

# SPEECH PERCEPTION AND SPEECH PRODUCTION

## Perception and Production of Syllable-Initial English [r] and [l] by English and Japanese Speakers

by

Elzbieta B. Slawinski

The Psychology Department, The University of Calgary

### Introduction.

Various studies have investigated a contribution of multiple acoustic cues to the perceptual distinction of [r-l] phonetic contrast in English (e.g., Dalston, 1975; Underbakke and Polka, 1988). The results of these studies showed that both spectral and temporal properties facilitate a distinction between [r] and [l] sounds. The onset frequency and transition of F3 in relation to F2 is a primary spectral difference needed for differentiation of [r] and [l] by English listeners. For [r], the F3 frequency at onset is low and therefore close to the F2 onset, while in the case of [l], F3 onset is high relative to F2 onset. Moreover, a short F1 transition duration is present at [l] sound and the long F1 transition corresponds to [r] sound in an initial syllable position. Thus, the presence of both spectral and temporal acoustic cues is important for [r] and [l] distinction in prevocalic position. Underbakke and Polka (1988) demonstrated that a trading relation exists between these cues. Thus, in order to enhance the perception of [r] and [l], the perceptual effects of changing one acoustic cue could be offset by changing the other cue in the opposing direction. The trading relation between temporal and spectral cues for the [r] and [l] contrast depends on language-universal phonetic processing constraints, and may be modified in second language acquisition.

As the Japanese language does not have a contrast between [r] and [l] in prevocalic position, these sounds are very difficult to be discriminated, both perceptually and productively, by Japanese adults. Japanese speakers, unlike native English speakers, do not perceive a synthesized [r-l] continuum categorically, and they do not make a distinction between those two sounds productively (Yamada and Tohkura, 1990). This study investigated how native speakers of Japanese, who are living in Canada for many years, perceive and produce Canadian English.

### Method.

**1. Subjects.** Ten native female speakers of Canadian English (age 20-35 years), and ten Japanese female speakers (age 20-53 years), who were residing in Canada, served as subjects. Japanese subjects could be divided into two categories: 2 females who started to acquire English in Canada at an age of 5 years; 8 females,

whose first contact with English was in Japanese school at age of 12 years.

**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 [r] sound, and the second series "l-cue" with a temporal pattern typical for [l] sound. Out of these series, the oddity discrimination tests were prepared (Underbakke and Polka, 1988). In each test six repetitions of six stimulus pairs were presented in triads; two stimuli were the same and one was different. 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 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 temporal dimension enhanced phonetic discrimination), d) two-cue conflicting (changes in temporal dimension suppressed phonetic discrimination).

**3. Procedure.** Subjects were tested individually on all four oddity discrimination tasks presented in a form of the computer game. Stimuli were presented via loudspeakers at approximately 70 dBA. During a production test, each subject was asked to produce three times the words "rake" and "lake". Recordings were made in an anechoic room, and speech samples were recorded on tape using a microphone B&K 4165, and DAT recorder, SONY DAT-75ES. 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.

### Results and discussion.

The pooled discrimination functions for the English speakers, presented on Figure 1, almost replicate the findings of Underbakke and Polka (1988), as performances in four oddity discrimination tasks are ordered: two-cue facilitating > one cue 'r' = one cue 'l' > two-cue conflicting. Such order of performances reflects the perceptual equivalence of spectral and temporal cues.

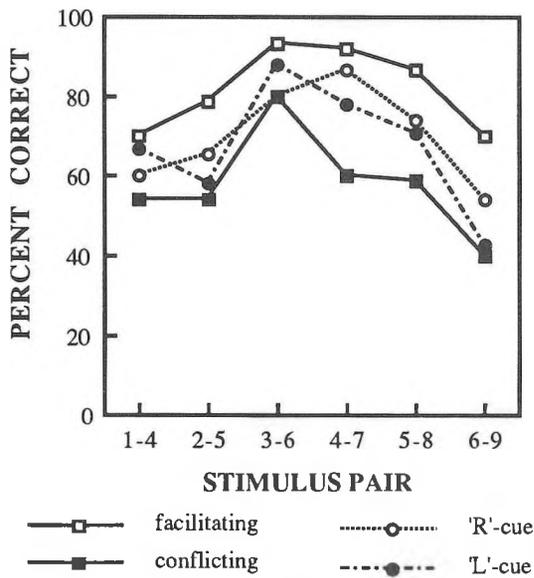


Figure 1.

