Foreign accent in English words produced by Japanese children and adults

Katsura Aoyama[1], James E. Flege[2], Susan G. Guion[3], Reiko Akahane-Yamada[4], Tsuneo Yamada[5]

[1] Texas Tech University Health Sciences Center, USA

[2] University of Alabama at Birmingham, USA

[3] University of Oregon, USA

[4] ATR Human Information Science Laboratories, Japan

[5] National Institute of Multimedia Education, Japan

E-mail: katsura.aoyama@ttuhsc.edu, jeflege@uab.edu, guion@oregon.uoregon.edu, yamada@atr.co.jp, yamada@nime.ac.jp

ABSTRACT

This study examined the pronunciation of English words produced by 16 native Japanese (NJ) adults and 16 NJ children living in the United States and groups of age-matched native English (NE) controls (16 each). Recordings were made twice, one year apart (Time 1, Time 2). The NJ participants had lived in the United States for 0.5 years at Time 1. Words produced by the four groups were rated for overall degree of foreign accent using a scale ranging from 1 ("strongest foreign accent") to 9 ("least foreign accent"). The NJ children's pronunciation of English, but not that of the NJ adults, was found to have improved significantly over the 1-year study interval. However, the results suggested that the rate at which children's pronunciation of a second language improves is not as rapid as is commonly believed.

1. INTRODUCTION

As observed by Snow [1], it is widely believed that "children can learn a foreign language more quickly, easily and well" than adults. However, contrary to this common view, some studies showed that children may not learn a second language (L2) easily or well. For example, in a study by Snow & Hoefnagel-Höhle [2], English-speaking adults outperformed children between the ages 3-10 years on various tasks in Dutch when tested within 6 months of their arrival in the Netherlands. In addition, recent studies indicated that adults who began learning their L2 in childhood differed from native speakers of L2 in terms of foreign accent ratings, although their overall degree of foreign accent was not as strong as other adults who begin learning the L2 in adulthood [3, 4].

Children's advantage in learning an L2 may be due to social factors rather than simply their chronological age. In the Snow & Hoefnagel-Höhle study cited earlier [2], for instance, all the children examined were enrolled in a school where the target L2 was used daily. The adult participants, on the other hand, typically spent their days in environments where the L1 (English) was used frequently.

Jia & Aaronson [5] proposed that children are generally more successful in learning an L2 because they are more likely to switch preference from their L1 to the L2 than adults. All the Chinese children examined by Jia & Aaronson preferred to use their L1 upon arrival in the U.S. However, children younger than 8 years of age switched their preference to English after one year, whereas children older than 12 years retained their preference for the L1. Jia & Aaronson hypothesized that such shifts in language preference cause younger learners to receive more L2 input than older learners.

The aim of this paper was to report changes in English pronunciation over one year by NJ adults and children. The NJ participants were tested approximately 0.5 years and 1.6 years after their arrival in the U.S. This permitted us to determine if adult-child differences exist at early stages of L2 learning.

2. METHOD

2.1. Participants

Sixteen NJ adults and 16 NJ children who were living in the U.S. participated. Most NJ participants were living in Houston or Dallas, Texas when tested. Sixteen adults and 16 children who spoke English as their native language also participated as a control group. The NE adults and children did not speak any language other than English. All 64 participants were tested twice with one year between (Time 1, Time 2). At Time 1, the NJ participants' average length of residence (LOR) in the U.S. was approximately 0.5 years; at Time 2 their LOR averaged 1.6 years (see Table 1). There were approximately equal numbers of female and male participants in each of the four groups. All participants were tested in a quiet room at their homes or at the University of Alabama at Birmingham.

ISBN 1-876346-48-5 © 2003 UAB

2.2. Elicitation Procedures

Twenty-six frequently occurring English words were recorded for each participant in the following manner. The participants saw a picture on the screen of a laptop computer and heard the corresponding word via a loudspeaker. An equivalent word in Japanese was displayed in Japanese orthography along with the picture in order to reduce uncertainty as to what word was to be produced. The 26 pictures were presented three times in a different random order for each participant. An auditory model was usually not provided to elicit the second and third tokens of the test words. The experimenter played out the auditory model of the word only when the participant was unable to say the word.

