

PERCEPTUAL CHANGE IN ADULTS: ACOUSTIC FACTORS AFFECTING THE IDENTIFICATION OF ENGLISH /r-l/ BY JAPANESE LISTENERS

John S. Logan, Kathy BharrathSingh
Department of Psychology, Carleton University, Ottawa

1. INTRODUCTION

Perception of the English phonemes /r/ and /l/ by Japanese listeners can serve as a model system to explore some of the factors underlying the capacity for perceptual change in adult listeners. Japanese listeners generally have difficulty perceiving the /r/ - /l/ distinction in English, especially when these phonemes are in word-initial positions (MacKain, Best, & Strange, 1981; Mochizuki, 1981). Logan, Lively, and Pisoni (1991) and Lively, Logan, and Pisoni (1993) demonstrated that Japanese listeners can learn to identify /r/ and /l/ more accurately after completion of a laboratory training procedure, indicating that adult listeners can redirect their attention to novel acoustic cues. However, many questions remain about how Japanese listeners learn to direct their attention to the acoustic cues necessary to identify /r/ and /l/.

A critical factor in the perception of /r/ and /l/ by Japanese listeners is phonetic context (Mochizuki, 1981; Sheldon & Strange, 1982). Performance is better when the phonemes are located in word-final positions than in word-initial positions. Acoustic analyses suggested that this effect is due to differences in the duration of /r/ and /l/ in these two contexts (Dissosway-Huff, Port, & Pisoni, 1982; Sheldon & Strange, 1982): /r-l/ segments located in word-final positions tend to be longer in duration than in word-initial positions. This explanation for the effects of phonetic context in the perception of /r/ and /l/ by Japanese listeners is referred to as the duration hypothesis.

Several factors, however, mitigate against the complete acceptance of the duration hypothesis. First, the evidence for it is primarily correlational and somewhat equivocal (i.e., /r-l/ duration does not account for all of the variation in perception by listeners). Second, Cantonese listeners, who also have difficulty perceiving the /r-l/ contrast show a positional asymmetry that is opposite to that observed in Japanese listeners (i.e., Cantonese listeners perceive /r-l/ in word-initial position more accurately than in word-final position).

Logan and BharrathSingh (1994) sought to directly evaluate the role of duration in the perception of /r/ and /l/ by Japanese listeners. A group of Canadian English (CE) and Japanese listeners were presented a series of naturally produced English words contrasting /r/ and /l/. The /r-l/ segments were either shortened or lengthened using a waveform editor. Results indicated that the duration manipulation did not have any effect on the identification performance for either group of listeners. Moreover, rating data from the CE listeners suggested that the quality of the stimuli may have been compromised by the waveform editing procedure. Accordingly, the null result that was obtained for the duration manipulation was not interpreted as indicating the duration hypothesis was incorrect but instead that the experiment did not provide an adequate test of the hypothesis.

The present experiment was designed to address some of the problems associated with using naturally-produced stimuli. Synthesized speech was used to permit the manipulation of /r-l/ duration without the difficulties inherent in modifying natural speech. In addition, it enabled a more fine-grained manipulation of the acoustic cues underlying duration by enabling the independent*

variation of /r-l/ steady duration and /r-l/ formant transition duration in order to determine whether one aspect of duration was more salient to listeners than another.

As well, the present experiment utilized an additional group of listeners. In addition to CE and Japanese listeners, Cantonese listeners were also tested. The inclusion of the Cantonese listeners provided an extra control group for comparison with the Japanese listeners. Moreover, because the Cantonese listeners also have difficulty perceiving /r/ and /l/, albeit in another phonetic context, they serve to demonstrate the potential interaction between acoustic information and the listener's phonological system.

2. METHOD

Stimulus preparation and experimental control utilized CSRE software (Jamieson, Ramji, Neary, & Baxter, 1990). RIGHT and LIGHT stimuli were synthesized using the Klatt software synthesizer (Klatt, 1980) as implemented in CSRE. Parameters for these stimuli were derived from a combination of Klatt's parameters and measurements from natural tokens. Each synthesized /t/ segment also included a naturally-produced burst.

