Acquisition of /r-l/phonemic contrast by Japanese children and adults

ELZBIETA B. SLAWINSKI
Psychology Department. The University of Calgary. (Canada)

I. INTRODUCTION

English native listeners and speakers use differences in both spectral and temporal acoustical cues to identify with a high efficacy members of the /r-l/ phonemic contrast. Specifically, the difference in the frequency of the onset of third formant (F3) between /r/ and /l/ is used in identification of members of this contrast. For /r/ the onset of F3 is close to the second formant (F2) and has a rising transition. For /l/ the onset of F3 is high relative to F2 and falls slightly towards the F3 of the following vowel (O'Connor, Gerstman, Liberman, Delattre, and Cooper, 1957). In addition to the spectral cue, the duration of the frequency transition of the first formant (F1) is used by English listeners in differentiation of prevocalic /r/ and /l/. The frequency transition of F1 in the /r/ is longer than that of the /l/'s frequency transition (Underbakke, Polka, Gottfried and Strange, 1988).

Japanese speakers, whose native language lacks the /r-l/ contrast in the initial prevocalic position, demonstrate perceptual and productive difficulties in distinction between those two phonemes (e.g., MacKain, Best, & Strange, 1981). Underbakke et al. (1988) compared the usage of acoustical cues in perceptual distinction of /r/ and /l/ phonemes by native English and Japanese adult listeners with different levels of exposure to English. Researchers found that an accuracy of discrimination of the /r/ and /l/ by English listeners and Japanese listeners, skilled in English, depended on a relation between spectral and temporal cues. On the other hand, Japanese listeners with a limited exposure to English were neither able to reliably identify /r/ and /l/, nor to equivalently perceive temporal and spectral cues. There was a distinct separation between performances of the two groups of Japanese listeners in terms of their perceptual abilities due to the difference in processing information for phonemic identity, which in turn might depend on duration and intensity of exposure to spoken English (Lively, Logan, & Pisoni, 1993).

The perceptual abilities of Japanese adult speakers with different degrees of exposure to English have been extensively studied (e.g., Lively, Logan, & Pisoni, 1993; Underbakke et al., 1988; Yamada & Tohkura, 1992). However, only a few studies have addressed spoken proficiency in the discrimination of the /r-l/ phonemic contrast (e.g., Slawinski & MacNeil, 1994). These studies demonstrate that spoken proficiency in a second language is a function of the age at which a first conversational contact with native speakers occurred. Ability to acquire the /r-l/ contrast decreases between 5 and 14 years of age (e.g., Yamada & Tohkura, 1992) and there is a positive correlation between productive and perceptual abilities in a distinction of /r/ and /l/ phonemes (Yamada, Strange, Magnuson, Pruitt, & Clarke III, 1994). None of the above studies has effectively addressed the issue of the relationship between the perceptual ability to integrate information of both spectral and temporal cues, and the usage of those cues in the productive distinction between the English /r/ and /l/ by 3 to 7 years old Japanese children and by Japanese adults.

The reported study included 2 experiments: perceptual and productive. Because the focus of perceptual experiment was to ascertain differences in the perceptual abilities of children of different ages and learners with different exposure to English, it was of interest to assess the productive capabilities of these individuals. To this end a second experiment of the study was conducted to determine if the same pattern of differences, observed in perception task, revealed itself in a production task. Thus, the goal of the study was to explore a relation between the perceptual ability to integrate spectral and temporal cues and a presence of these cues in spoken /r/ and /l/ phonemes.

