TR－A－0100 30

# The Effect of Experiment Variables on the Perception of American English /r,l/ for Japanese Listeners
## - *Response Categories, Classification of Subjects, and Stimulus Range* -

*Reiko A. YAMADA and Yoh'ichi TOHKURA*

# 1991. 2. 6

# The effects of experiment variables on the perception of American English /r,l/ for Japanese listeners
## - Response categories, classification of subjects, and stimulus range -

Reiko A. Yamada, and Yoh'ichi Tohkura

ATR Auditory and Visual Perception Research Laboratories, Kyoto 619-02, Japan

**ABSTRACT**

The effects of experiment variables associated with the experiment tasks on the identification of American English /r,l/ for native speakers of Japanese are investigated. The three variables studied are: response categories, classification of subjects, and stimulus ranges. Through experiments using synthesized American English /rait-lait/ series and naturally spoken stimuli, all three variables are revealed to affect the experiment results for Japanese listeners. The effects of these variables clarify some of the perceptual characteristics of /r,l/ for Japanese listeners. In the first experiment designed to determine the effects of response categories, it is suggested that Japanese listeners identify some of the stimuli on the /r-l/ continuum as /w/. In the second experiment, differences in perceiving synthesized stimuli are discussed by classifying the Japanese listeners. The results show a positive correlation between the perception of synthesized stimuli and naturally spoken /r,l,w/. Japanese listeners well able to identify naturally spoken /r, l, w/ perceive the synthetic stimulus series categorically but they still have three categories, /r/, /w/ and /l/ on the series. In the third experiment, the stimulus range shows a striking effect on the identification results for Japanese listeners. The /r/ and /l/ perception is strongly affected by the stimulus range, whilge the /w/ perception less so. This indicates that Japanese listeners tend to make a relative judgement between /r/ and /l/. The experiment variables studied in this paper are of great importance when considering modelling the acquisition process of non-native phonemes.

Many precious studies have revealed that phoneme perception is modified by the linguistic environment. However, the process of acquiring phonemes, i.e. the way phoneme perception is actually modified by the linguistic environment, has not been investigated so far. This paper is concerned with the variables associated with experiment designs and tasks which affect the results when studying the acquisition of phonemes in a second language (L2). The process of acquiring phonemes in L2 is much different from that in the first language (L1), because of the difference in age of acquisition and the existence of the first language phonological system. When studying the process of acquiring phonemes in L1, there are serious methodological problems because the ability to perform experimental tasks varies with age. Although some of such methodological problems can be avoided when studying the process of acquiring phonemes in L2, other variables, such as age when learning L2, quantity and quality of learning, etc., still affect the process. However, studying how such variables affect the acquiring process is thought to be efficient for further understanding of the nature of phoneme acquisition.

Studies of how native speakers of Japanese learn the phonemic distinction between American English (AE) /r/ and /l/ sounds provide clues to solving the problems of studying phoneme acquisition, because the influence of the phonemes in L1 on the acquisition of phonemes in L2 can be restricted. That is, in the phonological system of Japanese, the AE /r/ and /l/ contrast is not distinctive. Furthermore, neither AE /r/ nor AE /l/ resemble any Japanese phonemes. Thus, AE /r/ and /l/ contrast is difficult to acquire if the contrast is not learned before a certain age (Cochrane, 1980; Yamada and Tohkura, 1989). Most Japanese speakers have considerable difficulty in acquiring those two sounds even though they start learning English in junior high school at about age 12.

Four phonemes, /w/, /r/, /l/, and /j/, are classified as liquid in AE (O'Connor et.al., 1957; Lisker, 1957). In the Japanese phonemic system, there occurs no phoneme noted as /l/. Although there exists a phoneme noted as /r/ in Japanese (J /r/), this J /r/ is classified as a stop or flap depending on its vowel context. It is much different from AE /r/ in its acoustic features, and is actually often confused perceptually with /t/ or /d/ for AE listeners (Price, 1981). In Price's study, AE listeners identified J /r/ more often as a flap (/t/ or /d/; 72%) than as /r/ (20%) or /l/ (7%). Two other liquids, /w/ and /j/, occur in Japanese. Japanese /w/ and /j/ are not equal but rather similar to AE /w/ and /j/, respectively, and these AE /w/ and /j/ are hardly confused perceptually with other phonemes by Japanese listeners by Japanese listeners than AE /r/ and /l/.

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 (Goto, 1971; Liberman et al. 1973; Miyawaki et al. 1973; McKain et al. 1981; Mochizuki. 1981; Shimizu & Dantsuji. 1983; Strange & Dittmann. 1984). Trading relations between the spectral cue and the temporal cue were also investigated cross-linguistically (Polka et al. 1985,; Underbakke et al. 1988). Furthermore, Mann (1986) demonstrated that Japanese listeners respond to the different acoustic patterns which convey /r/ and /l/. However, there still remain some important issues which have to be studied in detail. When studying the perception of phonemes in a non-native language, results of experiments can be influenced by many variables associated with the tasks and stimuli used in perceptual experiments, even though these variables have little or no effect on the perception of phonemes in the native language. However, the effects of these variables have been often ignored in the previous cross-linguistic studies, despite their importance, especially when studying the acquisition process of non-native phonemes. This paper is concerned with the effects of such variables on the perception of AE /r/ and /l/ for Japanese speakers, focusing on the following viewpoints.

(1) "Don't Japanese listeners perceive other sounds on the /r-l/ synthetic stimulus continuum?" There is a possibility that they perceive some sound besides /r/ and /l/ on the /r-l/ series even when Americans do not. In the previous studies (MacKain et al. 1981; Strange, and Dittmann 1984), two peaks were often observed in the discrimination results for Japanese listeners. These two peaks were inconsistent with the response results of the identification tests, in which listeners were asked to choose one of the alternatives, /r/ or /l/. From Mochizuki's study (1981) using the synthesized /r-l/ series, Japanese listeners identified more stimuli as /w/ than American listeners did when they were asked to choose one of three response categories, /r/, /l/, and /w/. Authors' preliminary investigation, in which listeners were asked to identify the

stimuli without being limited to certain response choices, also revealed Japanese listeners identified more stimuli as /w/ at higher response rates than American listeners did. Therefore, it is reasonable to assume the existence of the /w/ category on an /r-l/ continuum. When we hypothesize that Japanese listeners perceive the three phonemes, /r/, / l/, and /w/ on the /r-l/ continuum, the two peaks observed in the discrimination test results can be interpreted consistently with the identification test results.

(2) "How should individual differences of Japanese listeners be controlled?" When studying the perceptual characteristics of non-native phonemes which have not been acquired completely by the subjects but, on average, rather poorly, it is of great importance to study the details of how, and to what extent these phonemes are acquired. As it is assumed that the subjects differ in their ability to identify non-native phonemes, their perceptual characteristics differ individually to a large extent. The experimental results have to be discussed, not on average, but by taking account of the individual differences among Japanese listeners. Some proper classification of subjects is needed to observe the acquisition process precisely.

(3) "Does the stimulus sets in the experiment session influence perception by Japanese listeners?" There is a possibility that Japanese listeners make a relative judgement among the stimuli presented in the experiment session. If they do, the results might be considerably influenced by the stimulus set, for example, the range of stimulus variation given in the experiment session.

To answer these questions, we performed three types of experiments. Experiment 1 was performed to study issue (1), by comparing the two cases of identification test results with respect to response categories; a forced choice between two categories, /r/ and /l/, and a forced choice among three categories, /r/, /l/, and /w/ for the same stimuli. In Experiment 2 issue (2) was studied by classifying the subjects according to the identification score of naturally-spoken /r/, /l/, and /w/ sounds. Experiment 3 was designed to study the effects of the stimulus set on the identification. The following two identification test results were compared: (1) the identification results when all the stimuli on a synthetic /r-l/ series were presented and (2) the results when those when only the stimuli on a part of the series were presented.

# EXPERIMENT 1

## Purpose

There is a possibility that Japanese listeners have a /w/ category on the AE /r-l/ continuum. If they do, further investigation by considering /w/ category is needed to describe the acquisition of /r/ and /l/. As this issue is thought to be very important for further study of /r/ and / l/ perception for Japanese listeners, the hypothesis that Japanese listeners have the /w/ category on /r-l/ continuum has to be examined first. For this purpose, Experiment 1 was designed to compare the results of the following three tests: (1) an identification test with two response categories, /r/, and /l/, (2) an identification test with three response categories, /r/, /l/, and /w/, and (3) a discrimination test.

## Method

**Stimuli.** A pair of English words, "right" and "light" was selected as a minimal pair for the initial consonant. A synthetic /rait-lait/ ("right-light") continuum was generated by the Klatt cascade formant synthesizer (Klatt; 1980) used in this experiment. The continuum consists of 19 stimuli (ST1 through ST19), and Fig. 1 provides a schematic spectrographic representation of the initial CV portion /rai-lai/ for the stimuli. The acoustic parameters for both ST1 (idealized "right") and ST19 (idealized "light") were derived from the naturally spoken /rait/ and /lait/ uttered by a native male speaker of AE. Temporal patterns of stimulus spectra were represented by first five formant frequencies (F1 through F5) and bandwidths. To extract these formant parameters and to synthesize the stimuli, we used a speech research software known as SPIRE (Zue and Cyphers, 1985; Cyphers, 1985). The parameters thus obtained were modified by repetition of trial listening to synthesized stimuli to provide synthesized /rait/ and /lait/ of clear and high quality.

