Perception & Psychophysics
1975, Vol. 18 (5), 331-340

An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English

KUNIKO MIYAWAKI
University of Tokyo, Tokyo, Japan

WINIFRED STRANGE and ROBERT VERBRUGGE
University of Minnesota, Minneapolis, Minnesota 55455

ALVIN M. LIBERMAN
Haskins Laboratories, New Haven, Connecticut 06510

JAMES J. JENKINS
University of Minnesota, Minneapolis, Minnesota 55455

and

OSAMU FUJIMURA
University of Tokyo, Tokyo, Japan

To test the effect of linguistic experience on the perception of a cue that is known to be effective in distinguishing between [r] and [l] in English, 21 Japanese and 39 American adults were tested on discrimination of a set of synthetic speech-like stimuli. The 13 "speech" stimuli in this set varied in the initial stationary frequency of the third formant (F3) and its subsequent transition into the vowel over a range sufficient to produce the perception of [r*a*] and [l*a*] for American subjects and to produce [r*a*] (which is not in phonemic contrast to [l*a*]) for Japanese subjects. Discrimination tests of a comparable set of stimuli consisting of the isolated F3 components provided a "nonspeech" control. For Americans, the discrimination of the speech stimuli was nearly categorical, i.e., comparison pairs which were identified as different phonemes were discriminated with high accuracy, while pairs which were identified as the same phoneme were discriminated relatively poorly. In comparison, discrimination of speech stimuli by Japanese subjects was only slightly better than chance for all comparison pairs. Performance on nonspeech stimuli, however, was virtually identical for Japanese and American subjects; both groups showed highly accurate discrimination of all comparison pairs. These results suggest that the effect of linguistic experience is specific to perception in the "speech mode."

One way to examine the effect of linguistic experience on the perception of speech is to compare the discrimination of phonetic segments by two groups of speakers: one group speaks a language in which the segments under study are functionally distinctive, the other does not. In that circumstance, a difference in the ability to discriminate can be attributed to the linguistic use of the distinction in the one case and lack of such linguistic use in the other.

Two cross-language studies of the kind described above are relevant to the experiment reported in this paper. One study deals with vowels (Stevens, Liberman, Studdert-Kennedy, & Ohman, 1969) and one deals with the voicing distinction in stops (Abramson & Lisker, 1970). In the vowel study, linguistic experience appeared to have no effect. Discrimination of synthetic vowels was the same for Swedish and American listeners, though the vowels were phonemically distinct for the one group and not for the other. The voicing distinction in stops yielded an opposite result. More accurate discrimination was observed at those positions on the stimulus continuum that corresponded to the different positions of the voicing boundary for the language spoken by the subjects (Thai or English).

This research was supported by grants to the following: Haskins Laboratories, National Institute of Child Health and Human Development (HD-01994); the Center for Research in Human Learning, National Institute of Child Health and Human Development (HD-01136) and the National Science Foundation (GB 35703X); and James J. Jenkins, National Institute of Mental Health (MH-21153). Dr. Liberman received support from the Japan Society for the Promotion of Science for his contribution to this research. The authors wish to express their appreciation to Thomas Edman for testing the second group of American subjects and to Arthur Abramson and Leigh Lisker for their assistance in constructing the speech stimuli. Robert Verbrugge is now at the Human Performance Center, University of Michigan, Ann Arbor; Osamu Fujimura is now the Head of the Linguistics and Speech Analysis Department, Bell Laboratories, Murray Hill, New Jersey. Requests for reprints should be sent to Winifred Strange, Center for Research in Human Learning, University of Minnesota, Minneapolis, Minnesota 55455.

The difference in discriminability obtained with vowels and stops may be related to articulatory, acoustic, and perceptual differences between these two classes of sounds. For the stops, the articulatory gestures are relatively rapid movements to and from closures of the vocal tract. For the vowels, the movements are slower and the vocal tract is more nearly open. The acoustic cues for the stops are, correspondingly, characterized by rapid changes in amplitude and frequency within a relatively short interval (Delattre, Liberman, & Cooper, 1955), while the cues for the vowels can be (and were in the experiment referred to above) associated with steady-state signals of longer durations (Fry, Abramson, Eimas, & Liberman, 1962). It may also be relevant that the cues for the stops are complexly encoded in the sound stream in the sense that they are merged on the same acoustic parameter with cues for succeeding (or preceding) segments, while in the case of vowels there can be (and were in the experiment referred to above) stretches of sound that carry cues for only one (vowel) segment (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). In the perceptual domain, two differences between vowels and stops have been found. First, in the comparison with steady-state vowels, stops show a greater tendency toward categorical perception (Fry et al., 1962; Liberman, Harris, Hoffman, & Griffith, 1957; Pisoni, 1973, 1975; Stevens et al., 1969; Vinegrad, 1972; Fujisaki & Kawashima, Note 1). Second, stops yield a larger right-ear advantage in dichotic listening tests, presumably due to a greater reliance on the left-hemisphere processing (Shankweiler & Studdert-Kennedy, 1967).