II. METHOD

2.1. Participants

Four groups of normally developing Japanese children and 3 groups of adults participated in the study. Each group consisted of 5 individuals with a varying number of males and females. The age range in the 3-year old group was 3:00 years (years:months) to 3:11 years (mean age: 3:4 years); for 4-year olds: 4:00 years to 4:5 years (mean age: 4:2 years); for 5-year olds: 4:9 years to 6:00 years (mean age: 5:5 years); for 7-year olds: 6:2 years to 7:10 years (mean age: 7:2 years). All children learned Japanese as their first language and their exposure to English and its intensity increased with age, Moreover, 5 Japanese adults with early and intensive exposure to English (mean age: 30 years), 5 Japanese adults with late and limited exposure to English (mean age: 35 years), and 5 adult native speakers of Canadian English (mean age: 32 years), participated in this study. All subjects had normal peripheral hearing (10 dB HL or better for 250 Hz to 6000 Hz) and no known articulatory problems.

2.2. Stimuli

Two continua of nine stimuli were generated on a KLSYN88a parallel/cascade synthesizer. Each continuum varied spectrally from F2 and F3 onsets' characteristics of 'rake' to those of 'lake', but had different temporal patterns. One continuum, the 'r-cue' had a temporal pattern typical for the /r/ sound; the second continuum, the 'l-cue', had a temporal pattern typical for an /l/ sound. Out of these continua, a perceptual oddity discrimination task was prepared. Four types of conditions, that varied on stimuli comparison pairs, were created: a. 'l-cue' (varying along the spectral dimension with a fixed 'l-temporal' pattern); b. 'r-cue' (varying along the spectral dimension with a fixed 'r-temporal' pattern); c. two cue facilitating (temporal cue supported spectral cue); and d. two cue conflicting (temporal cue conflicted with the spectral cue). All stimuli were three steps apart on the spectral dimension. Each condition consisted of six repetitions of six stimulus pairs presented in triads with two stimuli the same and one different. In addition, 2 pictures, that triggered production of words 'rake' and 'lake' were presented to subjects.

2.3. Procedure

The perceptual task was presented through loudspeakers at approximately 70 dBA via a MAC IIci computer in the form of a game. All responses were collected by the computer. A

short training session to familiarize subjects with the requirements of the test was administered, which consisted of four triads composed of the endpoint stimuli of a particular condition.

Six repetitions of each word ('rake' and 'lake') were recorded on tape in an anechoic room, using the microphone B&K 4165 and a 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 20 kHz, and the formant frequency trajectories were estimated by an LPC formant tracking method.

III. RESULTS AND DISCUSSION

The present study attempted to provide information about the acquisition of perceptual and productive use of acoustical cues in the distinction of English prevocalic /r/ and /l/ by Japanese children and adults who acquired English as a second language.

3.1. Experiment 1: Perception

Individual participants' data were transformed into percentage correct scores for each condition. Results from children and adults were analyzed separately. A repeated measures analysis of variance (ANOVA) with one between and 2 within variables (4 age groups X (4 conditions X 6 comparison pairs)) was performed on children data. There were significant main effects for age ($F(3,19)=5.81$, $p<.005$), condition ($F(3,57)=10.14$, $p<.0001$), and for comparison pairs ($F(5,95)=6.88$, $p<.0001$). The interactions between the age group and comparison pairs, and between condition and comparison pairs were also significant ($F(5,95)=2.09$, $p<.01$; $F(15,285)=2.17$, $p<.01$, respectively).

Simple main effects indicated that the 3 younger age groups did not significantly differ in their performance, however, their performance was significantly different from that of 7-year old ($F(1,9)=9.67$, $p<.01$; $F(1,9)=10.73$, $p<.01$; $F(1,9)=5.64$, $p<.05$, for 3-, 4-, and 5-year olds, respectively). Separate univariate ANOVAs with the age group as a between group variable revealed that the performances of four age groups were significantly different for the two cues facilitating condition ($F(3,19)=3.77$, $p<.05$; Fig. 1) and two cues conflicting condition ($F(3,19)=4.62$, $p<.01$; Fig. 1). Tukey's HSD showed that the overall performance of 7-year olds was significantly better on all conditions relative to the other 3 age groups.

