) investigated the interrelationships among physical ability, social functioning, and L2 proficiency and conducted correlation analyses for physical parameters, capacities of social cognition, and TOEIC scores together with English reading rate, as well as vocabulary size. Among the results, TOEIC scores showed a strong correlation with reading rate and some correlation with vocabulary. Also, vocabulary ability showed a negative correlation with one of the social parameters. Regarding the TOEIC scores, Sudo and Kaneko (2012) showed that Japanese general college students and returnees were significantly different in the total TOEIC scores and those in both Listening Section and Reading Section. Thus, TOEIC can be used as one of the indices to measure L2 proficiency.

The purpose of this study was to investigate the age factor in the acquisition of English by Japanese learners. Specifically, we attempted to find out the effects of age of arrival in the target language country together with length of residence on English proficiency of Japanese learners. We analyzed English proficiency from the viewpoints of production patterns and TOEIC scores. We focused on the two factors of production patterns: ISI durations and stressed vowel shortening. Based on the results of our previous studies, ISI durational control and stressed vowel shortening are shown to be the parameters of nativelikeness. Using TOEIC scores, we measured listening and reading abilities. We were interested in finding out the effects of an age factor on production patterns and listening/reading proficiencies. Since speaking and listening abilities were reported to have high correlations (Bozorgian 2012), we hypothesized that there are relationships between production patterns and listening proficiencies. We were also interested in the relationships between the English proficiency measured by TOEIC scores and the proficiency measured by production patterns. The
research questions addressed in the present paper are:
1) Does an age factor have an effect on various degrees of L2 attainment, depending on different features of production and perception?
2) Is there a “critical period” in the acquisition of English by Japanese learners?

In this study, we attempted to find out the effects of an age factor on different aspects of production and perception, testing the hypothesis of a “critical period” in L2 acquisition.

2. Methods

2.1 Subjects

We prepared three subject groups for the production experiment: 13 adult native speakers of American English (AMR: 8 males and 5 females), 24 Japanese returnees who had lived in the U.S. (RTN: 12 males and 12 females), and 13 Japanese learners of English who had neither lived nor studied abroad (JPN: 9 males and 4 females). All subjects from JPN started their English study when they entered a junior high school. The Japanese learners were all college students who belong to either a School of Health and Sports Science or a School of Medicine, and took the official TOEIC test for the measurement of their English proficiency.

The age of arrival in the U.S. for the subjects of RTN was 0 to 18. We divided RTN into two subgroups according to the age of arrival in the U.S. The age of arrival for RTN1 was 0 to 9; the age of arrival for RTN2 was 13 to 18 (RTN1: 17 subjects, RTN2: 7 subjects). We intentionally excluded the subjects from our analysis whose age of arrival was between 10 to 12, since the definition of puberty varies from study to study, thus making the age of 9 and 13 years cut-off points in this study. Thus, the subjects of RTN1 were supposed to be the subjects before the critical period, and those of RTN2, the subjects after the critical period. The subjects of RTN had been in the U.S. for over one year at least, and 12 years at the longest. We also divided the same RTN into two subgroups according to the length of residence in the U.S. The length of residence for RTN3 was six years or more (mean=9.3 years); the length of residence for RTN4 was one to three years (mean=1.5 years). Based on the data collected in this study and to make two subgroups of short versus long length of residence, we chose three years of length of residence as the cut-off point (RTN3: 11 subjects, RTN4: 13 subjects). The age of arrival and the length of residence for the subgroups of RTN are presented in Table 1.

Table 1 Age of arrival and length of residence for RTN.

| Age of arrival | Length of residence
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RTN1 (n=17)</td>
<td>RTN3 (n=11) RTN4 (n=13)</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

Note. RTN=a Japanese returnee who had lived in the U.S.; RTN1=a returnee whose age of arrival was 0 to 9; RTN2=a returnee whose age of arrival was 13 to 18; RTN3=a returnee whose length of residence was six to 12 years; RTN4=a returnee whose length of residence was one to three years.

2.2 Materials

Five English sentences were devised as the linguistic material for the production experiment (Table 2). The sentences differed in the number of nominally unstressed syllables that intervened between a target stressed syllable and the next stressed syllable. The number of unstressed syllables ranged from one to three. These sentences contained the ISIs within and between words (ISI1), except for the one-syllable ISI “Pete.” While the number of syllables in the ISI1 varied from sentence to sentence; the number of syllables in the following ISI (ISI2) was always two syllables. As shown in Table 2, the five sentences contained a target stressed vowel [i].

