

Age effect on acquisition of non-native phonemes: perception of English /r/ and /l/ for native speakers of Japanese

Reiko A. Yamada, & Yoh'ichi Tohkura

ATR Auditory and Visual Perception Research Laboratories

ABSTRACT

This study investigates the age effect on acquisition of American English (AE) /r/ and /l/ perception by native speakers of Japanese who have once been exposed to the AE speaking environment. A perceptual experiment designed to test the ability to identify naturally spoken /r, l, w/, and determine perceptual cues when identifying those phonemes using synthesized stimuli was performed for native AE subjects, and native Japanese subjects with and without the experiences of living in the U.S. The results show that some of the Japanese subjects who had resided in the U.S. acquired /r/ and /l/ perception, and that acquiring capability of acquiring decreases from 7 to 13 years of age.

1. INTRODUCTION

Many studies have revealed that phoneme perception is modified by the linguistic environment. The perception of American English (AE) /r/ and /l/ sounds for Japanese speakers is one of the strongest pieces of evidence that this is so. In the phonological system of Japanese, the AE /r/ and /l/ contrast is not distinctive, and neither AE /r/ and /l/ resemble any Japanese phonemes. Thus, most Japanese speakers have considerable difficulty in acquiring /r/ and /l/ contrast even though they start learning English in junior high school at about age 12.

Previous cross-linguistic studies using a synthetic /r-l/ stimulus series revealed that native speakers of Japanese had difficulties in perceptually differentiating

these two phonemes, and that they perceive the synthetic /r-l/ series continuously, even though native AE speakers perceive them categorically (e.g. [5, 6, 7, 9]). Furthermore, the perceptual cue for distinguishing /r/ from /l/ is different between AE speakers and Japanese speakers: AE speakers use F3 frequency as a predominant cue, and Japanese speakers use both F2 and F3 frequencies [12]. The effect of being exposed to an English speaking environment has also been studied [1]. This study revealed the effect of age on the /r,l/ acquisition. However, further control of the starting age and period of exposure are needed to understand the nature of acquisition process. Furthermore, the age of the subjects during participation in the experiment should also be controlled (e.g. it varied from 3 to 45 years of age in [1]), because the performance of children and adults may be expected to differ considerably.

This paper investigates the age effect on acquisition of AE /r/ and /l/ phonemes for native adults of Japanese by controlling the starting age and period of exposure to the AE speaking environment more precisely than previous studies. To determine the precise perceptual mode of the subjects, the identification tests not only of naturally spoken stimuli, which were designed to see overall identification ability, but also of synthesized stimuli, which were designed to investigate the perceptual cue, were performed. Furthermore, in this paper, the /w/ phoneme is considered in addition to /r/ and /l/, because Japanese listeners often identify some of the /r/ and /l/ sounds as /w/ [13].

2. STIMULI

Synthesized /rait-lai/ series generated by Klatt's cascade formant synthesizer, and naturally spoken stimuli were used. Figure 1 provides a synthetic spectrographic representation of the initial CV portion, /rai-lai/, for the synthesized stimuli. The acoustic parameters for idealized "right" and "light" were derived from the naturally spoken /rait/ and /light/ uttered by a native male speaker of AE. When generating the stimuli, three acoustic parameters, F2 and F3 onset frequencies and F1 transition duration, were varied. To construct the stimuli on F2-F3 plane, a variety of F2 and F3 onset frequency combinations were used. The F2 and F3 onset frequencies were varied independently from 800 Hz to 1400 Hz in 200 Hz steps, and 1200 Hz to 3000 Hz in 200 Hz steps, respectively. There were 37 combinations in total, excluding some contradictory combinations in which the F2 frequency was equal to or higher than the F3 frequency. F1 transition duration was varied from 70 ms to 16 ms in 6ms steps as the F3 onset frequency was varied from 1200 Hz to 3000 Hz. In all synthesized stimuli, the acoustic parameters for the vowel part /ai/ were common, and the duration of the /rai-lai/ part was fixed at 360 ms. The stimuli were synthesized and reproduced through 16-bit digital-analog conversion at a sampling frequency of 20 kHz and low-pass filtering with a cutoff frequency of 10 kHz. Several experiment sessions (i.e. with different stimulus randomizations) were recorded on a digital audio tape using a DAT recorder, SONY DTC-1000ES. Each session consisted of eleven blocks of ten trials and one block of one trial to make 111 trials in total. The 111 trials resulted from three randomly ordered repetitions of each of 37 stimuli. The inter-trial interval was 2 seconds, and the inter-block interval was 8 seconds. The block start signal was a beep sound recorded 2

