

PERCEPTUAL AND PRODUCTIVE DISTINCTION BETWEEN THE ENGLISH [R] AND [l] IN PREVOCALIC POSITION BY ENGLISH AND JAPANESE SPEAKERS.

Elzbieta B. Slawinski

The Psychology Department, The University of Calgary
2500 University Drive, Calgary, Alberta, Canada T2N 1N4

ABSTRACT

The present study examined the usage of spectral and temporal acoustical features in the production of the phonemic contrast between the [r] and [l] sounds in the prevocalic position by Japanese speakers with different perceptual skills in English. The acoustical analysis of recorded tokens demonstrated that Japanese speakers, who perceptually were able to integrate information carried by the spectral and temporal acoustical cues, were also able to use both of those cues in their productive distinction between the [r] and [l] sounds. A lack of ability to integrate the temporal and spectral cues of the perceptual task demonstrated by the second group of Japanese listeners was reflected in their poor productive distinction between the [r] and [l] sounds.

speakers were able to identify [r] and [l] reliably, and were able to treat the temporal and spectral cues as being perceptually equivalent.

While perceptual abilities of Japanese speakers with different exposure to English have been extensively studied by different researchers [2][3][6], none of these studies has addressed the issue of how ability to perceive spectral and temporal cues affects production of the [r-l] phonemic contrast. Thus, the present study examined the spectral and temporal features of the [r] and [l] sounds in the prevocalic position produced by Japanese speakers with different perceptual skills.

II. EXPERIMENT 1 - PERCEPTION

The goal of the experiment 1 was to examine Japanese and English speakers' ability to discriminate [r] and [l] sounds.

2.1. Method

2.1.1. Subjects. Five female monolingual native speakers of Canadian English, five female native speakers of Japanese with early exposure to English (five years of age), and five female native speakers of Japanese with late exposure to English (at twelve years of age or later) served as subjects. Speakers ranged in age from 20 to 50 years. Japanese speakers with early exposure to English were residing in Canada for 20 to 42 years. Japanese speakers with late exposure to English were resident in Canada for the past five to fifteen years. They had a good knowledge of grammar and vocabulary, but almost no conversational experience. All subjects had normal hearing (10 dB HL or better in a range of 250 Hz to 6000 Hz).

2.1.2. Stimuli. Two synthetic series of nine stimuli each, were generated using parallel/cascade synthesizer KLSYN88a. These series were interpolated in the same steps on the spectral dimension of F2 and F3 onset frequency from "rake" to "lake", but differed on the temporal dimension: one series "r-cue" carrying a temporal pattern typical for the [r] sound, and the second series "l-cue" with a temporal pattern typical for the [l] sound. Out

I. INTRODUCTION

A great number of studies have demonstrated that multiple acoustic cues contribute to the perceptual and productive distinction between phonemic contrasts. Thus, differences in spectral and temporal acoustical cues between [r] and [l], allow English speakers to discriminate this phonemic contrast. The [r] is characterized by the low F3 onset frequency, close to F2. In contrast, the [l] is characterized by the high F3 onset frequency [5]. Moreover, [r] is characterized by a longer duration of F1 frequency transition than that of the [l] [1][6]. The ability to integrate information, carried by both spectral and temporal cues, enables English speakers to discriminate efficiently between [r] and [l] sounds [4].

Because the contrast between [r] and [l] in prevocalic position does not exist in Japanese, Japanese speakers, unlike native English speakers, do not perceive a synthesized [r-l] continuum categorically, and as well they do not make a distinction between those sounds productively [7]. However, it has been shown [6], that after intensive training Japanese

of these series, the oddity discrimination tests were prepared [6]. All pairs were three steps apart on the spectral dimension. Four types of stimulus comparisons were prepared; a) one cue spectral-"l cue" (varying along the spectral dimension with fixed 'l'-temporal pattern), b) one cue spectral-"r cue" (varying along spectral dimension with fixed 'r'-temporal pattern), c) two-cue facilitating (changes in the temporal dimension giving enhanced phonetic discrimination), d) two-cue conflicting (changes in the temporal dimension giving suppressed phonetic discrimination).