Table 2 Linguistic materials for the production experiment.

<table>
<thead>
<tr>
<th>Number of syllables in a target ISI1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. P / ete pl / ays the p / iano.</td>
</tr>
<tr>
<td>ISI 1 ISI 2</td>
</tr>
<tr>
<td>2. P / ete can pl / ay the p / iano.</td>
</tr>
<tr>
<td>ISI 1 ISI 2</td>
</tr>
<tr>
<td>3. P / eter pl / ays the p / iano.</td>
</tr>
<tr>
<td>ISI 1 ISI 2</td>
</tr>
<tr>
<td>4. P / eter can pl / ay the p / iano.</td>
</tr>
<tr>
<td>ISI 1 ISI 2</td>
</tr>
<tr>
<td>5. P / eterson pl / ays the p / iano.</td>
</tr>
<tr>
<td>ISI 1 ISI 2</td>
</tr>
</tbody>
</table>

2.3 Procedures

We measured the Japanese learners’ English profi-
Table 3 TOEIC average scores (JPN & RTN).

<table>
<thead>
<tr>
<th>Subject group</th>
<th>Total (SD)</th>
<th>Listening (SD)</th>
<th>Reading (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPN Mean</td>
<td>422 (61) **</td>
<td>241 (24) **</td>
<td>181 (45) **</td>
</tr>
<tr>
<td>Range</td>
<td>335 - 505</td>
<td>210 - 280</td>
<td>110 - 250</td>
</tr>
<tr>
<td>RTN Mean</td>
<td>844 (103)</td>
<td>460 (35)</td>
<td>384 (77)</td>
</tr>
<tr>
<td>Range</td>
<td>630 - 985</td>
<td>390 - 495</td>
<td>235 - 490</td>
</tr>
</tbody>
</table>

** p < .01

Table 4 TOEIC average scores (RTN1 & RTN2).

<table>
<thead>
<tr>
<th>Subject group</th>
<th>Total (SD)</th>
<th>Listening (SD)</th>
<th>Reading (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTN1 Mean</td>
<td>867 (90)</td>
<td>471 (31) **</td>
<td>396 (68)</td>
</tr>
<tr>
<td>Range</td>
<td>710 - 985</td>
<td>395 - 495</td>
<td>265 - 490</td>
</tr>
<tr>
<td>RTN2 Mean</td>
<td>789 (121)</td>
<td>435 (32)</td>
<td>354 (95)</td>
</tr>
<tr>
<td>Range</td>
<td>630 - 955</td>
<td>390 - 465</td>
<td>235 - 490</td>
</tr>
</tbody>
</table>

** p < .01

proficiency by the official TOEIC test. We also carried out production experiments to investigate the learners’ proficiency in production of English sentences. We held one session for each subject, and the subjects were recorded individually. The utterance list of 25 sentences was arranged in a pseudo-random order. Before the recording, they were provided sufficient time for practice. The subjects were instructed to produce their best English-like utterance at a comfortable speaking rate and to read each sentence through without pausing. When they made a mistake, they were asked to repeat the sentence from the very beginning. For each subject, three repetitions of each sentence were selected from the total of five repetitions uttered. Spectrograms and waveforms were made from these recordings. We carried out acoustic analysis of the linguistic materials, measuring durations of target ISIs (ISI1) and following ISIs (ISI2). The duration of ISI1 was defined as the interval between the onset of the vowel [i] in “Pete”, “Peter”, and “Peterson” and that of the next stressed vowel in “play”, while that of ISI2 was defined as the interval between the onsets of the vowel in “play” and the next vowel. For the purpose of the normalization of the speaking rates, the ratio \( \frac{\text{RTN1}}{\text{RTN2}} \) was obtained. Furthermore, we measured the durations of the target stressed vowel, [i]. The onset of the vowels was defined as the instant a sharp rise appeared in the power of the first formant. Since the target vowel was followed by a stop [t] or a flap [r], the offset of the vowel was defined as the abrupt decrease in power corresponding to the stop closure or the change in power and formant structure corresponding to the flap.