Additional stimuli were constructed by varying two parameters in the original RIGHT and LIGHT stimuli: 1) the duration of the steady-state portion of the /r-l/ segments (referred to as the steady-state manipulation) and 2) the duration of the F1, F2, and F3 transition from the steady state to the /ai/ vowel (referred to as the transition manipulation). For the steady-state manipulation, two versions of the stimuli were constructed, the original and a shortened version. In the shortened version of RIGHT, the first 90 ms was removed from F1 and F2, and 95 ms was removed from F3. In the shortened version of LIGHT, the first 65 ms was removed from F1 and F2, and 70 ms was removed from F3. Note that although the shortened stimuli contained a shorter /r-l/ steady-state duration than the parameters specified by Klatt [1980], it seemed to be more "normal" sounding to CE listeners than the original stimuli based on Klatt's parameters. For the transition manipulation, three versions were constructed: the original, a moderately lengthened, and an extremely lengthened version. In the moderately lengthened version of RIGHT and LIGHT, F1, F2, and F3 transitions were lengthened by 30 ms. In the extremely lengthened version of RIGHT and LIGHT, F1, F2, and F3 transitions were lengthened by 60 ms. Altogether, 12 unique stimuli were produced using this procedure (2 words [RIGHT & LIGHT] X 2 steady-state durations [original and shortened] X 3 formant transition duration [unaltered, moderate, and extreme])

CE, Cantonese, and Japanese listeners were recruited from the Carleton University community. Six native speakers from each language group participated. CE listeners were monolingual English speakers whereas the Japanese and Cantonese listeners were bilingual (English plus their native language). All listeners reported no history of a speech or hearing disorder.

The stimuli were presented to listeners in a two-alternative forced choice (2AFC) identification task. Stimuli were presented over Sony MDR-P1 headphones at approximately 75 dB SPL. Listeners were seated in front of a computer monitor that displayed the printed forms of RIGHT and LIGHT. 250 ms later the auditory stimulus was then presented. The listeners were then required to move a cursor on the computer monitor by using a mouse to indicate whether they heard RIGHT or LIGHT. Each of the 12 stimuli was presented 15 times, making a total of 180 trials in the experiment. Stimuli were presented in a random order that varied from listener to listener. In addition to the identification task, each CE listener rated the quality the token on each trial using a 7-point scale, where 1 was "very clear", 4 "clear", and 7 "distorted" after each trial.

3. RESULTS

The CE listeners rated the stimuli as being clear (mean rating = 2.1). Moreover, no reliable differences among the stimuli were observed in the CE rating data. Thus, the stimuli used in the present experiment were consistently clear to CE listeners, regardless of durational manipulation. Analysis of the identification data indicated significant differences among the percentage of words correctly identified by the three groups of listeners, $F(2,180) = 58.02$, $p < .001$. The CE and Cantonese listeners identification performance (99.7, and 97.0%, respectively) exceeded that of the Japanese listeners (81.2%). In addition, there was a significant interaction between stimulus word (RIGHT vs. LIGHT) and steady-state duration (original vs. shortened), $F(1,180) = 9.45$, $p < .002$. This two-way interaction was subsumed, however, by a three-way interaction among group, stimulus word, and steady-state duration, $F(2,180) = 8.73$, $p < .001$, that is shown in Figure 1. The Japanese listeners identified the original steady-state version of /r/ more accurately than the shortened version whereas they identified the shortened version of /l/ more accurately than the original version of /l/. In contrast, the effect of duration and stimulus word was not important for CE and Cantonese listeners. No other statistically reliable effects were obtained.

4. DISCUSSION

The results of the present experiment provide partial support for the duration hypothesis. Although no effect of increased formant transition duration was found, increased steady state duration did seem to enhance performance for Japanese listeners, at least for /r/. Thus, /r-/ steady-state duration seems to be an important acoustic cue for Japanese listeners. However, the difference in the effect of steady-state duration for /r/ and /l/ is somewhat puzzling as most of the literature on the perception of /r/ and /l/ by Japanese listeners treat the two phonemes as equivalent.