To construct the stimuli on the /rait-lait/ continuum, three acoustic parameters, F2 and F3 onset frequencies and F1 transition were varied. These parameters are identical to some of the previous studies on /r-l/ perception (e.g. Polka and Strange, 1985). At the beginning of the syllable, the formant frequency was fixed at 83 ms (See Fig.1). This portion is called the initial steady state (ISS) for stimuli. F2 and F3 transition durations were kept constant and the formant frequencies were varied linearly within the transitions. From ST1 to ST19, the F2 and F3 onset frequencies were varied from 920 Hz to 1280 Hz in
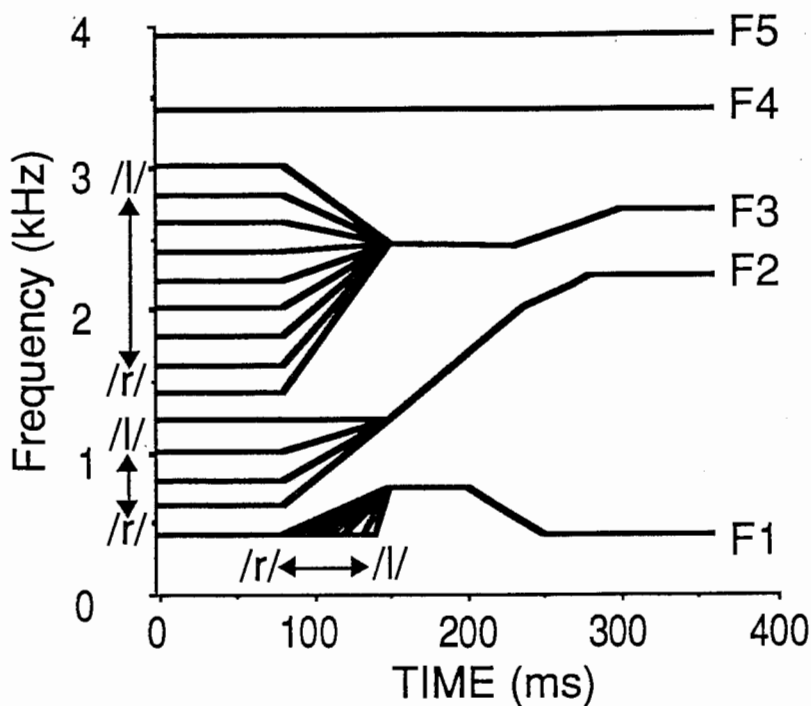
Figure 1   Schematic representation of frequency trajectories of F1 to F5 for the synthesized stimuli. The frequency of F2, that of F3, and temporal pattern of F1 were varied from/r/-characteristics to /l/-characteristic for the initial steady and transitional part of the stimuli.

20 Hz steps and from 1200 Hz to 3000 Hz in 100 Hz steps, respectively. With respect to F1, steady state and transition durations were varied across the continuum, though the onset frequency was fixed at 400 Hz. When F1 transition durations were varied from 67 ms to 13 ms in 3 ms steps, the F1 steady state duration varied from 83 ms (ST1) to 137 ms (ST19) also in 3 ms steps by keeping the total duration from the onset to the end of formant transition constant (i.e., 150 ms). On the other hand, the initial steady states for F2 and F3 were kept at 83 ms across the continuum. The duration of the /rai-lai/ part was 360 ms. The F1, F2 and F3 frequencies at the end of transitions (i.e., onset of the vowel part /ai/) were 750, 1220 and 2465 Hz, respectively. The F4 and F5 frequencies were kept at constant values, 3400 and 3950 Hz all through the stimuli. The fundamental frequency contour used to synthesize the stimuli was designed based upon the natural spoken / rait/ and /lait/, and this same contour was used for all stimuli. The original waveform of /t/ was copied from the /t/ part of the naturally-spoken utterance, /rait/, from which the synthesis parameters were derived, and was added to all the synthetic /rai-lai/ stimuli to make the /rait-lait/ continuum.

In a preliminary experiment, all 19 stimuli were presented to one native AE listener and seven native Japanese listeners. The listeners were instructed not to make a forced choice among some phoneme categories, such as /r/ and /l/, but to report what they heard as the initial consonant of each stimulus. The result showed that all the listeners perceived one of three phonemes, /r/, /l/, or /w/, only.

Throughout the experiments, 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. The experiment included sessions of identification tests and ABX discrimination tests. For identification tests, 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 ten blocks of ten trials to make 100 trials in total. The 100 trials resulted from five randomly-ordered repetitions of each of the 19 stimuli on the /rait-lait/ series with 5 dummy stimuli (ST1,5,10,15, and 19) at the beginning of the session. These dummy trials were intended as familiarization, and not scored. The inter-

trial interval (ITI) was 3 seconds, and the inter-block interval (IBI) was 8 seconds. The block start signal was a beep sound recorded 2 seconds prior to the beginning of each block.

For ABX discrimination tests, several sessions (i.e., with different randomizations of stimulus pairs) were recorded on a digital audio tape in triads. The 15 four-step comparison pairs (ST1 vs ST5, ST2 vs ST6, —, ST15 vs ST19) were chosen from the 19 stimuli, and each pair was arranged in triads in the four possible ABX permutations (i.e., ABA, ABB, BAA, and BAB) for a total of 60 triads. In each trial, one of the 60 triads was presented once with an inter-stimulus interval (ISI) of 1 second. Sixty trials were randomly arranged in six blocks of ten trials to make one session with an ITI of 3 seconds and an IBI of 8 seconds. The start signal was a beep sound recorded 2 seconds prior to the beginning of each block.

These stimuli were presented to listeners binaurally over headphones, STAX SR L Professional, in a sound-proof room at a fixed level of about 85 dB SPL at the peak intensity which was a comfortable level for the listeners.

**Subjects.** Five native speakers of Japanese (S1-S5) who had never resided abroad served as listeners. The average age of the listeners was 22, and ranged from 21 to 24. All listeners reported no history of hearing and speaking disorder.

**Procedure.** For identification tasks, two conditions, the RLW condition and the RL condition, were prepared. The RLW condition is a task with a forced choice among three categories, /r/,/l/, and /w/, and the RL condition is one with a forced choice between two categories, /r/ and /l/. Listeners were instructed to identify the word initial consonant, and to make a forced choice among the given categories by checking a corresponding response category on the answer sheet.

In the ABX discrimination tests, listeners were instructed to indicate whether the third item (X) in the triad matched the first one (A) or the second one (B) by checking a corresponding response category, "1" or "2", on the answer sheet.

Listeners participated mainly in 12 sessions on two different days. On each day, the experiment consisted of three sessions of identification tests, followed by another three sessions of ABX discrimination tests. For

three subjects (S1, S2, and S3), the identification tests were RL conditioned on the first day, and were RLW conditioned on the second day. For other two subjects (S4, and S5), the tests were RLW conditioned on the first day and RL conditioned on the second day.

## Results

The left panels in Fig.2 show identification results in the RL condition (top), RLW condition (middle) and ABX discrimination results (bottom), pooled across all five Japanese listeners. Identification performance is shown with response rates for each category, and discrimination performance is shown with accuracy in percent. Fifty percent is the chance level for discrimination.

Compared with the results given in the RL condition, those in the RLW condition were quite different. In the RL condition, where response categories were restricted to /r/ and /l/, the identification results seemed to indicate these five listeners had two distinctive categories, /r/ and /l/, because they identified ST1 to ST8 as /r/, and ST13 to ST19 as /l/ with high consistency. However, in the RLW condition, the frequent responses to /w/ appeared from ST7 to ST12, with a peak at ST9. The discrimination results showed one peak around ST5-ST9 or ST6-ST10. This peak corresponded more to the boundary between /r/ and /w/ in the RLW condition than that between /r/ and /l/ in the RL condition.

The center and right panels in Fig.2 show the individual results for two listeners, S1 (center) and S3 (right), who showed typical results about the /w/ category. Comparing the RLW-conditioned results and the discrimination results, it is clear that S1 had three distinctive categories, /r/, /w/, and /l/, on this synthetic series. S1 seemed to perceive this series categorically as /r/, /w/, and /l/, with high response rates for each category: ST1 to ST6 as /r/, ST9 to ST12 as /w/, and ST15 to ST19 as /l/, and two discrimination peaks, one around ST4-ST8 and another around ST10-ST14. By way of contrast, the results of the RL-conditioned identification seemed to indicate that the other listener, S3, perceived this synthesized continuum categorically as /r/ and /l/. She identified ST1 to ST8 as /r/ and ST10 to ST19 as /l/, with high response rates. The peak in the discrimination results corresponded to these identification results. However, from the results of the RLW-conditioned identification task, it is indicated that she had a distinctive category of /l/, but not of /r/. She
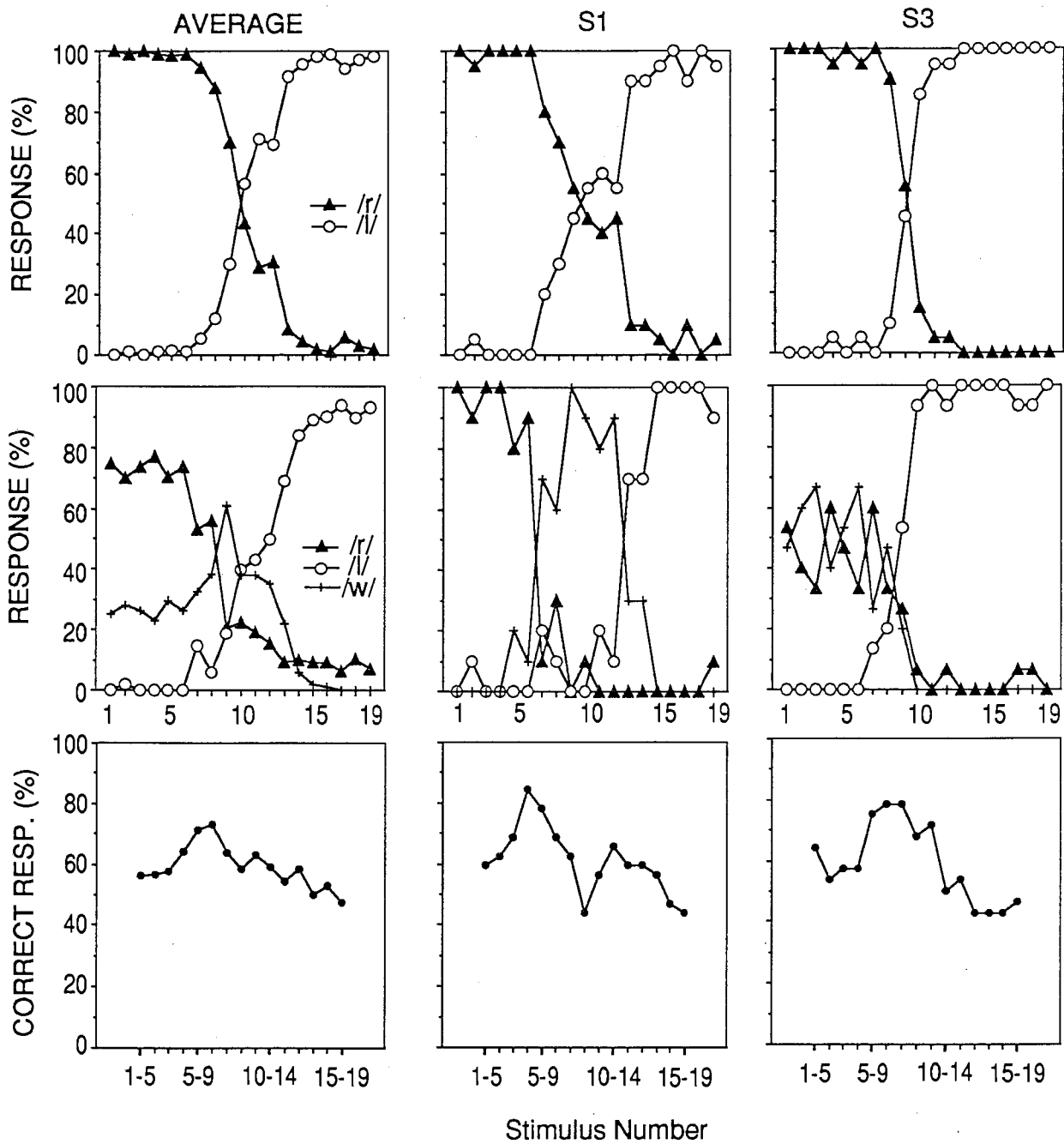
Figure 2    The pooled results in Experiment 1. The left panels show the results pooled across all the listeners while the middle and right panels show the individual results for S1 and S3, respectively. The upper panels represent the pooled identification rates for each stimulus in the RL-condition, where response categories were only /r/ and /l/. The middle panels represent the identification rates for each stimulus in the RLW-condition, where response categories were /r/, /l/, and /w/. The bottom panels represent the correct response rates for each stimulus-pair in the ABX discrimination test.