3. Results

As shown in Table 3, the English proficiencies of the Japanese learners were measured by the official TOEIC test. The TOEIC average score of JPN was 422 (Listening Section: 241, Reading Section: 181). The average score of the Japanese learners who had lived in the U.S. (RTN) was 844 (Listening Section: 460, Reading Section: 384). RTN showed much higher scores than JPN in both the total scores and the scores of the Listening and Reading Sections. The results of analysis of variance showed that the differences in the total TOEIC scores \( (F(1, 35) = 181.06, p < .01, \eta^2 = .84) \) and those in both the Listening Section \( (F(1, 35) = 415.39, p < .01, \eta^2 = .92) \) and Reading Section \( (F(1, 35) = 75.99, p < .01, \eta^2 = .68) \) between JPN and RTN were statistically significant.

In order to investigate the effect of the age of arrival—a factor related to the Critical Period Hypothesis (CPH) in the field of L2 acquisition, we divided RTN into two subgroups according to the age of arrival in the U.S.: RTN1 and RTN2. As shown in Table 4, the TOEIC average score of RTN1 was 867 (Listening Section: 471, Reading Section: 396). The average score of RTN2 was 789 (Listening Section: 435,
Reading Section: 354). RTN1 showed higher average scores in the total scores and the scores of the two sections than RTN2. The results of analysis of variance showed that the differences in the total TOEIC scores ($F(2, 34)=99.98, p<.01, \eta^2=.85$) and those in both the Listening Section ($F(2, 34)=252.94, p<.01, \eta^2=.94$) and Reading Section ($F(2, 34)=39.95, p<.01, \eta^2=.70$) among RTN1, RTN2 and JPN were statistically significant. Furthermore, in order to find the source of significant differences in the subject groups, multiple comparison tests (Tukey–Kramer) were carried out. The results showed that the differences in the total score and the score of each section between JPN and RTN1 and those between JPN and RTN2 were statistically significant ($p<.01$). The differences in the scores of the Listening Section between RTN1 and RTN2 were statistically significant ($p<.05$), while the differences in the total scores and those in the Reading Section between these subject groups were not.

Next, we divided the same RTN subjects into two subgroups according to the length of residence in the target language country. Table 5 shows that the TOEIC average score of RTN3 was 887 (Listening Section: 474, Reading Section: 414). The average score of RTN4 was 808 (Listening Section: 450, Reading Section: 358). The average total scores of RTN3 in addition to the average scores of Listening and Reading Sections were higher than those of RTN4. Concretely, there were eight RTN3s out of 11 subjects who scored higher than 800, while eight RTN4s out of 13 subjects did so. Although RTN3 showed higher average scores than RTN4, the results of multiple comparison tests (Tukey–Kramer) showed that these differences between RTN3 and RTN4 were not statistically significant.

Turning to the production of English sentences by the same subjects, we measured the durations of target ISIs (ISI1) and following ISIs (ISI2), as explained in Procedures. For the purpose of the normalization of the speaking rates, we obtained the ratio $\frac{ISI1}{ISI2}$. Figure 1 shows the results for values of the $\frac{ISI1}{ISI2}$ within a word produced by the three subject groups: AMR, RTN, and JPN (Sentences 1, 3, and 5 in Table 2).

In Figure 1, we could observe a durational tendency related to each of the subject groups. The increment in increase in the normalized ISI durations with the additional unstressed syllables differed among the three groups of subjects. RTN showed closer durational patterns to those of AMR. The normalized ISI durations increased by the largest increment for JPN, by a smaller increment for AMR, and by the smallest increment for RTN. The results of analysis of variance for the ISI durations within a word ($\frac{ISI1}{ISI2}$) showed that the differences in the normalized ISI durations among JPN, RTN, and AMR were statistically significant, regardless of the number of syllables in ISI1 (1 syllable: $F(2, 147)=7.40$, $p<.01, \eta^2=.09$; 2 syllables: $F(2, 147)=5.85$, $p<.01, \eta^2=.07$; 3 syllables: $F(2, 147)=5.16$, $p<.01, \eta^2=.07$). The results of multiple comparison tests (Tukey–Kramer) for the ISI durations within a word ($\frac{ISI1}{ISI2}$) showed that the differences in the normalized ISI durations between JPN and AMR were statistically significant, regardless of the number of syllables in ISI1 ($p<.01$). The differences between RTN and AMR were shown to be statistically significant only when ISI1 was one syllable ($p<.01$), but not significant in the case of two syllables and three syllables. The differences in durational patterns in terms of $\frac{ISI1}{ISI2}$ between JPN and
RTN were statistically significant when ISI1 was three syllables ($p<.05$).