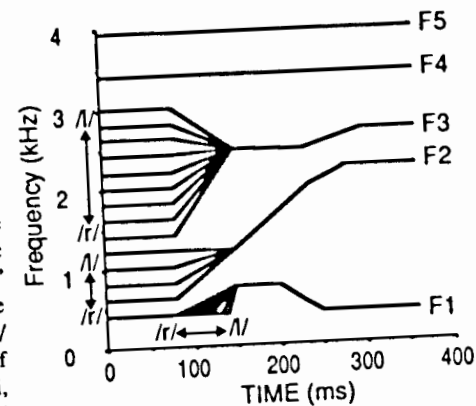


Figure 1 Schematic representation of frequency trajectories of F1 to F5 for the synthesized stimuli.

seconds prior to the beginning of each block. Naturally spoken stimuli contained sixteen combinations of English words. Each combination consisted of three words which were different from each other only in the initial consonant, i.e., /r/, /l/, or /w/ (e.g. "red", "led", and "wed"). The forty-eight words were spoken by two native AE speakers (one female and one male) to produce a total of ninety-six stimuli. They were recorded and converted from analog to digital at a 20-kHz sampling frequency with 16-bit accuracy. These stimuli were reproduced and were recorded on digital audio tape. In each session, each of 96 stimuli occurred once in random order to make 96 trials in total, and these 96 trials were arranged in nine blocks of ten trials and one block of six trials. Other conditions were identical to the identification tests of the synthesized stimuli.

3. SUBJECTS

One hundred and twenty native speakers of Japanese who have never lived abroad (Group J), 109 native speakers of Japanese who have resided in the U.S. (Group JE), and 9 native speakers of AE (Group A) served as subjects. Criterion for participation in the experiment as Group JE subjects was to fulfill all the following conditions: (1) native speaker of Japanese,

(2) had once lived on the U.S. mainland for more than 1 year, (3) had never lived in a foreign country other than the U.S., (4) speaks AE all the time at school, pre-school or kindergarten, or in business, (5) goes to school or conducts business under condition (4) at least 5 days a week, (6) received no special training for speaking AE in Japan. The start of their residence in the U.S. can roughly be thought to coincide with the start of their exposure to the AE speaking environment because English education in Japanese high schools is biased toward grammar, reading, and writing, and is mainly conducted by Japanese teachers. The age of the subjects in Group J is 19 on average, and ranged from 15 to 23, that in Group JE is 20 on average, and ranged from 13 to 40, and that in Group A is 25 on average, and ranged from 20 to 41. All the subjects reported no history of hearing or speaking disorder.

4. PROCEDURE

Each listener participated in two sessions of identification tests for synthesized stimuli, and one session of identification test for naturally spoken stimuli. In these tests, listeners were instructed to identify the word initial consonant, and to make a forced choice among the given categories regardless of the frequency of occurrence for each category through an entire session by checking a corresponding response category on an answer sheet. In the identification test for naturally spoken stimuli, listeners were also told that there might exist unfamiliar or meaningless words, but they should only identify the initial consonant.

5. RESULTS

After the identification rates for each stimulus were calculated, the values C_s and C_n

were obtained as perceptual ability scores of synthesized stimuli and that of naturally spoken stimuli, respectively. As AE listeners identify the stimuli whose F3 onset frequencies were higher than 2000Hz as /l/ and those which were lower as /r/ (Yamada & Tohkura, 1990), the C_s represents the averaged response rates of /l/ for the stimuli whose F3 onset frequencies were equal to or higher than 2000Hz and /r/ for the other stimuli ($0 \leq C_s \leq 1$). The C_n is the averaged correct response rates across all the naturally spoken stimuli ($0 \leq C_n \leq 1$).

The averaged C_s across Group A subjects was .91, and ranged from .75 to 1.00, that across Group J subjects was .48, and ranged from .21 to .76, and that across Group JE subjects was .74, and ranged from .32 to .99. The averaged C_n across Groups A was 1.00, that across Group J was .67, and ranged from .44 to .95, and that across Group JE was .87, and ranged from .55 to 1.00. In the histograms of both C_s and C_n values, two

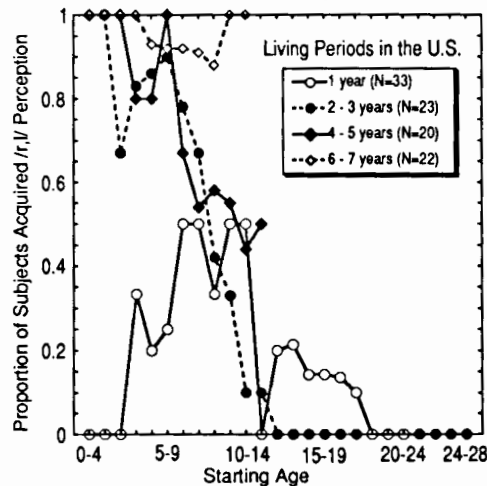


Figure 2 Capability of acquiring /r,l/ phonemes for Japanese speakers who have once been exposed to the AE speaking environment. The proportion of the subjects' number who have acquired /r,l/ perception for four groups of living periods (1 year, 2 - 3 years, 4 - 5 years, and 6 - 7 years) are represented as a function of starting age. As moving averages for each 5-year period are represented, the abscissa shows the average period.

peaks were observed in Group JE, even though only one peak was observed in Group A and J.