The ordering of the functions for the 7-year old was: two cues facilitating >' r-cue'>l-cue'>two cues conflicting. The differences between these functions were found to be significant ($F(3,15)=8.94$, $p<.001$). Moreover, it was found that the main effect for comparison pairs was significant ($F(5,25)=11.52$, $p<.0001$), as well as the interaction between conditions and comparison pairs ($F(15,75)=2.86$, $p<.001$). Japanese children of 3-, 4-, and 5-years of age showed a very different pattern of performance. There were no significant main effects for either stimulus condition or stimulus pairs; nor were any of the interactions significant. These groups of children did not perceive the temporal and spectral aspects of the phonemes equally and they used probably acoustical differences between these stimuli in order to discriminate them.

The performance of the Japanese adult listeners with early and intensive exposure to English was similar to that for native English adult listeners (Fig. 2). These two groups demonstrated an

Fig.1. Difference in perceptual discrimination between /r/ and /l/ by 4 groups of children on the two cues facilitating condition (solid line) and on the two cues conflicting condition (dashed line).

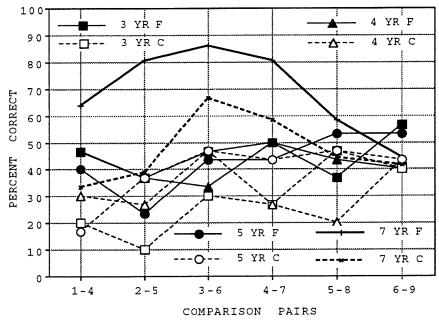
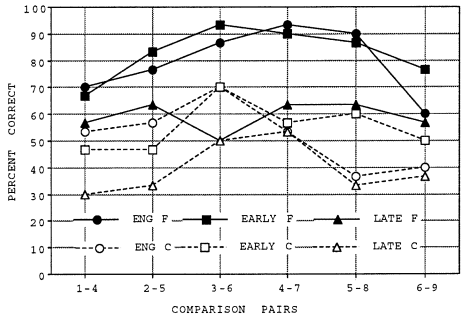


Figure 2. Difference in perceptual discrimination between /r/ and /l/ by three groups of adults on the two cues facilitating condition (solid line) and on the two cues conflicting condition (dashed line).



identical ordering of the discrimination functions: two cues facilitating > 'r-cue' = 'l-cue' > two cues conflicting. Repeated measures ANOVA showed that there were significant differences across the four conditions for English ($F(3,12)=12.7$, $p<.0001$) and for Japanese listeners ($F(3,12)=17.19$, $p<.0001$). The main effect for comparison pairs was also significant for English ($F(5,20)=9.771$, $p<.0001$) and for Japanese listeners ($F(5,20)=6.41$, $p<.001$).

Japanese listeners with late and limited exposure to English showed a very different pattern of performance when compared to Japanese listeners with early and intensive exposure (Fig. 2). There were no significant main effects for either stimulus conditions or stimulus pairs; nor were any of the interactions significant. Thus, this group of listeners, similarly to the 3-, 4-, and 5-year old children, did not perceive the temporal and spectral aspects of the signals equally and could not trade these cues one for another.

According to Repp (1983), the observed ordering of discrimination accuracy on comparison pairs in a vicinity of the phonemic boundary (better for comparison pairs near the phonemic boundary than for comparison pairs which included stimuli from the same category) by 7-years old children, English adults and Japanese adults with early and intensive exposure to English, corresponds to a perceptual equivalence of the temporal and spectral cues and it suggests that these listeners relied on phonemic similarity while performing the discrimination task. Moreover, it suggests that the relation between those cues is a result of a language-specific phonemic categorization process. The observation that the younger children (3-, 4-, and 5-years old) and Japanese adults with late and limited exposure to English were not able to distinguish between /r/ and /l/, even in the two cues facilitating condition, suggested that they relied more on acoustic dissimilarities between stimuli than on phonemic similarities, as they were showing similar discrimination between the middle range of comparison pairs and between those closer to phonemic templates.