2.1.3. Procedure. Subjects were tested individually on all four oddity discrimination tasks presented in the form of a computer game. Stimuli were presented through loudspeakers at approximately 70 dBA.

2.2. Results

The pooled discrimination functions for the English speakers almost replicated the findings of Underbakke and Polka (1988), as performances in four oddity discrimination tasks were ordered: two-cue facilitating > one cue 'r' = one cue 'l' > two-cue conflicting. This ordering of performances reflects the perceptual equivalence of spectral and temporal cues. Performance of Japanese speakers with long and/or intensive exposure to English was similar to that of English speakers. On the other hand, Japanese speakers without intensive exposure to English did not perceive temporal and spectral cues equally. They discriminated two-cue facilitating comparison pairs better than either one-cue comparison pairs or two-cue conflicting comparison pairs. Results of this experiment served as a basis for dividing Japanese subjects into two groups differing in the ability to integrate acoustical cues.

III. EXPERIMENT 2: PRODUCTION

The goal of the experiment 2 was to examine the use of both spectral and temporal cues in production of [r-l] phonemic contrast by speakers having different perceptual skills in English.

3.1. Method

3.1.1. Subjects. The same five monolingual native speakers of Canadian English, five native speakers of Japanese with good perceptual performance, and five native speakers of Japanese with poor perceptual performance served as subjects.

3.1.2. Procedure. During a production test, each subject was asked to produce three times the words with [r] and [l] in prevocalic position ("rake" and "lake"). Recordings were made in

an anechoic room, and speech samples were recorded on tape using a DAT recorder. The recorded speech samples were digitized at a 40 kHz sampling frequency with 16-bit amplitude accuracy. Speech samples were down-sampled to 10 kHz, and the formant frequency trajectories were estimated by an LPC formant tracking method.

3.2. Results

The acoustical analysis of recorded tokens demonstrated that [r] and [l] sounds produced by English speakers and by Japanese speakers with good perception differ in both temporal and spectral cues. Thus, a temporal cue such as the duration of F1 transitions produced by these speakers varied significantly for [l] and [r] sounds. Durations of the F1 transitions produced by Japanese with poor perception did not differ between [r] and [l] sounds (Figure 1).

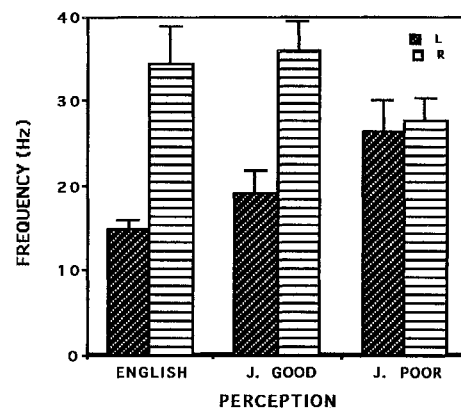


Fig 1. Mean durations of F1 transitions of [r] and [l] sounds produced by English speakers and Japanese speakers with good and poor perception of English. Standard errors are indicated by error bars.

Spectral cue such as the onset frequency of F2 transitions was significantly different in [r] and [l] tokens produced by the English and the Japanese with good perception speakers. On the other hand, Japanese with poor perception produced F2 at the same onset frequency for both these sounds (Figure 2).

All three groups displayed a significant difference in onset frequency of F3 transitions of [r] and [l] sounds. However, the absolute difference between onset frequencies of [r] and [l] was significantly smaller for Japanese speakers with poor perception than that of the two other groups (Figure 3). In addition, the F3 onset frequency of the [r] produced by both Japanese groups occurred at higher frequency than that of English speakers.

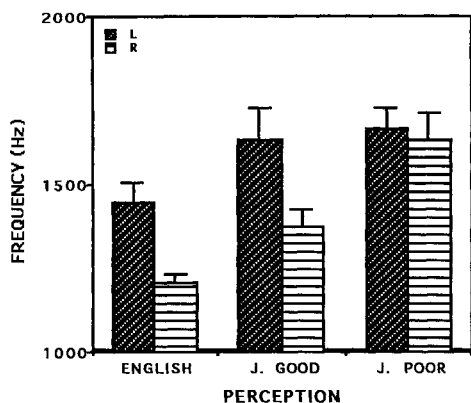


Fig 2. Mean frequency of F2 transition onsets of [r] and [l] sounds produced by English speakers and Japanese speakers with good and poor perception of English. Standard errors are indicated by error bars.