	Gender	Mean Age at T1	Mean LOR T1	T2
NE adults	7m 9 f	40.3 (4.7)	--	–
NJ adults	8m 8 f	39.9 (3.8)	0.5 (0.2)	1.6 (0.3)
NE children	10 m 6 f	10.6 (2.1)	--	–
NJ children	9m 7 f	9.9 (2.4)	0.4 (0.2)	1.6 (0.3)

Table 1: Characteristics of the NE and NJ participants. LOR = Length of residence in the United States in years. Age = Chronological age in years. Standard deviations are in parentheses.

The 13 words (*eight, neck, read, six, book, dog, light, wing, egg, fish, leaf, watch, write*) were chosen because all the participants were able to produce them without an auditory cue. The participants' third productions of each word was digitized (22.05 kHz, 16-bit resolution), then normalized for peak intensity (50% of the full scale). The third token of each word was examined because these tokens were considered to be closest to spontaneous speech.

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

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

The words in each of the strings contained a variety of initial consonants, final consonants and vowels. There was approximately 200 ms. of silence between the words. The word-strings sounded like a list of words because each word was produced in isolation. There was thus no intonation contour across the words.

2.3. Listener judgments

The three word-strings were rated by adult monolingual native speakers of English for overall degree of foreign accent. The 16 listeners (9 males, 7 females) had a mean age of 33 years. All were required to pass a pure-tone hearing screening (20 dB HL) prior to participating. None of the listeners had any special training in speech or language. Seven of the 16 listeners were from Alabama, and 9 were from Texas.

Each listener was tested individually in a sound booth. The word-strings were presented at a self-selected comfortable level via loudspeakers. Each listener participated in two sessions held on separate days, each lasting about 75 minutes.

The word-strings produced by the adults and children were presented in separate blocks, yielding six blocks of stimuli in all. Three blocks were presented on the first day, and the remaining three blocks were presented on the second day of testing. Half of the listeners rated blocks of stimuli produced by children followed by blocks of stimuli produced by adults; the remaining listeners rated the adults' blocks first. The order of the word-strings was counterbalanced within the adult or child blocks.

The listeners rated each word-string using a scale that ranged from 1 ("strongest foreign accent") to 9 ("least foreign accent"). The listeners were required to rate each word-string before moving on to the next word-string. The inter-stimulus-interval was 1 s. A stimulus could be re-played, but ratings could not be changed once given.

The 64 stimuli in each block (2 groups x 16 participants x 2 times of testing) were randomly presented three times each to each listener. A different random order was used in each of the six blocks for all 16 listeners. Fifteen extra trials were given as practice in the beginning of each block, but judgments of these trails were not analyzed. The median rating of the three presentations was obtained for each word-string and for each listener. These median ratings were used in analyses.

2.4. Analyses

The procedure just described yielded a total of 6144 median ratings for analysis (4 groups x 16 participants x 2 times of testing x 3 word-strings x 16 listeners). Intraclass correlations were high in all 6 blocks ($R = 0.76$ to 0.90, $p < 0.001$), suggesting that there was an acceptably high level of inter-rater agreement.

Two types of scores were derived from the ratings. The "talker-based" scores were computed for each participant by averaging over the ratings given by the 16 listeners. This yielded 384 talker-based scores (4 groups x 16 participants x 3 word-strings x 2 times). There was no guarantee that the results obtained in analyses of the talker-based scores would generalize to other listeners, so "listener-based" scores were also computed. These scores were computed by averaging over the scores given by each listener had given to the 16 participants in each group. This procedure yielded 384 listener-based scores (16 listeners x 4 groups x 3 word-strings x 2 times). A difference was not considered statistically significant unless it was significant in analyses of both talker-based and listener-based scores.

3. RESULTS

The talker-based scores for each participant at Time 1 and Time 2 were submitted to a (2) Age x (2) Native Language x (2) Time of Testing x (3) Word-string ANOVA. This analysis yielded main effects of Language (F(1,60) = 683.4, p < 0.01), Time (F(1,60) = 19.4, p < 0.01) and Word-string (F (2, 120) = 11.9, p < 0.01). It also yielded a significant three-way interaction, Language x Age x Time (F(1, 60) = 14.5, p < 0.01), and a two-way interaction, Language x Word-string (F(2, 120) = 8.5, p < 0.01). The four-way interaction was non-significant.

Analysis of the listener-based scores yielded similar results. These scores were also examined in (2) Age x (2) Native Language x (2) Time of Testing x (3) Word-string ANOVA. This analysis yielded main effects of Language (F(1,60) = 856.0, p <0.01), Time (F(1,60) = 161.0, p < 0.01) and Word-string (F (2, 120) = 14.0, p < 0.01). It also yielded a significant three-way interaction, Language x Age x Time (F(1, 60) = 69.8, p < 0.01) and a two-way interaction, Language x Word-string (F(2, 120) = 9.9, p < 0.01). The four-way interaction was again non-significant.