The results of the experiment 1 suggested that an acquisition of the perceptual discrimination of a phonemic contrast (at least /r-l/) in the second language depends on the intensity of exposure to this language, as longer (earlier) and more intensive exposure to a second language results in an improved perceptual ability relative to a shorter (later) and limited exposure. Thus, results of this study supported previous findings of Underbakke et al., (1988), and Yamada and

Tohkura (1992). Moreover, young children (3-years old), who used Japanese language as a mean of communication, did not discriminate English /r-l/ contrast. However, 7-year olds, whose spoken language in school was English, were very proficient in the perceptual distinction between /r/ and /l/.

In conclusion, Japanese children and Japanese adults with different degrees of exposure to English did not rely in the same manner on available acoustical cues needed for English prevocalic /r/ and /l/ discrimination. While discrimination by older Japanese children and adults with early and intensive exposure was strongly based on similarity or dissimilarity between a percept derived from the integrated information of spectral and temporal cues, the discrimination of Japanese younger children and adults with late and limited exposure to English relied to a lesser degree on the perceptual configuration of acoustical cues and was probably more influenced by acoustical dissimilarities.

3.2. Experiment 2: Production

In the second part of the study we investigated the productive use of temporal and spectral cues in the discrimination of the /r-l/ phonemic contrast by Japanese children, Japanese adults with different exposure to English, and English adults. Twelve words were analyzed per subject (six repetitions of 'lake' and 'rake') using Signalyze software (InfoSignal Ltd.). In each age group of Japanese children and adults' groups 60 words were analyzed (5 subjects x 6 repetitions x 2 consonants). A waveform editing program was used to determine the following characteristics of each word: (1) duration of F1 transition; (2) onset frequency of F2 formant transition; and (3) onset frequency of F3 formant transition. The onset of the transition was measured from a point of the apparent change in frequency. The offset of the transition was taken to be the point at which the steady state of the following vowel had started. A 15-ms half-Hamming window was used for all samples and samples were taken at pitch intervals.

Because a goal of this study was to explore the productive acquisition of the /r-l/ contrast in the second language, multiple univariate analyses of variance (ANOVA) with a type of group as a between-group variable were performed on the duration of F1 transition, and on the difference between frequency onsets of F3 and F2 formants for two examined consonants (/r/ and /l/). Follow-up analyses of the main effects with Tukey's HSD post hoc pairwise comparison tests were used to examine relevant results. Significant results at the $p<0.05$ level were reported. For all follow up tests Bonferroni adjustments were used to control the type 1 error rate.

Only the F1 transition duration for the /l/ was found to be significantly different between groups. The obtained difference was due to a long F1 transition produced by Japanese whose exposure to English started late and was limited. Durations of F1 transition for /l/, produced by members of all other groups, did not differ significantly, however, they were significantly shorter compared to that of the late learners ($F(6,35) = 4.20$, $p<.005$). Durations of the F1 transition for /l/, produced by all groups, except the late learners of English, were shorter compared to that for /r/. Furthermore, the difference between durations of F1 transitions in the /r/ and /l/ contexts was smaller for younger children than that for older children and adults (Fig. 3).

Univariate analysis of variance with group's type as a between group variable on a difference between onset frequency of F3 and F2 formants in /r/ context, revealed that this depen-

Figure 3. Durations of F1 transition in the /l/ and /r/ contexts produced by 3-, 4-, 5-, and 7-years old Japanese children and Japanese adults with different exposure to English. There are also shown the durations of F1 transitions for /r/ and /l/ produced by native English speakers.