appeared to identify each ST1 to ST8, which she labelled as /r/ in the RL-condition, as /r/ and /w/ evenly for about a 50 % response rate, respectively. On the other hand, she identified ST10 to ST19 as /l/ with a substantially high response rate in both the RL- and RLW- conditions. It is assumed that she perceived the stimuli categorically as /l/ and non-/l/ (i.e., /r/ or /w/). The peak observed in the discrimination results is considered to be a boundary between /l/ and non-/l/.

## Discussion

These experiment results lead us to understand the importance of considering /w/ category when studying the perception of AE /r/ and /l/ for Japanese listeners.

By taking account of the /w/ category, the discrimination performance can be well predicted from the identification performance. For instance, the location of the peak observed in the discrimination test results for the Japanese listeners on average, have never been well understood when the identification was discussed in the RL-condition. However it is revealed in the RLW-condition that the location of the peak corresponds to the /r/-/w/ identification boundary (Fig.2, left panels). This relationship between the identification results and discrimination results will be still more strongly supported when the individual results are studied. As a typical case of perceiving three phonemes, /r/, /w/, and /l/, on the present series, the results for listener S1 are shown in the middle panels in Fig.2. The two discrimination performance peaks can be explained only when the /w/ category is considered.

Furthermore, considerations of the /w/ category have led us not only to a better interpretation of the relationship between the identification and discrimination test results mentioned above, but also a further insight into perceptual mode of individual listeners. From the averaged and individual results it has been revealed that Japanese listeners often perceive a /w/ sound between /r/ and /l/ on the /r-l/ continuum. As a typical case in which the perceptual mode might be misunderstood unless the /w/ category is not considered in the identification test, the results for S3 are shown in the right panels in Fig.2. The results of RL-conditioned identification and discrimination may be indicate that the subject is capable of perceiving /r/ and /l/ categorically because her response rates in
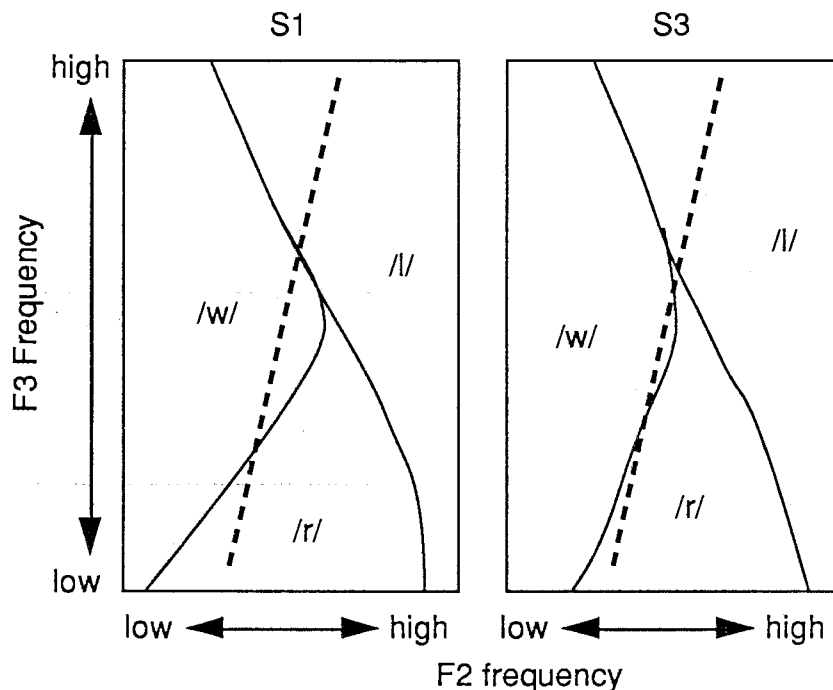


Figure 3. Schematic identification patterns on the F2-F3 plane for the listeners, S1 (left panel), and S3 (right panel). The identification boundaries between three responses, /r/, /l/, and /w/ are shown schematically with a plain line. The /r-l/ continuum used in the Experiment 1 is on the dashed line.

identifying /r/ and /l/ were as high as 100% and discrimination results showed one peak (Fig.2, right-top and -bottom panels). However, the results obtained from the RLW-condition identification test (Fig.2, right-middle panel) have given us an entirely new insight into the S3's perceptual mode. That is, what S3 can identify is not the contrast between /r/ and /l/ but that between non-/l/ and /l/.

The non-/l/ responses for S3 seem to rest on two assumptions; (1) the subject has neither distinctive /r/ nor the /w/ category, and (2) the /r/-side half of the present series lies around a boundary between the /r/ and /w/ categories. However, the second assumption has been supported by another test which was designed to see the characteristics of /w/ perception from a different viewpoint. All the listeners in this experiment (S1-S5) participated in this additional test, and in that test, the perceptual cues of AE /r/, /l/, and /w/ for Japanese listeners were investigated in a way similar to another of the authors' studies (Yamada & Tohkura, 1990). As is shown in Fig.3, the stimuli for an identification test were generated by F2 and F3 combinations which cover a reasonable area on an F2-F3 plane. Here, the /r-l/ series used in Experiment1 was a cross section of the plane. The identification results can be drawn as patterns of the /r/, /l/, and /w/ categories on the F2-F3 plane. The identification patterns for subject S3 are represented schematically with the present /r-l/ series as the cross section in Fig.3. This schematic representation shows that S3 has three distinctive categories, /r/, /l/, and /w/ on the F2-F3 plane, but the /r/-side half of the present series is located in the neibourhood of the boundary between /w/ and /r/, and the left half of the series lies in the /l/ category. Thus, these results support the second assumption about S3's responses. For most of the Japanese listeners, however, the present series does not lie close to a category boundary but crosses the /r/, /w/, and /l/ categories (e.g. S1).

When considering the /w/ category it is also important to note is that the experiment task with the /w/ category seemed more natural for Japanese listeners who even in the RL-condition, often claimed that they do perceive the /w/ sound for some stimuli. This, together with suggestions from previous studies, strongly motivated the hypothesis that Japanese listeners have the /w/ category on the AE /r-l/ continuum.

Despite the importance of considering the /w/ category described above, there still remain some issues to be discussed further. The results in Experiment1 might be much too complicated to give a simple understanding and description of the acquisition process regardless of new findings about the perceptual characteristics of the AE /r/ and /l/ for Japanese listeners because of the observed large individual differences among listeners. Note that the individual differences between two subjects S1 and S3 who each showed an entirely different perceptual mode (Fig.2, center and right panels). It is very difficult and sometimes useless to discuss only average results across listeners. How should these individual differences be dealt with? One solution to this problem would be to have a proper classification of the subjects as described in Experiment 2.

## EXPERIMENT 2

A large individual differences in the perception of synthesized AE /r-l/ continuum among Japanese listeners has been observed in previous studies and also in Experiment 1. Most Japanese subjects with no exposure to an English-speaking environment are unable to identify the synthesized continuum consistently, and perceive it continuously. However, some of them perceive the continuum categorically even though they have neither lived in a foreign country nor received any special English conversation lessons. For example, S1, one of the subject in the Experiment 1, perceived the continuum categorically, even though he had categories of not only /r/ and /l/, but also /w/. It is thought to be important to investigate the perceptual characteristics of such Japanese subjects, who may have rather distinct categories of /r/ and /l/, when the goal of the research is to understand the process of acquiring L2 phonemes, because the perceptual mode for such subjects may suggest how Japanese subjects' perceptual categories of /r/ and /l/ are formed, and how differently such categories are acquired by Japanese listeners compared to native listeners. However, when the averaged results of Japanese listeners are discussed, the data from subjects with distinct categories are buried.

### Purpose

The main purpose of this experiment is to observe such buried data by classifying them according to the identification ability of naturally spoken AE /r/, /l/, and /w/. In addition, the following three points will also be studied: the difference between performance of native AE listeners and that of Japanese listeners, the detailed data about /w/ perception for Japanese listeners, which was investigated in Experiment 1, using a larger population of subjects, and the correlation between the performance in the experiment using naturally spoken words and that using synthesized speech.

## Method

**Stimuli.** Both naturally spoken and synthesized stimuli were used. Synthesized stimuli were the same as used in Experiment 1. For naturally spoken words, sixteen word triads were used as speech materials. In each triad, words were different from each other only in the initial consonant, /r/, /l/, or /w/ (e.g. "red", "led", and "wed"). A total of forty-eight words spoken by two native AE speakers (one male and one female) to produce a total of ninety-six stimuli. They were recorded on tape using a DAT recorder, and digitized at a 20-kHz sampling frequency with 16-bit accuracy. For the identification test of naturally spoken stimuli, the stimuli were reproduced at a 20-kHz sampling frequency with a cutoff frequency of 10kHz, and several sessions with different random-izations of stimuli were recorded on a 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 identi-fication test of stimuli in Experiment 1. These stimuli were presented binaurally over headphones in the same way as in Experiment 1 at the fixed level which was comfortable for listeners.