A different picture emerges, however, from analyzes of the discrimination performance by Japanese, who were exposed to English being over 12 years of age. Their overall performance, for all conditions, was poorer than the English speakers (see Figure 2). Thus, these results might suggest that Japanese speakers did not integrate the spectral and temporal cues. Production results for these two groups of subjects (Figure 3) reflect their perceptual performances. English speakers use both cues, the spectral cue (difference between onset frequencies of F3 and F2), and the temporal cue (duration of F1 transition) in order to distinguish productively [r] and [l] sounds. Figure 3 shows that English [l] sounds (filled squares) are separated from an area of [r] sounds (filled circles).

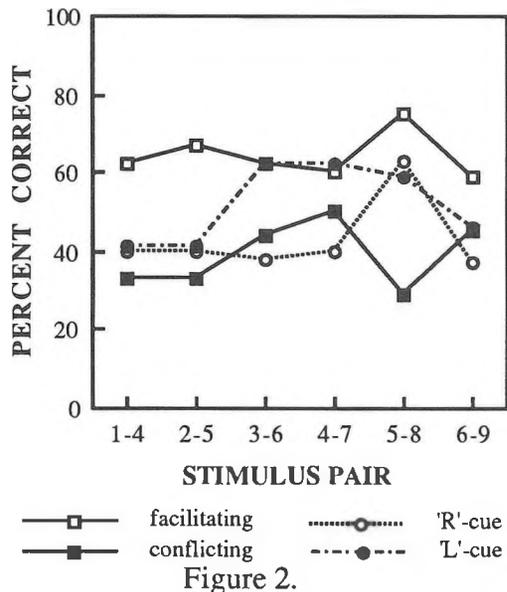
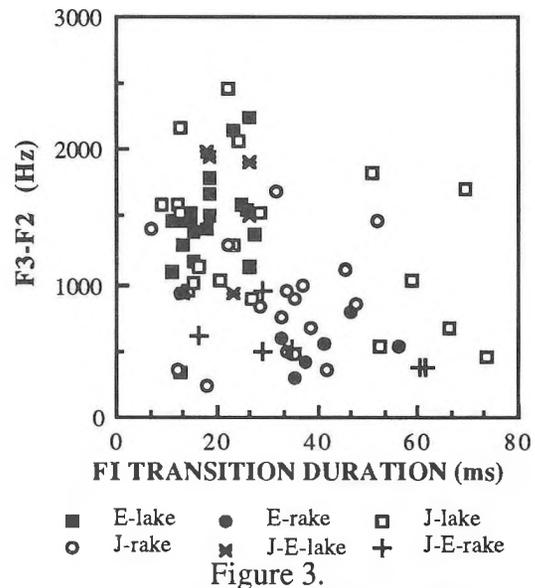


Figure 2.



Japanese speakers, unlike English speakers, use the spectral differences only to a certain extent. The [l] sounds (open squares) produced by Japanese are characterized by slightly larger difference between F3 and F2 formant frequency onsets, than that for the [r] sounds (open circles). However, Japanese speakers do not use the temporal cue. As shown on Figure 3, the areas containing [l] and [r] sounds overlap each other along the 'Transition duration' axis. Performance of Japanese speakers, with early exposure to English (marked on Figure 3 by "\*" -[l] sounds, and "+"-[r] sounds), resembles that of English speakers.

The above results indicate, that in an acquisition of phonemes in a second language the perception and the production are strongly related. Secondly, the acquisition of the [r] and [l] phonemic contrast by Japanese after 12 years of age is incomplete, unless a special training is applied.

#### References.

- Dalston, R.M. (1975). Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults. *J. Acoust. Soc. Am.* 57, 462-469.
- Yamada, R.A., and Tohkura, Y. (1990). Perception and production of syllable-initial English /r/ and /l/ by native speakers of Japanese. *Proc. ICSLP' 90*. pp.757-760.
- Underbakke, M., and Polka, L. (1988). Trading relations in the perception of /r/-/l/ by Japanese learners of English. *J. Acoust. Soc. Am.* 84, 90-99.

#### Acknowledgment.

This work was supported by Toronto Hospital for Sick Children.