One factor that may account for some of the differences between /r/ and /l/ identification performance by the Japanese listeners is stimulus construction. Recall that to produce the shortened version of RIGHT approximately 90 ms was removed from the steady portion of the /r/ segment whereas for LIGHT, only 65 ms was removed from the steady-state portion of /l/. This difference in the length removed was necessary because of the duration of the steady state portion in the original stimuli; if a 90 ms section was removed from LIGHT, it would have interfered with the formant transition in the /l/ segment. However, this difference in stimulus construction does not account for why the shortened version of /l/ was identified more accurately than the original version. Additional research is needed to clarify this result. One option is to simply remove an equivalent 65 ms section from both the /r/ and /l/ segments

Further work is underway to investigate the role of duration in perception of word-final /r-/l/. It will provide complementary information about the acoustic cues used by Japanese and Cantonese listeners in word-initial position. Additional work is also necessary to integrate findings obtained from duration manipulation experiments with the results of experiments in which formant frequencies are manipulated (e.g., Yamada & Tohkura, 1992).

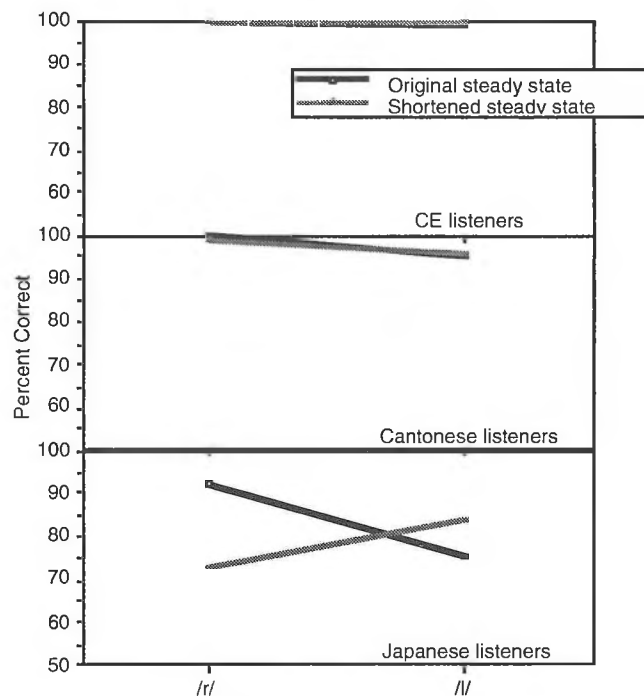


Figure 1. Identification performance (percent correct) for CE, Cantonese, and Japanese listeners presented synthesized RIGHT (/r/) and LIGHT (/l/) stimuli that varied in steady state duration.

5. REFERENCES

- Dissosway-Huff, P., Port, R.F., & Pisoni, D.B. (1982). *Research on speech perception: Progress Report No. 8*. Bloomington, IN: Indiana University.
- Goto, H. (1971). *Neuropsychologia*, *9*, 317-323
- Henly, E., & Sheldon, A. (1986). *Language Learning*, *36*, 505-521.
- Jamieson, D.G., Ramji, K.V., Neary, T.M. & Baxter, T.A. (1990). *Canadian Speech Research Environment (3.0)-User's manual*.
- Klatt, D.H. (1980). *JASA*, *67*, 971-995.
- Lively, S.E., Logan, J.S., & Pisoni, D.B. (1993). *JASA*, *94*, 1242 - 1255.
- Logan, J.S., & BharrathSingh, K. (1994). *Canadian Acoustics*, *22*.
- Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). *JASA*, *89*, 874-886.
- MacKain, K., Best, C.T., & Strange, W. (1981). *Applied Psycholinguistics*, *2*, 369-390.
- Mochizuki, M. (1981). *Journal of Phonetics*, *9*, 283-303.
- Sheldon, A. & Strange, W. (1982). *Applied Psycholinguistics*, *3*, 243-261.
- Yamada, R.A., & Tohkura, Y. (1992). *Speech perception, production and linguistic structure.*, Tokyo: Ohm, pp. 155-174.

6. ACKNOWLEDGMENTS

Work supported by NSERC. Thanks to Linda Polka for providing useful information for developing the stimulus parameters.