Regarding the ISIs between words, AMR and RTN showed similar durational patterns when the ISIs were between words. The results of analysis of variance showed significant differences among JPN, RTN, and AMR in the case of two syllables and three syllables (2 syllables: $F(2, 147)=9.46$, $p<.01$, $\eta^2=.11$; 3 syllables: $F(2, 147)=13.07$, $p<.01$, $\eta^2=.15$). The results of multiple comparison tests (Tukey–Kramer) showed that the differences between AMR and RTN were not statistically significant, while the differences between JPN and RTN were statistically significant in the case of two syllables and three syllables ($p<.01$). Also, the differences between AMR and JPN were statistically significant when ISI1 was two syllables and three syllables ($p<.01$).

In order to investigate the effect of the age of arrival—a factor related to the Critical Period Hypothesis on the production proficiency, we compared the durational patterns for RTN1 and RTN2. Figure 2 shows the durations of ISI within a word ($\frac{ISI}{2}$) for AMR, JPN, RTN1, and RTN2. In this figure, we could observe that AMR and RTN1 showed similar durational patterns, while the patterns for JPN and RTN2 were shown to be similar. The results of analysis of variance showed significant differences in the durations of ISI within a word ($\frac{ISI}{2}$) among JPN, RTN1, RTN2, and AMR (1 syllable: $F(3, 143)=10.26$, $p<.01$, $\eta^2=.18$; 2 syllables: $F(3, 139)=13.66$, $p<.01$, $\eta^2=.23$; 3 syllables: $F(3, 144)=5.98$, $p<.01$, $\eta^2=.11$). The results of multiple comparison tests (Tukey–Kramer) showed that the differences in $\frac{ISI}{2}$ between RTN2 and AMR were statistically significant in the case of one and two syllables ($p<.01$). The differences in $\frac{ISI}{2}$ between JPN and RTN1 were statistically significant when ISI1 was two syllables and three syllables ($p<.01$). The differences between JPN and RTN2 were shown to be statistically not significant. The differences in $\frac{ISI}{2}$ between RTN1 and RTN2 were statistically significant when ISI1 was three syllables ($p<.01$). The values of $\frac{ISI}{2}$ for the ISIs between words showed a similar tendency among the subject groups, which we observed in the normalized durations within words. The results of analysis of variance showed significant differences in the durations of ISI between words ($\frac{ISI}{2}$) among these subject groups (2 syllables: $F(3, 145)=4.54$, $p<.01$, $\eta^2=.09$; 3 syllables: $F(3, 144)=9.01$, $p<.01$, $\eta^2=.16$). The results of multiple comparison tests (Tukey–Kramer) showed significant differences in $\frac{ISI}{2}$ between JPN and RTN1 when ISI1 was two syllables and three syllables ($p<.01$).

Next, we attempted to observe the effect of length of residence in the target language country on production proficiency, and compared the values of $\frac{ISI}{2}$ for RTN3 and RTN4. Figure 3 shows the values of $\frac{ISI}{2}$ for ISI within a word for AMR, JPN, RTN3, and RTN4. In Figure 3, we observe that AMR and RTN3 showed similar durational patterns, while the patterns for RTN4 were relatively closer to those of AMR as compared to those of JPN. The results of analysis of variance showed significant differences among these subject groups (1 syllable: $F(3, 143)=12.29$, $p<.01$, $\eta^2=.20$; 2 syllables: $F(3, 138)=13.65$, $p<.01$, $\eta^2=.23$; 3 syllables: $F(3, 140)=17.88$, $p<.01$, $\eta^2=.28$). The results of multiple comparison tests (Tukey–Kramer) showed that the differences in $\frac{ISI}{2}$ between RTN4 and AMR were statistically significant in the case of one and three syllables ($p<.01$), and the differences between RTN3 and AMR were statistically significant only in the case of one syllable ($p<.05$). The differences in $\frac{ISI}{2}$ between JPN and RTN3 were statistically significant when ISI1 was two syllables and three syllables ($p<.01$). The dif-
A "Critical Period" in Second Language Acquisition Reconsidered

Table 6 Percentages of shortening for the target stressed vowels (vowel shortening from a one-syllable ISI to a two-syllable ISI and three-syllable ISI).