**Subjects.** Ten native AE listeners (Group-A: A1-A10), 36 native Japanese listeners who had no ex-perience living in a foreign country (Group-J: J1-J36), and three native Japanese listeners who had resided in the U.S.A. (Group-JE: JE1-JE3) served as subjects. The age of listeners in Group-A was 30.2 on average, and ranged from 23 to 40, and that in Group-J was 20.1 on average, and ranged from 18 to 22. All the Group-A subjects were raised in the U.S.A. One of them had been living in Japan

for two years and others for less than one year when they participated in this experiment. The age at the time of the experiment and the age range while living in the U.S.A. for the Group-JE listeners were 26 and 6 to 8 for JE1, 26 and 24 to 25 for JE2, and 23 and 21 to 22 for JE3, respectively. All listeners reported no history of hearing and speaking disorder.

**Procedure.** The experiment consisted of three types of tests: identification tests of synthesized stimuli, ABX discrimination tests of synthesized stimuli, and identification tests of naturally-spoken stimuli. The procedure for the identification tests and the ABX dis-crimination tests using synthesized stimuli were identical to those of the RLW-conditioned identification tests and the ABX discrimination tests in the Experiment 1.

The procedure for the identification test of naturally spoken stimuli was similar to that of synthesized stimuli. Listeners were asked to identify the word initial consonants, and to make a forced choice among the given response categories, /r/, /l/, and /w/, even when it was difficult to choose one. They were also told that there might exist unfamiliar or meaningless words.

Each listener participated in three sessions of the identification tests for the synthesized stimuli, six sessions of the ABX discrimination tests, and two sessions of the identification tests for the naturally spoken stimuli.

## Results

Figure 4 shows the identification and ABX discrimination results for the synthesized /rait-lait/ con-tinuum, pooled across all ten AE listeners (left panels), and 36 Japanese listeners (right panels). In Group-A, the listeners consistently identified six /r/-side stimuli ( ST1 to ST6 ) as /r/ and six /l/-side stimuli (ST14 to ST19) as /l/. The /w/ response appeared for ST7 to ST13, and peaked at ST11 even with low rates. These /w/ responses occurred for half of the AE listeners, but they did not occur at all for the rest of them. The results from the ABX discrimination tests in Group-A peaked at around ST10. In Group-J, though response patterns are inconsistent among listeners, they tended to identify seven /r/-side stimuli (ST1-ST7) as /r/, seven /l/-side stimuli (ST13-ST19) as /l/ at rates of 60 to 80 %, and some stimuli around ST10 as /w/ with low rates. As is seen in the response patterns for each category, the three responses changed
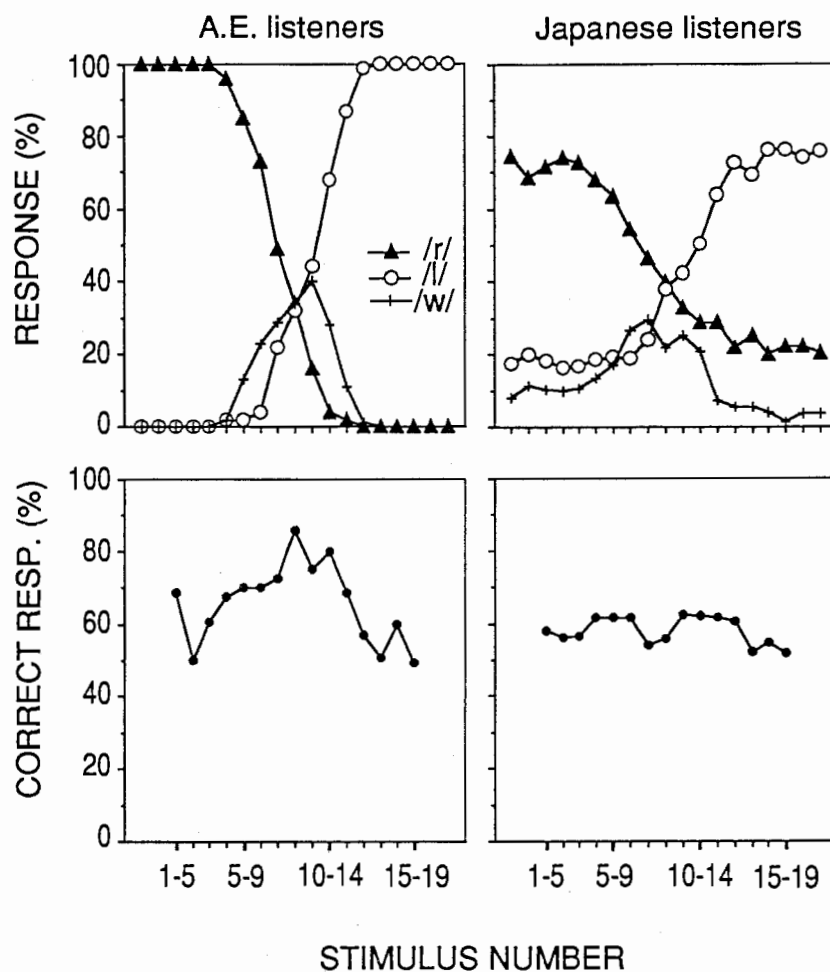
**Figure 4** The averaged results across A.E. listeners (left panels) and Japanese listeners (right panels) in Experiment 2. Pooled identification rates for each each stimulus are represented in the upper panels, and averaged correct response rates for each stimulus-pairs in the 4-step ABX discrimination test are represented in the lower panels.

gradually and continuously in Group-J. The results from the ABX discrimination tests in Group-J seemed flat, and showed no clear peak, and the response rates to all stimulus-pairs stayed slightly higher than the chance level of 50% accuracy.

The pooled correct response rates in identification tests for naturally spoken stimuli $(Cn)$ were 99.8% in Group-A, 65.4% in Group-J, and 85.3% in Group-JE. Group-J listeners were classified into the following four groups according to individual identification performance for the naturally spoken stimuli: Group J-50 (50% ≤, and < 60% in accuracy), Group J-60 (60% ≤, and < 70% in accuracy), Group J-70 (70% ≤, and < 80% in accuracy), and Group J-80 (80% ≤, and < 90% in accuracy). All the listeners in Group-J belonged to these four groups. The

number of listeners and pooled $Cn$ across listeners in each of four groups were 14 listeners and 54.9% in Group J-50, 10 listeners and 64.6% in Group J-60, 7 listeners and 73.1% in Group J-70, and 5 listeners and 84.1% in Group J-80.

Figure 5 shows the results of synthesized stimuli, pooled across the listeners in each of the four groups, according to the above definition. A positive correlation between the performance for the naturally spoken stimuli and that for the synthesized stimuli were observed. The response differentiation among the three categories, /r/, /l/, and /w/, in identification results was quite insignificant in Group J-50, and in Group J-60. The discrimination results seemed flat and had no peak in both Group J-50 and J-60. Compared with these two groups, the identi-
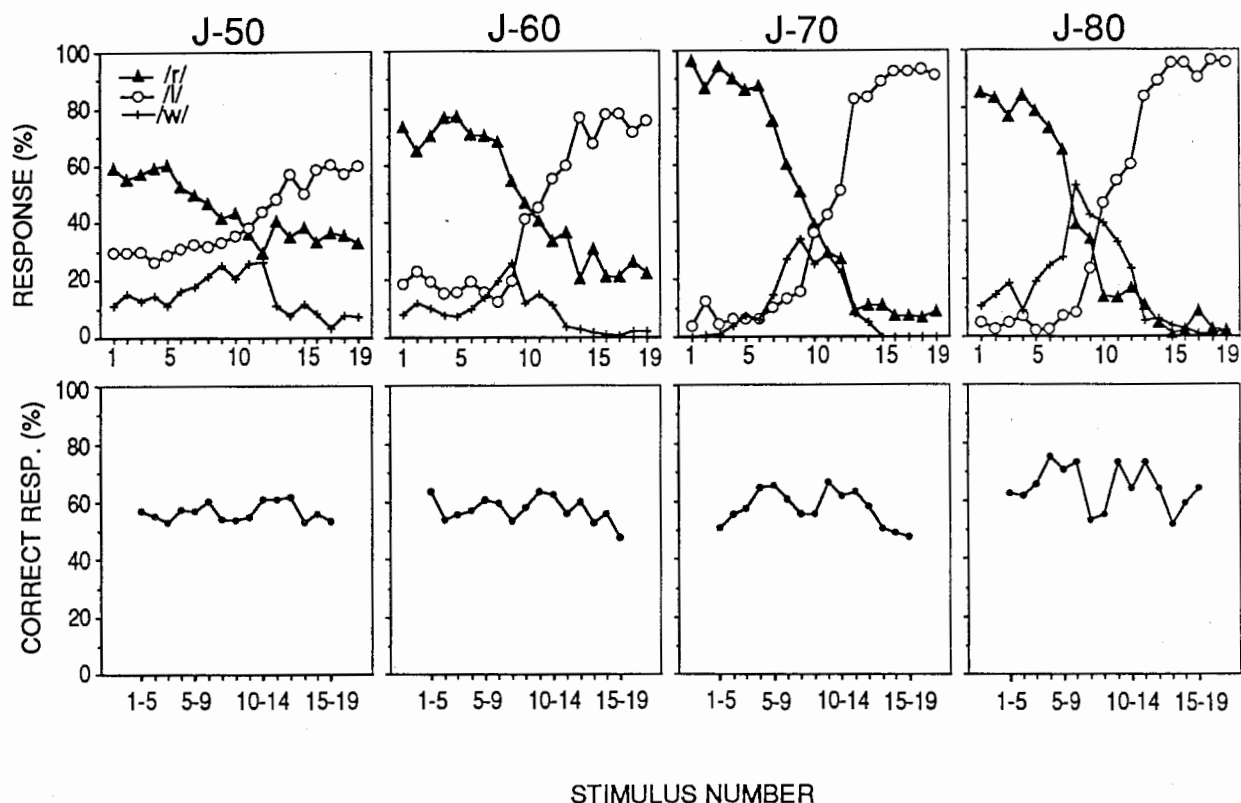
-10-

**Figure 5** The pooled identification rates (upper panels) and correct response rates in the 4-step ABX discrimination test (lower panels) for four groups of Japanese listeners, Groups J-50, J-60, J-70, and J-80, in Experiment 2. Japanese listeners are classified into above four groups according to the correct response rate in identifying the naturally spoken /r, l, w/ sounds. The J-50 group showed the lowest, and J-80 the highest accuracy in identifying naturally spoken stimuli.

fication results appeared much more consistent and gave higher rates in Groups J-70 and J-80. In these two groups two peaks were observed in the discrimination results. The peaks were slightly higher in J-80 than in J-70. For each group, it was found that the two peaks corresponded to the two identification boundaries, one between /r/ and /w/, and the other between /w/ and /l/.