The experiment reported here is intended to investigate the effect of linguistic usage on the perception of yet another class of phones, the liquids [r] and [l]. There are several reasons why an investigation of these phones is of interest.

First, the perception of [r] and [l] is an obvious choice for a cross-language study of Japanese and English, since the distinction between these phones is phonemic in English but not in Japanese. In syllable-initial position, which is the only context we will be concerned with, [r] and [l] are in minimal contrast in English, as in "red" vs. "led." The articulation of these phones is hard to characterize because reasonably stable acoustic results can be achieved by a variety of articulatory strategies. Typically, however, the English [r] in syllable initial position is articulated with the tongue tip turned up against the post-alveolar region of the hard palate—the lateral palato-lingual contact spreading medially without forming a closure—while the medio-dorsum of the tongue maintains a concave shape (Miyawaki, Note 2). A syllable-initial [l], on the other hand, is articulated with the tongue tip in contact with only the medial portion of the alveolar ridge, forming no palato-lingual contact laterally. In both cases, the voicing continues throughout the articulation (Heffner, 1952; Jones, 1956). Acoustically, a sufficient and important cue for the distinction between [r] and [l] is the initial steady-state and transition of the third formant. For [r], the third formant originates just slightly above the starting frequency of the second-formant transition, while for [l], it starts from a much higher frequency, equal to or even higher than the steady-state frequency of the third formant of the adjoining vowel (O'Connor, Gerstman, Liberman, Delattre, & Cooper, 1957).

In Japanese, [r] and [l] do not constitute a phonemic contrast. The phone that is referred to as a Japanese [r] is typically a loose alveolar stop in initial position or the so-called "flapped-r"—the tongue tip making a very brief contact with the alveolar ridge—in intervocalic position. To an American listener, the Japanese [r] often sounds like [d]. In some cases, the phone is produced with "lateral" articulation, usually with a tendency of retroflexing, and it might sound to an American like an [l] or an [r]. There is no apparent allophonic distribution of [r] and [l] in different contexts (Miyawaki, 1973).

Acoustically, in contrast to the American liquids, the Japanese [r] tends to have little or no initial steady state. The starting point and the transition of the third formant seem to vary unsystematically over a range of values sufficient to distinguish the American [r] and [l], although it appears that in most cases F3 assumes relatively lower values more like the American [r] than [l]. It is important to note that both English [l] and English [r] are perceived by Japanese speakers as the same consonant, their /r/, and there is no other English consonant that shares this characteristic in word-initial position.

Second, a cross-language study of [r] and [l] is of interest because these phones form an articulatory manner class (liquids) that is not only different from the two classes previously studied (stops and vowels), but in some ways intermediate between them. Thus, the liquids are not articulated with the complete closure of the vocal tract that characterizes the stops, nor with the open vocal tract of the vowels. Also, their articulation is not so fast as that of the stops, nor so slow as the vowels. As for their acoustic characteristics, liquids in initial position typically have short steady-state portions with an appreciable amount of sound energy preceding the formant transitions, while stop consonants have only transitions with little or no sound energy preceding them and vowels can be produced entirely with steady-state formants.

From the standpoint of distribution, liquids in English are intermediate between vowels and stops in terms of their phonotactic property, viz, vowel affinity (Fujimura, 1975). In Japanese /r/, the only liquid, behaves as a consonant from a functional point of view.

A third reason for a cross-language study of [r] and [l] is that it is quite easy to isolate the distinguishing acoustic cue for these phones. Thus, we can determine how the two language groups discriminate this cue, not only in a speech context, but also in isolation, when it is not perceived as speech. On this basis, we can judge whether the effect of linguistic experience, if any, is limited in the perceptual domain to speech or whether, alternatively, it extends to nonlinguistic auditory processes.

For these reasons, it is interesting to examine any difference between the Americans and the Japanese in the pattern of discrimination of this class of sounds. In addition, our study has a final point of interest in that it provides data relevant to some questions about tendencies toward continuous and categorical perception. So far, these questions have been asked about vowels and stops, but not about the liquids, the class of phones that we will study here.