<table>
<thead>
<tr>
<th>ISIs</th>
<th>Between words</th>
<th></th>
<th>Within a word</th>
<th></th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1→2</td>
<td>1→3</td>
<td>1→2</td>
<td>1→3</td>
<td></td>
</tr>
<tr>
<td>AMR</td>
<td>17</td>
<td>28</td>
<td>26</td>
<td>33</td>
<td>26</td>
</tr>
<tr>
<td>JPN</td>
<td>4</td>
<td>13</td>
<td>10</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>RTN</td>
<td>17</td>
<td>28</td>
<td>23</td>
<td>36</td>
<td>26</td>
</tr>
<tr>
<td>RTN1</td>
<td>15</td>
<td>26</td>
<td>20</td>
<td>36</td>
<td>24</td>
</tr>
<tr>
<td>RTN2</td>
<td>22</td>
<td>31</td>
<td>29</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>RTN3</td>
<td>17</td>
<td>29</td>
<td>25</td>
<td>38</td>
<td>27</td>
</tr>
<tr>
<td>RTN4</td>
<td>17</td>
<td>27</td>
<td>21</td>
<td>34</td>
<td>25</td>
</tr>
</tbody>
</table>

Note. 1→2=a one-syllable ISI to a two-syllable ISI; 1→3=a one-syllable ISI to a three-syllable ISI.

Percentages in Table 6 show the degree of vowel shortening for AMR, JPN, and RTN. That is, it shows the percentages of shortening for the target stressed vowels from a one-syllable ISI to a two-syllable ISI and three-syllable ISI. As shown in the table, the stressed vowels were shortened considerably when unstressed syllables were added to the ISIs in the case of AMR and RTN, while JPN exhibited a much smaller percentage of shortening. The average percentages of vowel shortening were 26% for AMR, 26% for RTN, and 12% for JPN. The percentage of the target vowel shortening was shown to be higher in the case of ISIs within a word, as compared to ISIs between words for all subject groups. The results of analysis of variance showed significant differences in the percentage of shortening among AMR, JPN, and RTN (ISIs within a word: one-syllable ISIs, F(2, 47)=6.23, p<.01, \( \eta^2=.21 \); one-syllable ISIs to three-syllable ISIs, F(2, 47)=7.68, p<.01, \( \eta^2=.25 \); ISIs between words: one-syllable ISIs to two-syllable ISIs, F(2, 47)=11.28, p<.01, \( \eta^2=.32 \); one-syllable ISIs to a three-syllable ISIs, F(2, 47)=7.33, p<.05, \( \eta^2=.24 \)). The results of multiple comparison tests (Tukey–Kramer) showed that the differences in the percentage of shortening between AMR and JPN were statistically significant in all cases (p<.01), except that the vowels from a one-syllable ISI to a three-syllable ISI between words were statistically significant.

Differences in vowel shortening between JPN and RTN4 were statistically significant when ISI was two and three syllables (p<.01). Regarding the difference in vowel shortening between RTN3 and RTN4, the statistical difference was only found when ISI was three syllables (p<.01).

In addition to the durational patterns related to each of the subject groups in terms of values of VQI, we obtained the average increment of increase in the normalized ISI durations from one-syllable ISIs to two-syllable ISIs. The average increment in the normalized ISI duration per syllable, as estimated from the correlation curve, shows that normalized ISI durations increased by a larger increment for JPN, rather than RTN and AMR. The results of analysis of variance showed significant differences in the subject groups (ISIs within a word: F(2, 140)=8.88, p<.05, \( \eta^2=.11 \); ISIs between words: F(2, 143)=16.06, p<.01, \( \eta^2=.18 \)). The results of multiple comparison tests (Tukey–Kramer) showed significant differences between JPN and RTN (ISIs within a word: p<.05; ISIs between words: p<.01). Regarding RTN1, RTN2, JPN, and AMR, analysis of variance showed statistically significant differences among the subject groups in the increment of increase both in ISIs within a word (F(3, 139)=8.78, p<.01, \( \eta^2=.16 \)) and ISIs between words (F(3, 142)=13.14, p<.01, \( \eta^2=.22 \)). The results of multiple comparison tests (Tukey–Kramer) showed significant differences between JPN and RTN1 (p<.01). Also, analysis of variance showed that there were significant differences among AMR, JPN, RTN, and RTN4 (ISIs within a word: F(3, 139)=5.88, p<.01, \( \eta^2=.11 \); ISIs between words: F(3, 142)=10.82, p<.01, \( \eta^2=.19 \)). RTN3 and JPN showed the statistically significant differences in the increment of increase both in ISIs within a word and ISIs between words (p<.01). RTN4 and JPN showed statistically significant differences in the increment of increase both in ISIs within a word (p<.01) and ISIs between words (p<.05).