The Group JE subjects were divided into two groups according to their C_s and C_n values as follows: acquired group (subjects whose C_s and C_n values are; $0.75 \leq C_s$, and $0.90 \leq C_n$), and non-acquired group (the other subjects). In order to observe the correlation between the acquisition performance and the age of exposure to the AE speaking environment, the probabilities of acquired group subjects among subjects who have started living in the U.S. at the same age were calculated. JE subjects were classified into groups according to their living periods, and the following four groups was represented in Figure 2: subjects lived in the U.S. for 1 year, 2 - 3 years, 4 - 5 years, and 6 - 7 years. As the living conditions (starting age and period of residence) are not fully controlled, and the number of subjects for each data point was not insufficient, we plotted the moving averages under the following conditions: the average age upon taking up residence was 5 years, and the shift period was 1 year. Age noticeably age affected acquisition performance. The acquisition probability decreased rapidly from 7 to 13 years of age. This result is especially obvious in the 2 - 3 years old, in which living conditions are better controlled than in the other groups. Eleven subjects have resided in the U.S. more than 8 years, only one of them, who have resided in U.S. for 8 years from 25 years old, failed to acquire /r,l/ perception.

6. DISCUSSION

Showing that result that the acquisition probability decreased with age was consistent with many previous studies of phoneme acquisition (e.g. [1, 2, 3, 4, 10]). The age when the capability of acquisition decreases in the present study was also similar to results of previous studies on second language production (e.g. [8, 11]). The relationship between acquisition of perception and that of production is of great inter-

est. The productions of /r,l/ phonemes for all the subjects in the present experiment were recorded after the perception test. We also plan to analyze the production characteristics of the present subjects and to study the relationship between acquisition of perception and that of production. In addition, further efforts to obtain data from Japanese subjects with greater variations in living conditions are required.

7. References

- [1] Cochrane, R.M. (1980). "The acquisition of /r/ and /l/ by Japanese children and adults learning English as a second language." *J. Multiling. Multicult. Development*, 1, 331-360.
- [2] Flege, J. (1990). "Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language." *J. Acoust. Soc. Am.* 89, 395-411.
- [3] Flege, J. and Eefting, W. (1987). "Production and perception of English stops by naive Spanish speakers." *J. Phonet.* 15, 67-83.
- [4] Mack, M. (1989). "Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals." *Percept. Psychophys.* 46, 187-200.
- [5] McKain, K.S., Best, C.T. & Strange, W. (1981). "Categorical perception of English /r/ and /l/ by Japanese bilinguals." *Appl. Psycholinguist.*, 2, 369-390.
- [6] Miyawaki K., Strange, W., Verbrugge, R., Liberman, A.M., Fujimura, O. & Jenkins, J. (1975) "An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English." *Percept. Psychophys.*, 18, 331-340.
- [7] Mochizuki, M. (1981). "The identification of /r/ and /l/ in natural and synthesized speech." *J. Phonet.* 9, 283-303.
- [8] Oyama, S. (1976). "A sensitive period for the acquisition of a nonnative phonological system." *J. Psycholinguistic Res.* 5, 261-283.
- [9] Strange, W., & Dittmann, S. (1984) "Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English." *Percept. Psychophys.* 36, 131-145.
- [10] Tahta, S., Wood, M., & Loewenthal, K. (1981). "Foreign accents: factors relating to transfer of accent from the first language to a second language." *Lang. Speech* 24, 265-272.
- [11] Tahta, S., Wood, M., & Loewenthal, K. (1981). "Age changes in the ability to replicate foreign pronunciation and intonation." *Lang. Speech*, 24, 363-372.
- [12] Yamada, R.A., & Tohkura, Y. (1990). "Perception and production of syllable-initial English /r/ and /l/ by native speakers of Japanese." *Proceedings of ICSLP '90*, 757-760.
- [13] Yamada, R.A., & Tohkura, Y. (1991). "The effect of experiment variables on the perception of American English /r,l/ for Japanese listeners: Response categories, classification of subjects, and stimulus range." submitted to *Percept. Psychophys.*