The correlation between the identification performance of naturally spoken stimuli and that of synthesized stimuli across subjects was estimated as follows. As a value, $Cs$, which represents response consistency for all 19 stimuli on the synthesized series, was defined as the averaged value of the highest response rates across the 19 stimuli. Spearman's rank-order correlation between $Cs$ and $Cn$ across Group-J listeners showed a significant positive correlation ( $r = + .60$, $p < .0004$ ).

In Group-JE, the correct identification rates of naturally spoken stimuli were 97.9% for JE1, 81.7% for JE2, and 70.8% for JE3. On average, the relationship between results of naturally spoken stimuli and those of

synthesized stimuli is similar to that observed in Group-J. However, there were also observed large individual differences among these three listeners in Group-JE. JE1 perceived the synthesized /r-l/ continuum categorically like AE listeners did, JE2 perceived it non-categorically (i.e., with low identification rates and without clear peaks in the discrimination results), and JE3 identified the stimuli with higher consistency than JE2 did but without any clear peaks in the discrimination results.

## Discussion

Comparing the perceptual mode of AE listeners and that of Japanese listeners on average across all the listeners in each group (Fig.4), a simple conclusion is obtained: AE listeners perceive the /r-l/ series categorically, but Japanese listeners perceive the same series continuously. This conclusion is the same one which has been revealed in the previous studies. However, by

-11-

analytic observation of the experiment results, the following findings about the perceptual characteristics of the /r-l/ series for Japanese listeners were obtained. (1) Perceptual mode of Japanese listeners was clarified by observing the classified data according to the listeners' identification ability of naturally spoken stimuli. (2) The assumption that Japanese listeners perceive /w/ on the /r-l/ series which had been found in the Experiment 1 was confirmed. (3) The results of the experiment using naturally spoken stimuli are highly correlated to those of the experiment using the synthesized stimuli.

With respect to the perceptual modes for each classified listener groups shown in Fig.5, in Groups J-50 and J-60 whose ability to identify the naturally spoken words was low, the responses in the identification test showed no consistency, and those in the ABX discrimination test stayed almost at the chance level. In Groups J-70 and J-80, who showed higher identification ability in the test using naturally spoken words, the responses in the identification test seemed consistent, and discrimination peaks appeared. These peaks seemed higher in Group J-80 than that in Group J-70. These results imply that the acquisition process of /r/ and /l/ contrast for Japanese listeners. Japanese listeners who have poorly acquired the /r/ and /l/ contrast perceive the /r-l/ series continuously as has been proved in many previous studies. However, some of the Japanese listeners acquire the /r/ and /l/ contrast so well that they can identify the stimuli near the prototype with high accuracy. However, even these Japanese listeners still perceive a /w/ sound on the series.

The perception of /w/ is supported by the results of the ABX discrimination test. There appeared two peaks in Groups J-70 and J-80. On the whole, these results confirm the finding of Experiment 1, that /w/ is perceived by Japanese listeners who have acquired AE /r/ and /l/ rather well. In contrast, Group-A showed (Fig. 4) one broad peak in the discrimination result. This broad peak resulted from the fact that half of the Group-A listeners perceived some stimuli on the continuum as /w/ and others did not. The two notable peaks observed in Japanese skilled groups (J-70 and J-80) might be caused by the shift of the category boundary between /w/ and other two phonemes, /r/ and /l/, from that of AE listeners. As has been discussed in Experiment 1, the present stimulus series crosses the /w/ category for Japanese listeners. It was also revealed that the present series slightly crosses the /w/ category for AE listeners by the

authors (Yamada & Tohkura 1990). However, it should be noted that no prototype /w/ occurs on the series. The acoustic characteristic which distinguishes /w/ sound from other liquids is the low frequency of F2 (Lisker, 1957). The value of the F2 frequency for the present stimulus series is higher than that for /w/. Therefore, the stimuli which were identified as /w/ by Japanese listeners are thought to be far from the prototype /w/. The rather low level of the identification rate for the /w/ category is presumably related to this characteristic of the stimuli.

Even after such detailed investigation, there still remain two possibilities for the origin of the /w/ responses for Japanese listeners; (1) the effect of the Japanese /w/ category, and (2) the effect of the response bias caused by using a forced-choice task among three response categories. In the first assumption, the /w/ responses for Japanese listeners are thought to be affected by the fact that among the present three categories, only /w/ is similar to the Japanese /w/. Presumably, this /w/ category acquired as one of the L1 phonemes is thought to be so robust as to have a strong influence on category learning of L2 at older ages. In the second assumption, Japanese listeners tend to choose three response categories at similar occurrences in their answers, because they don't have distinctive categories of those phonemes and cannot make absolute judgements. In contrast, AE listeners are hardly affected by the stimulus set because they have distinctive categories. As a result, Japanese listeners are more easily affected by response bias caused by adding /w/ to the response categories than are AE listeners. Further investigation based on these assumptions is needed in order to understand the /w/ responses for Japanese listeners. The second assumption is somewhat associated with the experiment task, which is the main issue in this paper. Accordingly, the effect of response bias was investigated by studying the stimulus set effect in Experiment 3.

When the success in classifying subjects is considered from a different viewpoint, it indicates a positive correlation between the performance in the experiment using synthesized stimuli and that using naturally spoken stimuli. This correlation has been revealed to be significant statistically. In general, however, it depends on the quality of synthesized speech, because there is a possibility that the experiment results are affected by the quality of synthesized stimuli especially when studying non-native phoneme perception. Further investigation is needed to see the general correlation

between the perception of synthesized stimuli and that of naturally spoken stimuli

The results also showed that there are large individual differences in Japanese listeners' perception, even though the quality and quantity of their English education in high school was similar, and none of them had lived abroad. This indicates that such large individual differences among Japanese listeners depend not only on the quality and the quantity of their English education but also on their individual attitude to such education. In the previous study (MacKain, et al.; 1981), Japanese subjects who had moved to the U.S.A. as adults were classified into two groups (i.e., experienced, and non-experienced) according to the quality and quantity of their English conversation experience in the U.S. The results showed that the experienced group perceived /r/ and /l/ more like Americans do than the non-experienced group. However, this classification of subjects with respect to their English conversation experience might be too crude to investigate the individual perceptual characteristics of the subjects. Moreover, there still might be large individual differences in /r,l/ perception within the experienced and non-experienced groups. The effect of the quality and quantity of their employment of English in the U.S. on perception could be discussed more efficiently by taking account of the subjects' ability to perceive /r/ and /l/ contrast when they began living in the U.S.

The experience of living in the U.S. is an important variable when studying /r/ and /l/ perception of Japanese listeners. JE1, a 26-year-old subject who had resided in the U.S.A. from age 6 to 8, showed the high identification ability of 97.9% correct in the test using naturally-spoken stimuli. The results from the test using the synthesized stimuli revealed that JE1 does not have the /w/ category on the /r-l/ continuum used in the experiment (i.e., she had no /w/ response in the identification test, and there was one peak in the ABX discrimination test). In contrast, the other two experienced listeners, JE2 and JE3 who had lived in the U.S., from age 24 to 25, and from age 21 to 22, respectively, perceived the series non-categorically. This result indicates that JE2 and JE3 could not have acquired the /r/ and /l/ contrast from their experience in the U.S. The perceptual mode difference between JE1 and the other Japanese listeners in Group JE and Group J might be brought about by the fact that JE1 resided in the U.S. from age 6 to 8 while, when the others resided in the U.S., they were older (21 to 22, and 24 to

25). JE1 is thought to be exposed to the English speaking environment young and long enough to acquire the perception of /r/ and /l/ equivalent to a native AE speakers. As the present study includes only three subjects who had lived in the U.S., the relationship between the acquisition of /r/ and /l/ contrast and the age and period of exposure to the AE speaking environment is still open to question. A detailed investigation to answer this question may lead us to a better understanding of how speech perception is modified by the linguistic environment.

## EXPERIMENT 3

### Purpose

The AE listeners perceive the synthesized /r-l/ series categorically, but most of the Japanese listeners do not perceive it categorically as has been revealed in previous studies and the above two experiments in this paper. This finding implies the possibility that Japanese listeners' results are affected by the stimulus range presented in the experiment session even though the AE listeners' results are not affected because the AE listeners' judgement is thought to be more absolute than the Japanese listeners' judgement. The main purpose of this experiment is to observe the effect of the stimulus range on the results for AE listeners and Japanese listeners. In addition, the following issues will be also argued. (1) Do Japanese listeners tend to choose given response categories relatively so that response occurrences for each category are balanced regardless of the stimuli in a given set? (2) Is there any difference between the effect of the stimulus range on the /r/ and /l/ responses and that on /w/ response? (3) Do the /w/ responses for Japanese listeners which were observed in Experiments 1 and 2 appear as the effect of response bias by adding /w/ to the response categories? In order to argue these issues, the relationship between the judgement and the stimulus range was studied.

### Method

Stimuli. The synthetic /rait-lait/ series which consisted of 17 stimuli (ST1 through ST17) was used. The onset frequency of F2 and F3 was varied from 960 Hz

to 1280 Hz in 20Hz steps, and 1400Hz to 3000Hz in 100Hz steps, respectively. These F2 frequency and F3 frequency conditions for ST1 to ST17 in this experiment correspond to ST3 to ST19 in Experiments 1 and 2. The durations of F1 steady state and transition were kept at 80msec and 70msec for all stimuli. Other conditions for the stimuli are identical to those used in Experiments 1 and 2.