In the listener analysis, a two-way interaction Language x Age (F(1,60) = 9.6, p < 0.01), and a three-way interaction Language x Time x Word-string (F(2, 120) = 5.12, p < 0.01) were also significant. These interactions will not be examined further, because they were not significant in the talker analysis.

In the following section, only the differences that were significant in the analyses of both talker- and listener-based scores will be discussed. The mean values obtained for each group in the talker and listener analyses were, of course, the same. However, the two analyses yielded different standard deviations, *F*-values, and *p*-values. Only the *F*- and *p*-values generated in analyses of the talker-based scores will be presented here due to space limitations.

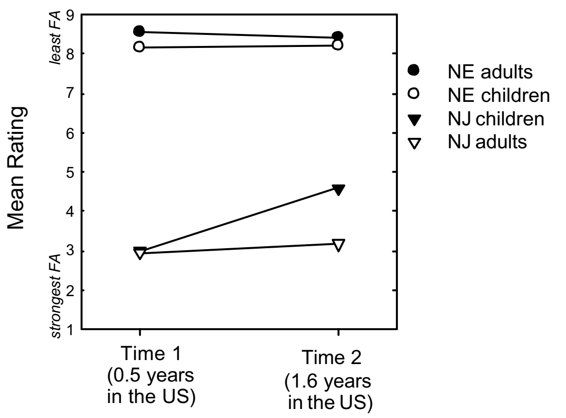

Figure 1: Mean foreign accent ratings given by the NE listeners to English word-strings produced by the NE participants or by the NJ participants.

The Language x Age x Time interaction was explored by averaging the ratings for three different word-strings for each participant group at Time 1 and Time 2 (see Figure 1). The simple effect of Time was significant for the NJ children (F (1,15) = 17.9, p < 0.01), but not the NJ adults, the NE adults or the NE children (F (1,15) = 3.5 to 4.7, p > 0.05). This suggested that the NJ children's scores improved from Time 1 to Time 2 (*means* 3.0 and 4.6 respectively), but not the NJ adults' scores (*means* 2.9 and 3.2 respectively).

The simple effect of Age was significant for the NE group at Time 1 (F (1,30) = 7.3, p < 0.05), suggesting that the NE children's scores (*mean* 8.2) were significantly lower than the NE adults' (*mean* 8.6) at Time 1. At Time 2, the difference between the NE children's and adults' scores was non-significant (*means* 8.2 vs. 8.4, F (1,30) = 2.4, p > 0.1). The simple effect of Age was non-significant for the NJ participants at Time 1 (*means* children 3.0, adults 2.9, F (1,30) = 0.07, p > 0.1). However, at Time 2, the NJ children's scores were significantly higher than the NJ adults' (*means* 4.6 and 3.2, F (1,30) = 8.6 p < 0.01).

The simple effect of Language was significant for all Age x Time combinations (F (1,30) = 60.8 to 811.1, p < 0.01), indicating that the NJ adults' and children's scores were significantly lower than the NE adults' and children's at Time 2 as well as at Time 1.

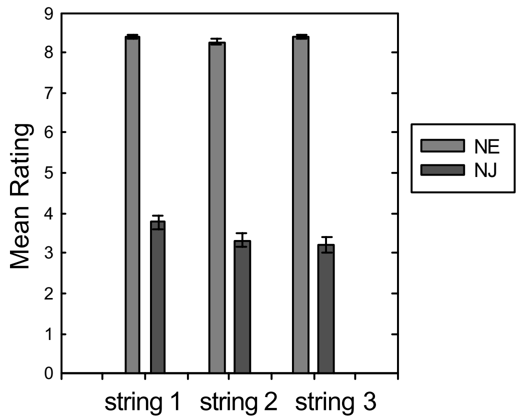

Figure 2: Mean foreign accent ratings for three different English word-strings produced by the NE and NJ participants.

The Language x Word-string interaction was explored by averaging over age (adult, child) and time (Time 1, Time 2) for the NE and NJ groups (Figure 2). For NE participants, the differences among the ratings for the three word-strings were non-significant (F (2, 31) = 2.1, p > 0.1). However, differences among the three word-strings were significant for the NJ participants (F(1, 31) = 12.4, p < 0.01). Tukey's tests indicated that the ratings for word-string 1 was significantly higher than the ratings for strings 2 and 3 (*means* 3.8 vs. 3.3 and 3.2, respectively). This suggested

ISBN 1-876346-48-5 © 2003 UAB

that foreign accent may be more evident in some words than others, and that the listeners' evaluations of a non-native speaker might change somewhat depending on the words that are being considered.