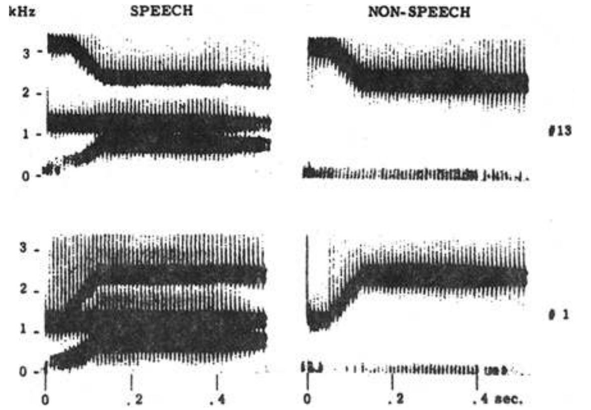

Figure 1. Spectrograms of speech and nonspeech stimuli—[la] upper and [ra] lower.

METHOD

Stimulus Materials

A series of 15 three-formant speech patterns was generated with the parallel-resonance synthesizer at Haskins Laboratories. The structure of the third formant (F3) varied over a range sufficient to produce perception of the consonant-vowel syllables, [ra] and [la]. The stimuli consisted of three contiguous parts: an initial 50-msec steady-state portion, a 75-msec transition of the formant frequencies between the initial and final steady states, and a final steady-state vowel portion of 375 msec duration.

The 15 stimuli differed only in the frequency values of the third formant within the initial steady-state and transition portions. Initial steady-state values of F3 varied in 15 roughly equal steps from 1,362 to 3,698 Hz. Transitions of the formant frequency were linear functions of time from each initial steady-state value to the common steady-state value of 2,525 Hz for the vowel.

Frequency values of the first formant (F1) and second formant (F2) were identical for all 15 stimuli. F1 was set to a frequency of 311 Hz during the initial steady state, then was changed linearly during the transition to a frequency of 769 Hz for the vowel. F2 remained at a constant frequency of 1,232 Hz throughout the entire syllable.

Within the final 400 msec of each syllable, amplitudes of F2 and F3 were set to -3 and -15 dB relative to F1, respectively. The amplitude of F1 at its onset was -12 dB relative to its final value and increased as a decelerated function over the first 100 msec of the syllable. F2 amplitude over the first 100 msec was -3 dB relative to its final steady-state value. F3 amplitude remained constant throughout the syllable. Superimposed on these amplitude values was an overall amplitude contour on the first 50 msec of the syllable, which began 15 dB below its final value and rose linearly. The syllable had a gradually falling fundamental frequency contour from 114 to 96 Hz.

For comparison with the speech patterns, a set of nonspeech stimuli was generated which consisted of the 15 different F3 patterns in isolation. The stimuli were generated by setting the F1 and F2 amplitudes to zero throughout the syllable, so the resultant F3 patterns may be considered as acoustically identical to the F3 patterns *within* the speech stimuli. These stimuli did not sound like speech, but rather light high-pitched glissandos followed by a steady pitch. Figure 1 illustrates the two pairs of examples at nearly extreme F3 values. Stimuli are numbered consecutively with the lowest F3 initial value labeled "1."

Two types of tests were constructed from the speech stimulus set: an identification test and an oddity discrimination test. The former was constructed by recording the speech patterns one at a time in random order with a 1.5-sec interstimulus interval. Each stimulus appeared 10 times for a total of 150 trials. Trials occurred in blocks of 30 with a 5-sec interval between blocks. This test will be referred to as the identification test.

For oddity discrimination tests, Stimuli 14 and 15 were deleted.[1] Ten pairs of stimuli were selected such that each pair (AB) differed by three steps (i.e., 1-4, 2-5, . . . , 10-13). For each pair, triads were constructed by duplicating one stimulus of the pair; all six permutations (AAB, ABA, BAA, ABB, BAB, BBA) were generated. Thus, the oddity test consisted of 60 triads, six permutations for each of 10 comparison pairs. The triads were recorded in random order with a 1-sec interstimulus interval and a 3-sec intertriad interval. Two such randomizations were recorded on audio tape for presentation to subjects. These will be referred to as Speech Tests 1 and 2, respectively.

Oddity discrimination tests of the nonspeech F3 patterns were constructed in the same way as the speech tests. This was accomplished by substituting the corresponding F3 stimulus for each speech stimulus. Thus, the pairing of stimuli and order of triads was the same as that in the speech tests. The two randomizations of 60 triads each will be referred to as Nonspeech Tests 1 and 2, respectively.