There were three stimulus set conditions: /r/-side set, /l/-side set, and full set. The /r/-side set and the /l/-side set consisted of ten /r/-side stimuli (i.e. ST1 to ST10), and ten /l/-side stimuli (i.e. ST8 to ST17), while the full set consisted of all 17 stimuli on the series. The stimuli in each set were arranged in random order with five time repetitions. At the beginning of each session, twenty dummy stimuli which consisted of the stimuli in each set were added for familiarization. Several sessions (i.e. randomizations) per set were recorded on digital audio tape under conditions identical to those in Experiments 1 and 2.

**Subjects and procedure.** Seventy-three native speakers of Japanese who had never lived in foreign countries, and 5 native speakers of AE served as listeners. All the listeners were naive for this kind of experiments on /r/ and /l/ perception. The Japanese listeners were mainly undergraduate students. The age of AE listeners was 25 years old in average, and varied from 20 to 30 years old. All the listeners reported no history of hearing and speaking disorder.

Each listener participated in two sessions. The total of Japanese listeners were divided into three groups, each of which participated in one stimulus condition in the first session; 27 listeners for the /r/-side set (/r/-side condition), 24 listeners for the /l/-side set (/l/-side condition), and 22 listeners for the full set (full condition). This full condition is the control to see the difference between the response pattern in the first session and that in the second session for Japanese listeners. The five AE listeners were divided into two groups, consisting of 3 and 2 listeners, respectively. One group (3 listeners) participated in the /r/-side set (/r/-side condition) and the other (2 listeners) participated in the /l/-side set (/l/-side condition) in the first session. In the second session, all the listeners participated in the full set.

The stimuli were presented through headphones (STAX SR Lambda Professional) in a sound-proof room.

The listeners were instructed to judge the initial consonant of each stimulus, and forced to choose one of the three response categories, /r/, /l/, or /w/. They were told that they had to choose one of the response categories they heard in each trial regardless of the frequency of occurrence for each category through an entire session.

**Results**

**The /r/-side condition.** The identification curves for the /r/-side condition were shown in Fig. 6. For the AE listeners, identification rates for each stimuli in the /r/-side set are similar to those in the full set. They identified ST1 to ST6 as /r/ with high response rates of 93.3% to 100%, ST7 as /r/ with 73.3%, ST8 as /r/ with about 50% in both /r/-side and full set sessions. This indicates that the stimulus set has only a small influence on the identification results for AE listeners.

For the Japanese listeners, in contrast, identification judgement was strongly affected by the stimulus set. Identification rates for the stimuli in the /r/-side set were much different from those in the full set. The Japanese listeners identified ST1 to ST5 as /r/, /l/, and /w/ with response rates of about 60%, 20%, and 20%, respectively, in the full range set. However, they identified the same stimuli (ST1 to ST5) differently in the /r/-side set. On the whole, they identified them less frequently as /r/ (40%-50%), and more frequently as /l/ (30-40%) than they did in the full set. When the /r/, /l/, and /w/ identification curves are compared to each other, the /w/ identification curve for ST1 to ST10 for Japanese listeners was much more similar between two sets than those of the /r/ and /l/ identification curves. Table 1 shows the differences in the identification rates for each of ST1 to ST10 between two sessions for Japanese listeners. The differences between the two rates were evaluated statistically by a Wilcoxon signed-rank test. The /r/-identification rates were significantly lower in the first /r/-side set than in the second full set for ST5 ( $z=2.9$, $p < .005$ ), ST6 ( $z=2.1$, $p < .05$ ), and ST8 ( $z=3.1$, $p < .002$ ). The /l/-identification rates were significantly higher in the first /r/-side set than in the second full set for ST1 ( $z=2.3$, $p < .02$ ), ST2 ( $z=2.4$, $p < .02$ ), ST3 ( $z=2.1$, $p < .04$ ), ST5 ( $z=2.2$, $p < .03$ ), ST6 ( $z=2.4$, $p < .02$ ), ST9 ( $z=2.0$, $p < .05$ ), and ST10 ( $z=2.7$, $p < .007$ ). In the /w/-identification rates, in contrast, the significant difference was observed only for ST8 ( $z=2.1$, $p < .04$ ).

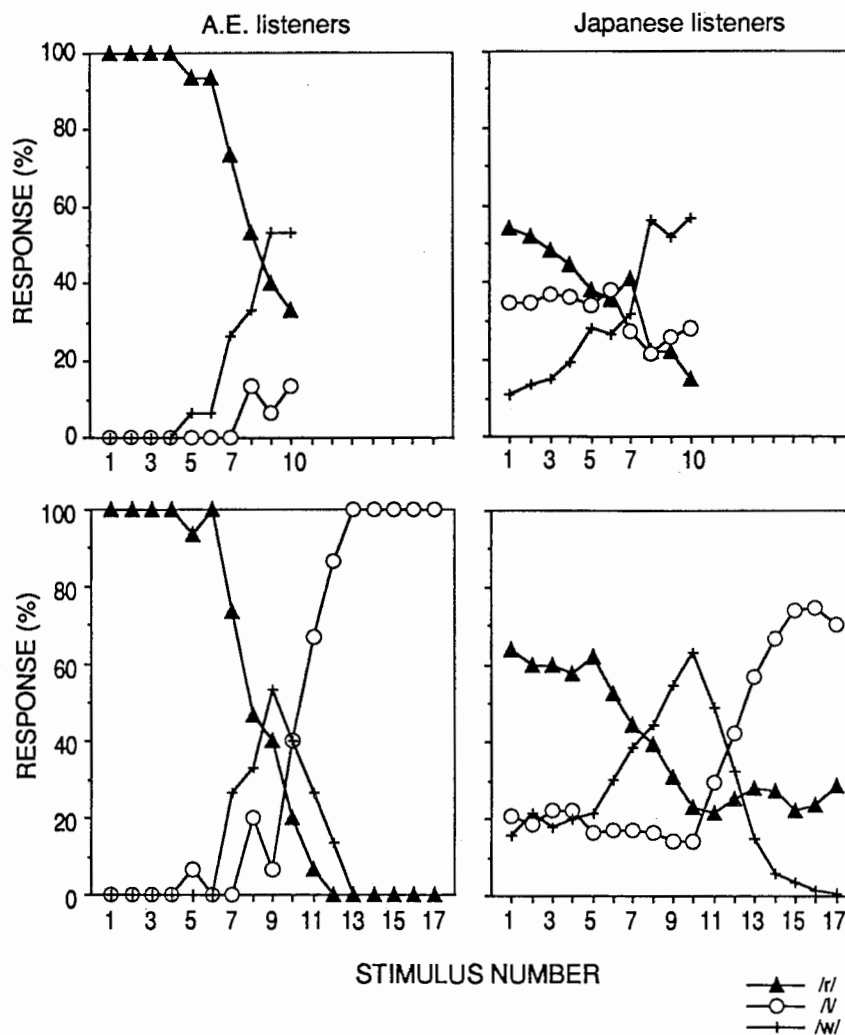For the purpose of comparing the overall re-

**Figure 6** The results of /r/-side condition for A.E. listeners (left panels) and those for Japanese listeners (right panels) in Experiment 3. The upper panels show the pooled identification rates for each response category in the first /r/-side session, where only /r/-side ten stimuli (ST1 to 10) are presented to the listeners. The lower panels show the pooled identification rates in the second full set session, where all the stimuli on the continuum are presented.

sponse rates in each set, the averaged response rates for each response category across all the stimuli (ST1 to ST10) in the /r/-side set, ST1 to ST10 in the full set, and all the stimuli (ST1 to ST17) in the full set were shown in Table 2. The distribution of observed /r/, /l/, and /w/ response rates pooled across all the stimuli (ST1-ST10) in the first session (D1) was compared to that for ST1-ST10 in the second session (D2_reduced) and ST1-ST17 (D2_full) in the second session using the chi-square test (df=2). The chi-square score between D1 and D2_reduced was 1.7 and lower than that between D1 and D2_full ($\chi 2$ = 43.9, p < .001 ) for A.E. listeners. In contrast, for the Japanese listeners, the chi-square score between D1 and

D2_reduced was 5.7 ( p < .1 ) and higher than that between D1 and D2_full ( $\chi 2 = 0.7$ ).

**The /l/-side condition.** The identification curves for the /l/-side condition were shown in Fig. 7. For the AE listeners, overall identification characteristics for the stimuli in the /l/-side set were similar to those for the same stimuli (ST8 to ST17) in the full set. A slight difference between the two sets was observed: the identification boundary between /r/ and /w/ shifted from between ST10 and ST11 in the /l/-side set to around ST9 in the full set. They identified the /l/-side 6 stimuli (i.e., ST12 to ST17) with high response rates of 90% to 100% in both the /l/-

-15-

Table 1 Difference in the identification rates for each of the common stimuli between the first session and the second session (ST1-ST10 in the /r/-side condition, ST8-ST17 in the /l/-side condition, and ST1-ST17 in the full condition) for Japanese listeners in Experiment 3. The response rate for each stimulus in the 2nd session was subtracted from that for the same stimuli in the first session.

| | Difference (%) | | | | | | | | |
| | /r/-side condition | | | /l/-side condition | | | full condition | | |
| | /r/ | /l/ | /w/ | /r/ | /l/ | /w/ | /r/ | /l/ | /w/ |
|---|---|---|---|---|---|---|---|---|---|
| ST1 | -9.6 | 14.1* | -4.4 | – | – | – | -13.6 | 8.2 | 5.5 |
| ST2 | -8.1 | 16.3* | -8.1 | – | – | – | 0.0 | 1.8 | -1.8 |
| ST3 | -11.9 | 14.8* | -3.0 | – | – | – | -4.5 | 3.6 | 0.9 |
| ST4 | -13.3 | 14.1 | -0.7 | – | – | – | -11.8* | 5.5 | 6.4 |
| ST5 | -24.4** | 17.8* | 6.7 | – | – | – | -10.9 | 4.5 | 6.4 |
| ST6 | -17.0* | 20.7* | -3.7 | – | – | – | 5.5 | 3.6 | -9.1 |
| ST7 | -3.7 | 10.4 | -6.7 | – | – | – | 9.1 | 0.9 | -10.0 |
| ST8 | -17.0** | 5.2 | 11.9* | 20.8* | -5.0 | -15.8* | -0.9 | 10.9 | -10.0 |
| ST9 | -8.9 | 11.9* | -2.9 | 14.2 | 1.7 | -15.8 | 1.8 | 1.8 | -3.6 |
| ST10 | -8.1 | 14.1** | -5.9 | 15.0 | -4.2 | -10.8 | 10.0* | 1.8 | -11.8* |
| ST11 | – | – | – | 18.3* | -13.3* | -5.0 | 3.6 | 11.8* | -15.5* |
| ST12 | – | – | – | 20.0* | -19.2* | -0.8 | 7.3 | 20.9* | -28.2** |
| ST13 | – | – | – | 18.3 | -17.5* | -0.8 | 1.8 | 13.6 | -15.5* |
| ST14 | – | – | – | 10.8 | -13.3 | 2.5 | -8.2 | 21.8** | -13.6** |
| ST15 | – | – | – | 10.8 | -10.0 | -0.8 | 6.4 | 2.7 | -9.1* |
| ST16 | – | – | – | 15.8* | -20.0* | 4.2 | -1.8 | 3.6 | -1.8 |
| ST17 | – | – | – | 20.8* | -20.8* | 0.0 | 1.8 | -1.8 | 0.0 |