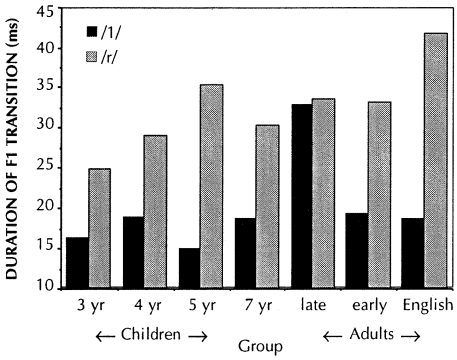
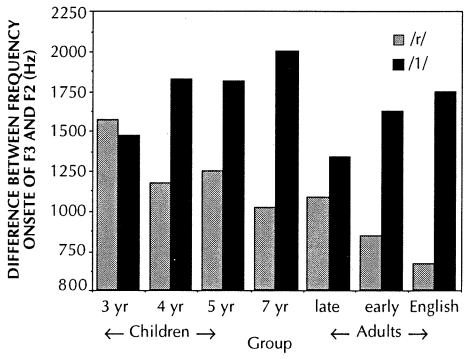


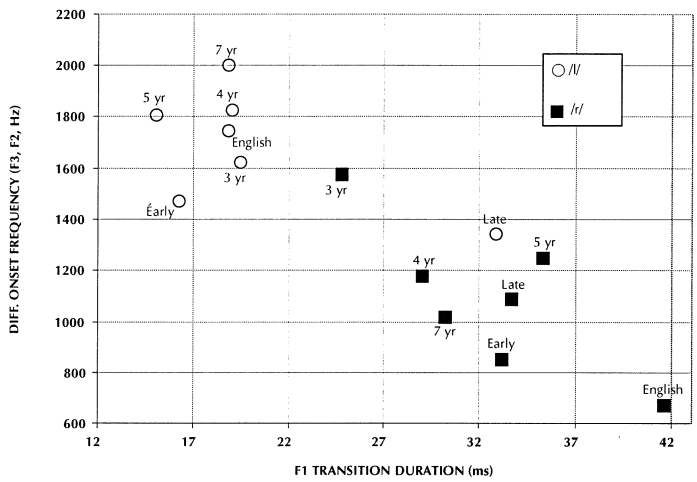
Figure 4. The difference between onset frequencies of F3 and F2 formants in the /l/ and /r/ contexts produced by 3-, 4-, 5-, and 7-years old children and adults with different exposure to English. There is also shown the difference between onset frequencies of F3 and F2 formants for /r/ and /l/ for native English speakers.



dent variable was significantly different between groups ($F(6,35)=4.41$, $p<0.001$). Seven years old children, Japanese adults with early and intensive exposure to English and native speakers of English produced significantly smaller difference between onset frequencies of F3 and F2 formants in /r/ context compared to that produced by younger children and Japanese adults with late and limited exposure to English. The difference between onset frequencies of F3 and F2 formants in the /l/ context did not differ significantly among all groups. However, it was observed that the difference between frequency onsets of the F3 and F2 formants increased as a function of age and/or exposure to English for the produced /l/ phoneme, while at the same time this difference decreased with age and/or exposure to English for the /r/ phoneme. In particular, the difference between the 'F3- F2 onset formants' in the /l/ and /r/ contexts was enhanced in a production of 7-year olds and Japanese adults with early and intensive exposure to English (Fig. 4).

Spoken English phonemes /r/ and /l/ differ in their spectral and temporal characteristics: F1 transition for /l/ is shorter than that for /r/ and difference between onset frequency of F3 and F2 formant transitions is greater for /l/ than that for /r/. Results of this study indicated that all groups, but 3-years old children and adults with late and limited exposure to English, produced very distinct /r/ and /l/ phonemes (Fig. 5). The /r/ produced by 3-year olds carried information close to this for /l/, while, adults with late and limited exposure to English produced /l/ that had characteristics closer to these of /r/.

Figure 5. Mean values of spectral and temporal characteristics of /r/ and /l/ produced by 3-, 4-, 5-, and 7-year olds and adults with different exposure to English. There are also shown mean values of native English speakers.