Turning to the target vowel shortening, Table 6 shows the degree of vowel shortening for AMR, JPN, and RTN. That is, it shows the percentages of shortening for the target stressed vowels from a one-syllable ISI to a two-syllable ISI and three-syllable ISI. As shown in the table, the stressed vowels were shortened considerably when unstressed syllables were added to the ISIs in the case of AMR and RTN, while JPN exhibited a much smaller percentage of shortening. The average percentages of vowel shortening were 26% for AMR, 26% for RTN, and 12% for JPN. The percentage of the target vowel shortening was shown to be higher in the case of ISIs within a word, as compared to ISIs between words for all subject groups. The results of analysis of variance showed significant differences in the percentage of shortening among AMR, JPN, and RTN (ISIs within a word: one-syllable ISIs to two-syllable ISIs, F(2, 47)=6.23, p<.01, \( \eta^2=.21 \); one-syllable ISIs to three-syllable ISIs, F(2, 47)=7.68, p<.01, \( \eta^2=.25 \); ISIs between words: one-syllable ISIs to two-syllable ISIs, F(2, 47)=11.28, p<.01, \( \eta^2=.32 \); one-syllable ISIs to a three-syllable ISIs, F(2, 47)=7.33, p<.05, \( \eta^2=.24 \)). The results of multiple comparison tests (Tukey–Kramer) showed that the differences in the percentage of shortening between AMR and JPN were statistically significant in all cases (p<.01), except that the vowels from a one-syllable ISI to a three-syllable ISI between words were statistically significant.
The differences between RTN and JPN were statistically significant in all cases \( (p < .05) \), except for the vowels from a one-syllable ISI to a two-syllable ISI within a word being significant \( (p < .05) \), while those between AMR and RTN were not significant.

Regarding the effect of the age of arrival, we compared the target vowel shortening for RTN1 and RTN2. Figure 4 shows the percentage of shortening of stressed vowels in the ISIs between words for AMR, JPN, RTN1, and RTN2. As shown in the figure, we observe that the subgroups of RTN—RTN1 and RTN2—and AMR showed a similar tendency in the percentage of the target vowel shortening. The average percentage of vowel shortening from a one-syllable ISI to a two-syllable ISI was 18% for RTN1 (ISIs within a word: 20%, ISIs between words: 15%). In the case of RTN2, the average percentage of vowel shortening was 26% (ISIs within a word: 29%, ISIs between words: 22%). The results of analysis of variance showed significant differences among the subject groups (ISIs within a word: one-syllable ISIs to two-syllable ISIs, \( F(3, 46) = 5.13, p < .01, \eta^2 = .25 \); one-syllable ISIs to three-syllable ISIs, \( F(3, 46) = 4.61, p < .05, \eta^2 = .23 \); ISIs between words: one-syllable ISIs to two-syllable ISIs, \( F(3, 46) = 8.97, p < .01, \eta^2 = .37 \); one-syllable ISIs to three-syllable ISIs, \( F(3, 46) = 4.96, p < .05, \eta^2 = .24 \)). The results of multiple comparison tests (Tukey–Kramer) showed that the differences between AMR and RTN1, those between AMR and RTN2, and those between RTN1 and RTN2 were not statistically significant. The difference in the percentage of vowel shortening between RTN1 and JPN was statistically significant in the one-syllable ISIs to two-syllable ISIs between words \( (p < .01) \), and those in the one-syllable ISIs to three-syllable ISIs within a word and between words \( (p < .05) \). The one-syllable ISIs to two-syllable ISIs within a word did not show statistical differences between RTN1 and JPN. Also, the differences in the percentage of vowel shortening between RTN2 and JPN were statistically significant (one-syllable ISIs to two-syllable: \( p < .01 \); one-syllable ISIs to three-syllable ISIs: \( p < .05 \)).