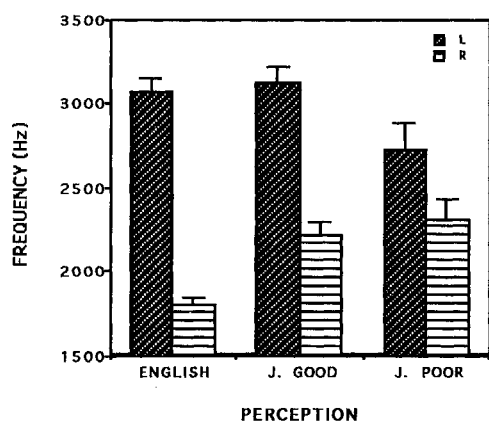


Fig 3. Mean frequency of F3 transition onset of [r] and [l] sounds produced by English speakers and Japanese speakers with good and poor perception of English. Standard errors are indicated by error bars.

Spectral cues such as F2 and F3 onset frequencies differed for [r], [l], and [w] produced by English speakers. Points corresponding to these onset frequencies are clearly separated for those three sounds (Figure 4).

F3 and F2 onset frequencies for [r], [l], and [w] produced by Japanese with poor perception are displayed on Figure 5.

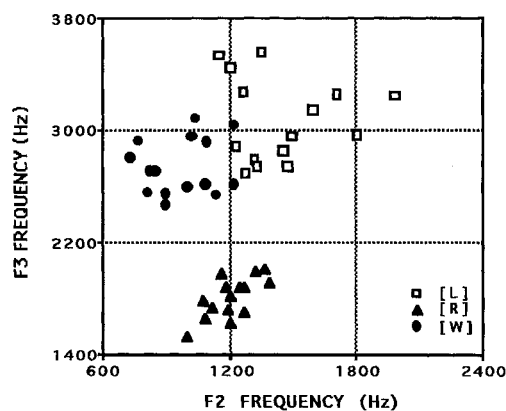


Fig 4. F3-F2 plane for [r], [l], and [w] sounds produced by English speakers.

Points corresponding to F3 and F2 onset frequencies of [w] for these speakers are clustered in a manner similar to those of English speakers. However, points corresponding to onset frequencies of [r] and [l] produced by this Japanese group are intermixed and are dispersed within a region corresponding to regions of both [r] and [l] sounds produced by English subjects.

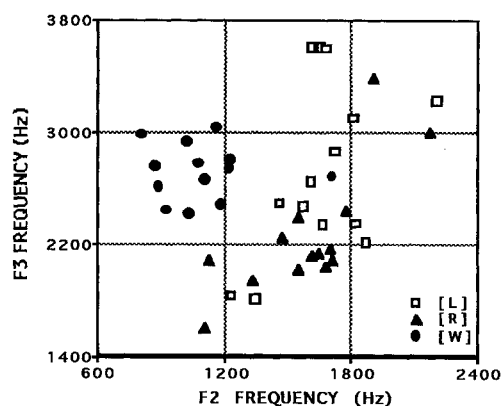


Fig 5. F3-F2 plane for [r], [l], and [w] sounds produced by Japanese speakers with poor perception.

Japanese speakers with good perception produced [l] and [w] sounds with F3 and F2 onset frequencies similar to those produced by English speakers. However, points corresponding to F3 and F2 onset frequencies of [r] are more spread out than those for English speakers (Figure 6).

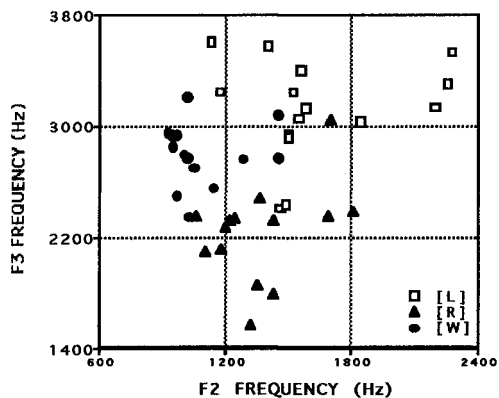


Fig 6. F3-F2 plane for [r], [l], and [w] sounds produced by Japanese speakers with good perception.