Thus, the high perceptual proficiencies on the oddity discrimination tests demonstrated by 7-year olds and adults with early and intensive exposure to English, were mirrored by a differentiation of /r/ and /l/ phonemes in their production. The perceptual discrimination accuracy of /r/ and /l/ sounds by these groups did not rely on detection and classification of acoustical differences only, but was based on the process of attending to patterns of acoustical cues that differentiate this phonemic contrast (Pisoni, Logan, & Lively, 1991), and subsequently comparing them to the multidimensional representation of phonemic percept. Furthermore, an awareness of relations between spectral and temporal cues in the /r/ and /l/ allowed them to use these cues in the production of /r/ and /l/. On the other hand, the poor perceptual performance of 3-years old children and adults with late and limited exposure to English was associated with a lack of productive differentiation between /r/ and /l/.

Different pattern was observed for groups of 4- and 5-year olds, who could productively differentiate the /r/ and /l/, however, they could not perceptually distinguish these phonemes. According to Repp (1983), perceptual discrimination performance depends on an ability to integrate information about acoustical cues, that is the result of a language-specific phonemic categorization process. An effective category formation is a lengthy process that depends on the amount of environmental variability necessary for generalization (Lively, Logan, & Pisoni, 1993), as well as on the ability to pay attention to a few acoustical cues at the same time. Thus, it is probable that listeners with a limited exposure to variability of English pronunciation did not generate clear percepts of /r/ and /l/. Acoustical analyses of temporal and spectral cues of produced /r/ and /l/ by these children indicated also that these cues had values between those typical for English phonemes. Thus, 4- and 5-years old children produced the /r/ that carried acoustical cues similar to these of their /l/, however,their /r/ was in a range of possible English /r/.

Does perception of phonemic contrast precede a productive distinction of /r/ and /l/, or does productive distinction between these phonemes precede perceptual acquisition of the phonemic contrast? The results of reported study are not conclusive. The results of this study suggest that probably adults have to acquire perceptual proficiency in discriminating /r/ and /l/ in order

to be able to distinguish them productively. With regard to children, in addition to their exposure to English a cognitive development probably plays a role in perceptual and productive acquisition of /r/ and /l/. Thus, from the results discussed here and elsewhere, apparently there exists a strong relationship between the perceptual ability to process information about phonemic percept, and its productive execution.

REFERENCES

Lively, S.; Logan, J. S.; Pisoni, D. B. 1993. Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *J. Acoust. Soc. Am. 94*, 1242-1255.

MacKain, K. S.; Best, C.T.; Strange, W. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals, *Applied Psycholinguistics, 2*, 369-390.

O'Connor, J. D.; Gerstman, L. J.; Liberman, A. M.; Delattre, P. C.; Cooper, F. S. 1957. Acoustic cues for the perception of initial /w, r, l/ in English. *Word, 13*, 25-43.

Pisoni, D. B.; Logan, J.; Lively, S. 1991. Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In *Development of speech perception: The transition from recognizing speech sounds to spoken words*, in H. Nusbaum, J. Goodman MIT, Cambridge, MA.

Repp, B. 1983. Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech Communication, 2*, 341-362.

Slawinski, E.; MacNeil, J. F. 1994. The relationship between perceptual and productive discriminations of English [r] and [l] sounds by Japanese speakers. *International Journal of Psycholinguistics, 10*, pp.1-23.

Underbakke, M.; Polka, L.; Gottfried, J.; Strange, W. 1988. Trading relations in the perception of /r/-/l/ by Japanese learners of English. *J. Acoust. Soc. Am. 84*, 90-100.

Yamada, R. A.; Tohkura, Y.1992. The effects of experimental variables on the perception of American English /r, l/ by Japanese listeners. *Perception and Psychophysics, 52*, 376-379.

Yamada, R. A.; Strange, W.; Magnuson, J. S.; Pruitt, J. S.; Clarke, W. D. 1994. The intelligibility of Japanese speakers' production of American English /r/, /l/, and /w/, as evaluated by native speakers of American English. *Proceedings ICSLP' 94*. 2023-2026.