Regarding the effect of the length of residence, we compared the target vowel shortening for RTN3 and RTN4. Figure 5 shows the percentage of shortening of stressed vowels in the ISIs between words for AMR, JPN, RTN3, and RTN4. As shown in the figure, we observe that the subgroups of RTN—RTN3 and RTN4—and AMR showed a similar tendency in the percentage of the target vowel shortening. The average percentage of vowel shortening from a one-syllable ISI to a two-syllable ISI was 21% for RTN3 (ISIs within a word: 25%, ISIs between words: 17%). In the case of RTN4, the average percentage of vowel shortening was 19% (ISIs within a word: 21%, ISIs between words: 17%). The results of analysis of variance showed significant differences among the subject groups in vowel shortening (ISIs within a word: one-syllable ISIs to two-syllable ISIs, \( F(3, 46) = 4.15, p < .05, \eta^2 = .21 \); one-syllable ISIs to three-syllable ISIs, \( F(3, 46) = 4.80, p < .01, \eta^2 = .24 \); ISIs between words: one-syllable ISIs to two-syllable ISIs, \( F(3, 46) = 7.23, p < .01, \eta^2 = .32 \); one-syllable ISIs to three-syllable ISIs, \( F(3, 46) = 4.66, p < .05, \eta^2 = .23 \)). The results of multiple comparison tests (Tukey–Kramer) showed that the differences between AMR and RTN3, those between AMR and RTN4, and those between RTN3 and RTN4 were not statistically significant. The differences in the percentage of vowel shortening between RTN3 and JPN were statistically significant in all cases \( (p < .01) \), except in one-syllable ISIs to two-syllable ISIs within a word and...
A "Critical Period" in Second Language Acquisition Reconsidered

one-syllable ISIs to three-syllable ISIs between words ($p < .05$). Also, the differences in the percentage of vowel shortening between RTN4 and JPN were statistically significant in all but one case ($p < .05$ in all cases except in one-syllable ISIs to two-syllable ISIs between words: $p < .01$).

4. Discussion

In the present study, we examined the L2 production of Japanese learners in terms of ISI durational control and stressed vowel shortening. In both parameters of production, we observed significant differences between the JPN without the experience of living in the target language country and RTN with the experience. The experience of living in the target language country was shown to be a crucial factor in the acquisition of production patterns. The quantity and quality of L2 input are different between JPN and RTN. Regarding the quantity of input, Asher (1977) reported that on average, children, by age 6, have spent a minimum of 17,520 hours (8 hours a day) listening to their L1, while Japanese L2 learners, by age 18, have spent from 1,000 to 2,000 hours (20 minutes a day) of contact with English as L2 (Takefuta 1997). In addition to significant differences in quantity of input, it is obvious that the quality of input is different when learners are surrounded by native speakers of the target language, in contrast to when they are not exposed to native speakers in everyday life. In the present study, the subjects of RTN had lived in the U.S. for over one year at least. This suggests that for successful acquisition, a minimum of one-year living might be required as the amount of necessary exposure to the target language, which is in accordance with the results of our study with junior high school students (Sudo and Kaneko 2005). Several studies on short-term studying abroad programs have reported positive effects on the participants such as global awareness, socio-cultural adaptation, and academic success, but not immediate effects on speaking and listening abilities of the participants (Antonakopoulou 2013, Kurt, Olitisky and Geis 2013, Xu, DeSilva, Neufeldt and Dane 2013), lending support to the necessity of a minimum longer period for the improvement of those abilities.

Regarding an early-late effect on L2 production, we observed that RTN1, rather than RTN2, showed a significant amount of approximation to the production patterns of native speakers. We found significant differences in the ISI durational patterns between the subgroups of RTN which were classified according to the age of arrival, RTN1 being the subgroup of RTN whose age of arrival was before puberty and RTN2, the subgroup whose age of arrival was after puberty. This significant early-late age effect in the ISI production is in line with the results of the previous studies (Baker and Trofimovich 2005, Darcy and Kruger 2012, Flege et al. 1999, Guion 2003, McCarthy et al. 2013) which reported the early-late effect on production. These results confirm that there is a critical period for the acquisition of this specific feature of production.