Thus, acoustical analyses of [r] and [l] in the initial position produced by English speakers indicated that these sounds are separated because of different values of F3 onset frequency and F1 transition duration. On the other hand, Japanese with poor perception do not produce [r] and [l] sounds with distinct values of F3 onset frequency and duration of F1 transition, and therefore a boundary between clusters corresponding to these sounds cannot be drawn on the F3 versus F1 transition duration plane (Figure 7).

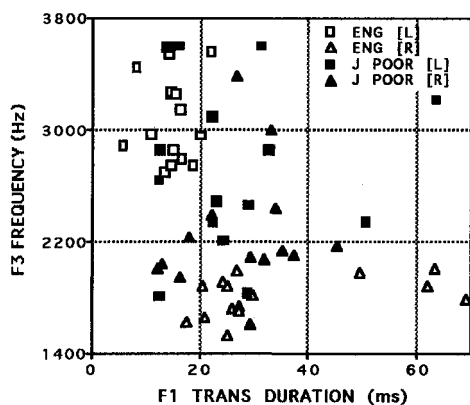


Fig 7. F3 onset frequency versus F1 transition duration for [r] and [l] sounds produced by English speakers and Japanese speakers with poor perception.

IV. DISCUSSION

The results obtained in this study demonstrated that in order to produce phonemic contrast it is necessary to distinguish phonemes perceptually. The

prevocalic [r] and [l] sounds produced by English speakers, who integrate both temporal and spectral cues during perception of the phonetic identity of these sounds, carry information about both of these acoustical cues. Furthermore, Japanese listeners, who perceptually were able to integrate information carried by both of these acoustical cues, were also able to use both of those cues in their productive distinction between the [r] and [l]. Their perceptual and productive performances were not significantly different from those of Canadian English subjects and were significantly different from those of the second group of Japanese listeners. A lack of ability to integrate the temporal and spectral cues in the perceptual task demonstrated by the second group of Japanese listeners was reflected in their poor productive distinction between the [r] and [l] sounds: the duration of F1 transition and the F3 onset frequency did not differ in the produced [r] and [l] sounds. The difference in production demonstrated by two Japanese groups, is probably due to the difference in processing information for phonemic identity, which in turn depends on an exposure to English. However, an ability to integrate the spectral and temporal acoustical cues is a function of a phonemic process that can be modified during the acquisition of a second language.

REFERENCES

- [1] R.M.Dalston, "Acoustical characteristics of English /w,r,l/ spoken correctly by young children and adults." J. Acoust. Soc. Am. 57, pp. 462-469, 1975.
- [2] J.S.Logan, S.E.Lively, and D.B.Pisoni, "Training Japanese listeners to identify English /r/ and /l/: A first report." J. Acoust. Soc. Am. 89, pp. 874-886, 1990.
- [3] K.Miyawaki, W.Strange, R.R.Verbrugge, A.M.Liberman, J.J.Jenkins, and O.Fujimura, "An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English." Perception and Psychophysics, 18, pp. 331-340, 1975.
- [4] L.Polka, and W.Strange, "Perceptual equivalence of acoustic cues that differentiate /r/ and /l/." J. Acoust. Soc. Am. 78, pp. 1187-1197, 1985.
- [5] J.D.O'Connor, L.J.Gerstman, A.M.Liberman, P.C.Delattre, and F.S.Cooper, "Acoustic cues for the perception of initial /w,r,l/ in English." Word, 13, pp. 25-43, 1957.
- [6] M.Underbakke, L.Polka, J.Gottfried and W.Strange, "Trading relations in the perception of /r/-/l/ by Japanese learners of English." J. Acoust. Soc. Am. 84, pp. 90-100, 1988.
- [7] R.A.Yamada, and Y.Tohkura, "Perception and production of syllable-initial English /r/ and /l/ by native speakers of Japanese." Proc. ICSLP'90, pp. 757-760, 1990.