* significant at p < 0.05
** significant at p < 0.01

Table 2  Pooled percentages of /r/, /l/, and /w/ identification across the stimuli for each condition in Experiment 3 are shown. For the /r/-side and /l/-side conditions, the followin g three observed percentages are shown; (1) pooled across all the stimuli in the first session, (2) pooled across common stimuli between the two sessions in the second session, and (3) pooled across all the stimuli in the second session. For the full condition, pooled precentages across all the stimuli in each of two session are shown. The chi-square values between the frequencies in the first session and each of the observed frequencies in the second session are also presented .

| | A.E. listeners Response rate (%) | | | $\chi^2$ | J. listeners Response rate (%) | | | $\chi^2$ |
|---|---|---|---|---|---|---|---|---|
| | /r/ | /l/ | /w/ | | /r/ | /l/ | /w/ | |
| /r/-side condition | | | | | | | | |
| 1st session (/r/-side set) | | | | | | | | |
| ST1-ST10 | 78.7 | 3.3 | 18.0 | - | 37.2 | 31.8 | 31.0 | - |
| 2nd session (full set) | | | | | | | | |
| ST1-ST10 | 77.3 | 7.3 | 15.3 | 1.7 | 49.4 | 17.9 | 32.7 | 5.7$^\S$ |
| ST1-ST17 | 45.9 | 42.7 | 11.4 | 43.9*** | 39.5 | 34.9 | 25.6 | 0.7 |
| | | | | | | | | |
| /l/-side condition | | | | | | | | |
| 1st session (/l/-side set) | | | | | | | | |
| ST8-ST17 | 26.0 | 65.0 | 9.0 | - | 38.9 | 39.3 | 21.8 | - |
| 2nd session (full set) | | | | | | | | |
| ST8-ST17 | 13.0 | 67.0 | 20.0 | 8.5* | 22.4 | 51.4 | 26.2 | 6.5* |
| ST1-ST17 | 48.2 | 39.4 | 12.4 | 13.5** | 37.0 | 36.7 | 26.3 | 0.6 |
| | | | | | | | | |
| full condition | | | | | | | | |
| 1st session (full set) | | | | | | | | |
| ST1-ST17 | - | - | - | - | 36.8 | 40.9 | 22.3 | - |
| 2nd session (full set) | | | | | | | | |
| ST1-ST17 | - | - | - | - | 37.1 | 34.1 | 28.8 | 1.4 |

*   significant at $p < 0.05$
**  significant at $p < 0.01$
*** significant at $p < 0.001$
§   significant at $p < 0.1$
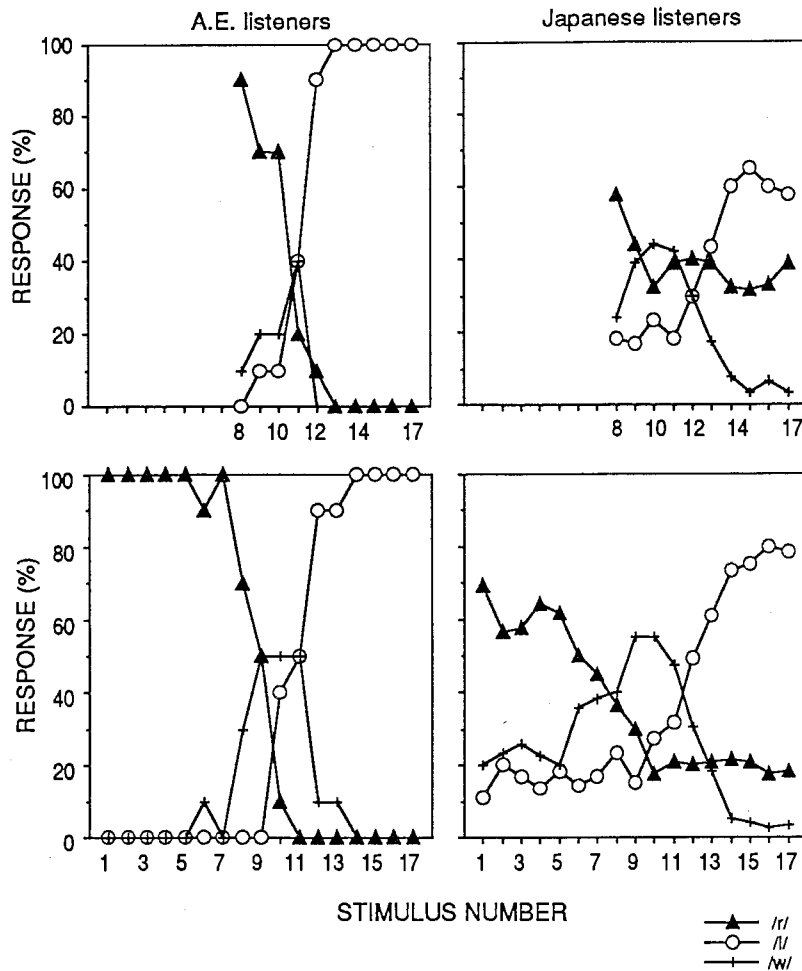
**Figure 7** The results of /l/-side condition for A.E. listeners (left panels) and those for Japanese listeners (right panels) in Experiment 3. The upper panels show the pooled identification rates for each response category, /r/, /l/, and /w/, in the first /l/-side session, where only /l/-side ten stimuli (ST8-17) are presented to the listeners. The lower panels show the pooled identificatoin rates in the second full set session, where all the stimuli on the continuum are presented.

side and full-side sessions. In general, we can conclude that AE listeners' identification is hardly influenced by the stimulus set .

For the Japanese listeners, identification judgement was strongly affected by the stimulus set as was also observed in the /r/-side condition. The identification rates for the stimuli in the /l/-side set were much different from those in the full set. They identified ST10 to ST17 as /r/ at about a 20% response rate in the full set, but at a 30%-40% response rate in the /l/-side set. Instead, they identified the same stimuli (ST10 to ST17) less frequently as /l/ in the /l/-side set than they did in the full set. Again, the /w/ identification curve for ST8 to ST17 was much more similar between two sets than the /r/ and /l/ identification curves. The differences in the identification rates for each of ST8 to ST17 between two sessions are shown

in Table 1. The /r/-identification rates were significantly higher in the first /l/-side set than in the second full set for ST8 ( z=2.4, p<.02 ), ST11 ( z=2.4, p<.02 ), ST12 ( z=2.3, p<.03 ), ST16 ( z=2.0, p<.05 ), and ST17 ( z=2.5, p<.02 ). The /l/-identification rates were significantly lower in the first /l/-side set than in the second full set for ST11 ( z=2.4, p<.02 ), ST12 ( z=2.3, p<.03 ), ST13 ( z=2.0, p<.05 ), ST16 ( z=2.5, p<.02 ), and ST17 ( z=2.5, p<.02 ). In the /w/-identification rates, in contrast, the significant difference was observed only for ST8 ( z=2.0, p<.05 ).

The averaged response rates for each response category pooled across all the stimuli in the first session (ST8-ST17) from the identification frequencies pooled across ST8-ST17 and ST1-ST17 in the 2nd session are shown in Table 2. The distribution of observed /r/, /l/, and /w/ response rates pooled across all the stimuli (ST8-

ST17) in the first session (D1) was compared to that for ST8-ST17 in the second session (D2_reduced) and ST1-ST17 (D2_full) in the second session using the test for independence. The chi-square score between D1 and D2_reduced ( $\chi 2$ = 8.5, p < .05 ) was lower than that between D1 and D2_full ( $\chi 2$ = 13.5, p < .01 ) for A.E. listeners. However, for Japanese listeners, in contrast, the chi-square score between D1 and D2_reduced ( c2 = 6.5, p < .05 ) was higher than that between D1 and D2_full ( $\chi 2$ = 0.6).

**The full condition.** The identification curves of the full condition for were shown in Fig. 8. Only Japanese listeners participated in this condition. On the whole, the identification curves were much more similar between two sessions than in the former two conditions (Fig.6-8). The /r/ judgement was more consistent between two sessions than /l/ or /w/ judgement. Significant differences in /r/-identification rates were observed only at ST4 ( z=2.2, p<.04 ), and ST10 ( z=2.3, p<.02 )(Table 1). The /l/ judgement and /w/ judgement were different between sessions at around ST11 to ST14. Significant differences in /l/-identification rates were observed at ST11 ( z=2.3, p<.03 ), ST12 ( z=2.4, p<.02 ), and ST14 ( z=3.1, p<.003 ), and those in /w/-identification were observed at ST10 ( z=2.2, p<.03 ), ST11 ( z=2.1, p<.04 ), ST12 ( z=2.9, p<.004 ), ST13 ( z=2.3, p<.02 ), ST14 ( z=2.8, p<.006 ), and ST15 ( z=2.1, p<.04 ). The listeners identified these stimuli around ST11-ST14 as /l/ at higher response rates, and as /w/ at lower response rates in the first session than in the second session. No significant difference was observed between the distribution of /r/, /l/, and /w/ for pooled identification rates across all the stimuli in the first session and that in the second session (Table 2, $\chi 2$ = 1.4).

## Discussion

From the results of this experiment, the following three findings have been obtained about /r-l/ perception for Japanese listeners.