Turning to another parameter of production, stressed vowel shortening, we found that an early-late age effect was not significant on this parameter of production patterns. Concretely, we did not observe a significant difference among the subgroups of RTN, RTN1, 2, 3, and 4. The two factors of age of arrival and length of residence in the target language country did not result in significant differences in the production patterns of stressed vowel shortening. This result suggests that this specific feature of production can be acquired after the critical period, supporting the results of the studies (Bongaerts et al. 1995, Neufeld 1977, 1979, 1980), which reported even late learners can acquire a native-like production.

We have presented a difference in age effect on the two production parameters. The results indicate that the degree of difficulty in acquiring durational control of stressed vowels is not so high, compared with the acquisition of the ISI durational control. This suggests that the acquisition of unstressed syllables in an ISI is more challenging than that of stressed syllables for Japanese learners of English, offering further support to the results of our study reporting difficulty in acquiring durational control of unstressed syllables (Sudo and Kaneko 2005).

In addition to production parameters, we also measured the English proficiency of Japanese learners by the TOEIC scores. We observed significant differences in the total TOEIC scores and those in both the Listening Section and the Reading Section between JPN and RTN. The results of the TOEIC scores showed that the experience of living in the target language country was a factor which can facilitate the acquisition of listening and reading proficiencies. Regarding the subgroups of RTN, which were classified according to the age of arrival, we observed significant differences in the TOEIC scores between the two subgroups. Concretely, a significant difference between RTN1 and RTN2 was observed in the Listening Section, but not in the Reading Section. The results indicated that the factor of age can be complemented by other factors of learners in
reading ability, while the age factor is a decisive factor in the case of listening ability. This result of the early-late age effect on listening ability is in accordance with the results of previous studies (Flege et al. 1999).

In this study, we attempted to find out why previous studies on an age effect reported different results about an early-late factor. We offered data showing that L2 Japanese learners have different degrees of attainment of L2 skills, depending on which feature of production is the target of acquisition. Stressed vowel shortening was shown to be a feature which can be acquired after puberty, while ISI durational control, especially for the unstressed syllables in the ISI, showed an effect of the early-late dichotomy. Regarding listening and reading abilities measured by the TOEIC, listening proficiency was shown to have an early-late age effect, while reading proficiency could be acquired after puberty. To sum up, the results of this study lend partial support to the existence of a “critical period” in L2 acquisition.

The factor of length of residence in the target language country was shown to have an effect on ISI durational control, but not on compensatory vowel shortening. We observed that RTN3 with the length of residence of six years or more, rather than RTN4, showed a significant amount of approximation to the production patterns of native speakers. The factor of length of residence was shown to be not significant in the TOEIC scores of RTN3 and RTN4, which is in accordance with the results of Tahta et al. (1981), revealing that length of residence was not identified as a significant predictor. In other words, we observed significant differences in the TOEIC scores of the Listening Section between RTN1 and RTN2, but not between RTN3 and RTN4. Although the number of the subjects in this study was not large, and the age of arrival and the length of residence were not quantitatively matched with each other, the results can still be interpreted as indicating that the age of arrival is a factor that overrides the length of residence in regard to the English proficiency measured by the TOEIC scores. Even if the length of residence was not long, some of the learners showed proficient listening and reading abilities. It should be noted that this factor could be complemented by other factors of learners, such as motivation, aptitude, attitude, personality among others, provided that the learners have the experience of living in a target language country, regardless of the length of time. Here, we should consider some limitations of this study. There remains a possibility that the different degrees of significance observed in the age of arrival and the length of residence may be attributed to the combined effects of the two factors. Further research, with a larger number of subjects, is needed, especially to analyze the main factor in different degrees of acquisition attainment.

5. Conclusion

The present study investigated the hypothesis of an age effect, the Critical Period Hypothesis, on L2 production and listening/reading abilities by Japanese learners of English, and showed that a critical period exists for some aspects of L2 acquisition, but not for others. Concretely, we found a critical period for the acquisition of listening ability and ISI durational control, including production of unstressed syllables; we also found that specific features of production, such as stressed vowel shortening and reading ability could be acquired after the critical period.

Acknowledgement

This research was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology [25370700].

References


A "Critical Period" in Second Language Acquisition Reconsidered


Munhall, K., C. Fowler, S. Hawkins and E. Saltzman (1992)

(Received Sep. 5, 2014, Accepted Oct. 14, 2015)