4. DISCUSSION

The results of this study suggested that the NJ children's pronunciation of English improved over one year of residence in the U.S., whereas the NJ adults' pronunciation did not change measurably. There are several possible bases for this finding. First, children may indeed learn foreign languages 'more quickly and easily' than adults [see 1]. However, it is important to note that the NJ children continued to receive much lower ratings than the NE children even after 1.6 years of U.S. residence. Children may learn an L2 more rapidly than adults, but their learning may not be as rapid as is commonly believed.

Second, the adult-child difference seen at Time 2 may have been due to differing amounts of exposure to English received by the NJ adults and children. The NJ children all attended English-medium schools, whereas some of the NJ adults had little opportunity or need to speak English on a daily basis. The NJ children were 6 to 13 years of age (*mean* 9.9 years) at Time 1. Based on Jia & Aaronson's findings, it is possible that at least some of them switched their language preference from Japanese to English, or felt equally comfortable using their L1 and L2 by Time 2. The improvement in foreign accent ratings among the NJ children may have been due in part to the increased input that accompanies a switch in language preference.

A somewhat unexpected finding of this study was a difference in ratings among the three word-strings produced by the NJ participants. The higher mean ratings on word-string 1 suggested that the NE listeners judged string 1 to be less foreign-accented than the other two strings. All words examined here were monosyllabic, and each word-string contained a variety of initial and final consonants. One possible reason for the difference in ratings may be the initial /w/. Word-strings 2 and 3 both contained words beginning in /w/ (*wing, watch*), but Word-string 1 did not. Both Japanese and English possess a /w/, but these sounds are phonetically different in the two languages. Unlike English /w/, the /w/ of Japanese is less lip-rounded [6]. Previous research suggested that the phonetic differences between L1 and L2 categories are less likely to be discerned when there is an 'equivalent' L1 category [7]. As a result, L2 sounds that have an equivalent sound in learners' L1 may become a source of foreign accent. One possibility suggested by our findings is that the difference between English /w/ and Japanese /w/ was small and the NJ participants perceived them as 'equivalent'. If they indeed used Japanese /w/ in English words, it might have contributed to the detection of foreign accent. However, this issue should be examined in greater detail in future research.

5 SUMMARY

In summary, the results of this study suggested that NJ children's pronunciation of English improved significantly over one year of residence in the U.S. NJ adults' pronunciation was not observed to change measurably in the same time period. The results also suggested that, despite the improvement over time noted for the NJ children, their pronunciation of English remained foreign-accented after 1.6 years of U.S. residence.

6. ACKNOWDEDGEMENTS

This research was supported by NIH Grant DC 00027. The authors thank Linda Legault for her help.

REFERENCES

[1] C. Snow, "Relevance of the notion of a critical period to language acquisition," in *Sensitive periods in development: interdisciplinary perspectives*, M. H. Bornstein, Ed. Hillsdale, NJ: Lawrence Erlbaum, 1987, pp. 183-209.

[2] C. E. Snow and M. Hoefnagel-Höhle, "The critical period for language acquisition: evidence from second language learning," *Child development*, vol. 49, pp. 1114-1128, 1978.

[3] T. Piske, I. R. A. MacKay, and J. E. Flege, "Factors affecting degree of foreign accent in an L2: A review," *Journal of Phonetics*, vol. 29, pp. 191-215, 2001.

[4] G. H. Yeni-Komshian, J. E. Flege, and S. Liu, "Pronunciation proficiency in the first and second languages of Korean-English bilinguals," *Bilingualism: Language and cognition*, vol. 3, pp. 131-149, 2000.

[5] G. Jia and D. Aaronson, "Age differences in second language acquisition: the dominant language switch and maintenance hypothesis," in *Proceedings of the 23rd Annual Boston University Conference on Language Development*, vol. 23. Somerville, MA: Cascadilla Press, 1999, pp. 301-312.

[6] T. Vance, *An introduction to Japanese phonology.* New York, NY: State University of New York Press, 1987.

[7] J. E. Flege, "The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification," *Journal of phonetics*, vol. 15, pp. 47-65, 1987.