First, the range of stimuli on the /r-l/ continuum presented to the listeners in one experiment session significantly affects the identification judgement of Japa-
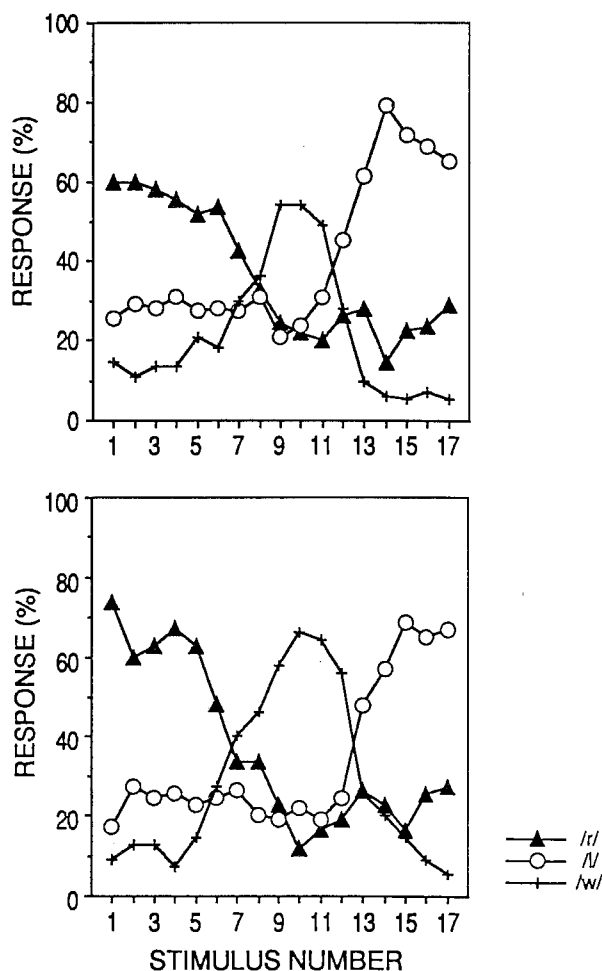


Figure 8 The results of full condition for Japanese listeners in Experiment 3. The pooled identification rates for each reponse category, /r/, /l/, and /w/, in the first session (upper panel), and those in the second session (lower panel) are represented. All 17 stimuli on the continuum are presented in both sessions.

nese listeners while it does not affect that of AE listeners. In the /r/-side condition, where the range of the stimuli are different between two sessions, the identification results for the /r/-side stimuli which are common for both sessions were significantly different from each other for Japanese listeners, but similar for A.E. listeners. Japanese listeners identified the /r/-side stimuli as /r/ at a lower rate and as /l/ at a higher rate in the /r/-side set than in the full set. In the /l/-side condition, identification judgement for Japanese listeners was also different between the two sessions. They identified the /l/-side stimuli as /r/ at a higher rate and as /r/ at a lower rate in the /l/-side set than in the full set. These results indicate that Japanese listeners' identi-

-19-

fication judgement for the /r-l/ continuum is strongly affected by the range of the stimuli. In the full-condition, a significant difference in identification judgement between two sessions was observed around ST11-ST14 for Japanese listeners. This difference has to be discussed further because the full condition is a control condition to see whether the results are influenced by the session order. As the stimuli ST11-ST14 belong to the /l/-side of the continuum, the difference between two sessions in the /l/-side condition is compared with that in the full condition. As a result of the comparison, it is noted that the directions of the difference in the full and the /l/-side conditions are opposite to each other. In the /l/-side condition, a higher /l/ and lower /w/ identification were observed in the second session than in the first session. For the full condition, in contrast, a lower /l/ and higher /w/ identification were observed in the second session than in the first session. Thus, these results in the full condition discussed above even strengthen the influence of the stimulus range on listeners' identification.

Second, Japanese listeners identify some stimuli as /w/ rather independently of the stimulus range than they identify /r/ and /l/. When observing the range effect on Japanese listeners' results according to the response categories, the /r/ and /l/ responses are much more affected by the stimulus range than the /w/ responses. The /w/ identification rates were not significantly different between sessions in both /r/-side and /l/-side conditions, where the range of the stimuli differed. These results suggest that the /w/ responses observed in Experiments 1 and 2 are not the result of response bias caused by adding /w/ to the response categories, because the /w/ responses are rather independent of the stimulus range.

Third, Japanese listeners tend to make a relative judgement between /r/ and /l/. In both /r/-side and /l/-side conditions, the identification patterns for the /r/, /l/ and /w/ categories in the first session were similar to those for the full range stimuli in the second session, but different from the partial identification patterns for the /r/-side and /l/-side stimuli in the second session. If Japanese listeners are capable of making an absolute judgement between /r/ and /l/, the identification patterns obtained in the /r/-side and the /l/-side conditions have to be similar to the partial identification patterns for the same /r/-side and /l/-side stimuli presented among the full range stimuli. In fact, A.E. listeners' identification rates in the first session were rather similar to partial identification rates for the /r/-side

and /l/-side stimuli than to global rates for the full range stimuli in the second session (Table 2). In conclusion, as results of relative judgement between /r/ and /l/, identification for each stimulus varies depending upon the stimulus range but the global identification patterns by Japanese listeners always tend to be kept similar regardless of the stimuli presented in experiment sessions. However, for the A.E. listeners the judgement between /r/ and /l/ appears rather absolute and identification for each stimulus is independent of the stimulus range.

In regard to /w/ identification, it is noted that the stimuli categorized into /w/ for Japanese listeners are included in both the /r/-side and /l/-side conditions. In contrast with /r/ and /l/ identification, /w/ identification seemed rather stable and was hardly affected by the stimulus range. This tendency suggests that Japanese listeners make an absolute judgement between /w/ and non-/w/ and a relative judgement between /r/ and /l/ among the stimuli categorized into non-/w/.

In general, the relationship between the stimulus range and judgement is a salient case of contextual effects, which have been important issues in psychophysics (e.g. Algom and Marks, 1990) after the adaptation-level theory by Helson (1964). In the speech perception area, the psychophysical contextual effect (not the vocalic context) has been studied in vowel discrimination (e.g. Repp & Crowder, 1990). In contrast, in experiments designed to measure the subjects' phonemic categories, the psychophysical contextual effect has not been argued so far. This might be because the phonemes are perceived categorically and judgement is hardly affected by the psychophysical context. In the non-native phoneme perception, this contextual effect is thought to be of great significance. In previous studies, the experiments appeared to be designed to empirically balance the number of stimuli identified as each response category for native listeners, and the main issue for those studies was the location of phonemic boundaries. However, such a large effect of the stimulus range for Japanese listeners' identification observed in the present results leads us to conclude that, in future non-native phoneme perception experiments, it is important to discuss the identification pattern for each phoneme on a given acoustic space in addition to the absolute location of phoneme boundaries.

-20-

## Conclusion

In this paper the effects of some variables which are associated with experiment tasks on the perception of AE /r/ and /l/ for native speakers of Japanese were investigated. Through several experiments, these effects were revealed to have striking influences on the experiment results. By investigating the effects, the following findings about the perceptual characteristics of A.E. /r,l/ for Japanese listeners were obtained:

(1) Japanese listeners identify some of the stimuli on the /r-l/ continuum as /w/.

(2) Japanese listeners have large individual differences in perceiving A.E. /r,l/. However, by classifying the Japanese listeners according to their identification ability of naturally spoken /r,l,w/, Japanese listeners' perceptual mode for the /r,l/ is clarified.

(3) Japanese listeners with high ability to identify naturally spoken /r, l, w/ perceive the synthetic /r-l/ stimulus series categorically. However, these skilled Japanese listeners still have three categories, /r/, /w/ and /l/, on the series.

(4) The stimulus range has a significant effect on the identification results for Japanese listeners. The /r/ and /l/ identification judgement. is strongly affected by the stimulus range, but the /w/ perception is far less affected. When the stimulus range presented to the listeners in one experiment session is narrowed, Japanese listeners' response for each stimulus change due to a relative judgement between /r/ and /l/.

## REFERENCES

Cochrane, R. M. (1980). The acquisition of /r/ and /l/ by Japanese children and adults learning English as a second language. *J. Multilingual and Multicultural Development*, 1, 331-360.

Cyphers, D.S. (1985). Spire: a speech research tool. MS Degree Thesis MIT.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia*, 9, 317-323.

Helson, H. (1964). *Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior* (Harper & Row, New York).

Liberman, A.M., Miyawaki, K., Jenkins, J.J., & Fujimura, O. (1973). Cross-Language Study of the Perception of the F3 Cue for [r] vs. [l] in speech-and nonspeech- like patterns. *Journal of Acoustcal Society of Japan*, 29, 315.

Lisker, L. (1957) Minimal cues for separating /w,r,l,y/ in intervocalic position. *Word* 13, 256-267.

Mann, V.A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r". *Cognition*, 24, 169-196.

MacKain, K.S., Best, C.T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Appl. Psycholinguist*, 2, 369-390.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Fujimura, O., and Jenkins, J. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception &. Psychophysics*, 18, 331-340.

Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetcs*, 9, 283-303.

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

Polka, L., & Strange, W. (1985). Perceptual equivalence of acoustic cues that differentiate /r/ and /l/. *Journal of Acoustical Society of America*, 78, 1187-1197.

Price, P.J. (1980) A cross-linguistic study of flaps in Japanese and in American English. Ph.D. Thesis, University of Pennsylvania.

Repp, B.H., & Crowder, R.G. (1990). Stimulus order effects in vowel discrimination. *Journal of Acoustical Society of America*, 88, 2080-2090.

Shimizu, K., and Dantsuji, M. (1983). A study of the perception of /r/ and /l/ in natural and synthetic speech sounds. *Studia Phonologica*, XVII, 1-14.

Strange, W., & Dittmann, S. (1984) Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, 36, 131-145.

Underbakke, M., Polka, L., Gottfried, T.L., & Strange, W. (1988). "Trading relations in the perception of /r/-/l/ by Japanese learners of English." *Journal of Acoustical Society of America*, 84, 90-100.

Yamada, R.A., & Tohkura, Y. (1989). Perceptiual

characteristics of English syllable-initial /r, l/ for Japanese listeners. *Journal of Acoustical Society of America,.*Suppl. 1, NN13.

Yamada, R.A., & Tohkura, Y. (1990). Perception and production of syllable-initial English /r/ and /l/ by native speakers of Japanese. Proceedings of ICSLP90, 757-760.

Zue, V.W., and Cyphers, D.S. (1985). The MIT Spire system. in Proceedings of Speech Tech (Media Dimensions, New York), pp.277-279.