For purposes of familiarizing the subjects with the stimuli, two additional recordings were generated. The speech familiarization tape contained the following sequences: the speech stimulus set presented in succession from No. 1 to No. 15, the set repeated in reverse order, the patterns presented in random order with each stimulus occurring two times, and Stimuli 4 and 10 presented five times each. (The latter were judged to be the "best" tokens of [ra] and [la] by an experienced phonetician.) A nonspeech familiarization tape included a set of randomly presented F3 patterns, with each stimulus occurring twice and the two nonspeech patterns, Stimuli 4 and 10, recorded five times each. All experimental materials were then rerecorded and the second-generation recordings used in the experiment.

Subjects

Subjects were 39 native speakers of American English and 21 native speakers of Japanese. The American subjects, undergraduate students at the University of Minnesota, were tested at different times and under somewhat different procedures. Nineteen of the American subjects were students in introductory psychology classes offered during the summer session; this group is referred to as Americans I. The remaining 20 subjects were students in introductory psychology classes during the regular fall quarter; they are referred to as Americans II. The students received monetary reimbursement and extra credit points toward their course grade. All subjects reported having normal hearing.

The Japanese subjects were students and staff at the University of Tokyo. Every member of the group had received at least 10 years of formal English language training. Two subjects had lived abroad from the age of 12 years to 16 years. K.M. attended English-speaking schools in Ceylon; S.A. attended school in Germany. Data obtained for these subjects are discussed separately in the results. (It should be understood that English teaching in Japan usually tends to stress reading and writing; conversational English is not emphasized.) Subjects were paid for their participation in the experiment. All subjects reported having normal hearing.

Procedure

The experimental procedures were basically the same for all three groups of subjects, Americans I and II and Japanese. This section describes the basic procedure; in Appendix A, detailed procedural information for each group is given. The experiment consisted of three parts: familiarization, discrimination tests, and identification tests (for the Americans only).

Familiarization. The procedure for speech familiarization was as follows: Subjects listened to the ordered series without being told what speech sounds were represented. They were then informed that the stimuli were several instances of the English syllables [ra] and [la], and were presented the random series. Finally, they heard the five repetitions of Stimuli 4 and 10, which were described by the experimenter as the "best" instances of the two syllables.

For nonspeech familiarization, subjects were told that the stimuli were "related" to the speech sounds, but would probably not sound like speech. They heard the random series and were asked to describe them as best they could. They were then presented the repetitions of Stimuli 4 and 10 and asked if they could tell them apart easily.

Discrimination. Subjects were told that they would hear triplets of sounds in which two were always identical and one different, and they were to indicate on printed score sheets whether the different one occurred first, second, or third in the triad. They were instructed to respond on every triad, even if they had to guess. They were told they could use any criterion to make the difference judgment.

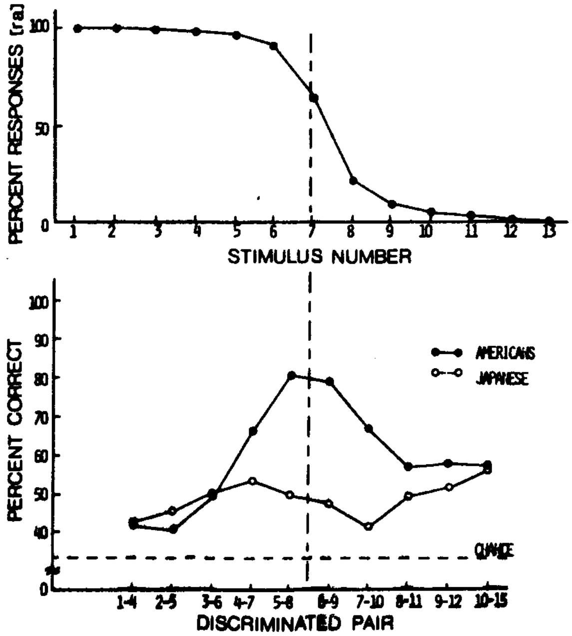

Figure 2. Upper graph: Pooled identification of speech stimuli by Americans. Lower graph: Pooled discrimination by Americans (closed circles) and Japanese (open circles).

All subjects completed two repetitions of Tests 1 and 2 (240 trials) for each stimulus set (speech and nonspeech) on the first day of testing. On Day 2, subjects were reminded of the procedure and again completed two repetitions of Tests 1 and 2 for each stimulus series. Thus, subjects completed a total of 480 trials, 48 judgments for each AB comparison pair, for both the speech stimuli and the nonspeech stimuli.

Identification (American subjects only). On the third day of testing, the American subjects were instructed to listen to each speech stimulus, and mark down on printed score sheets whether the syllable began with an "r" or an "l." They were told to identify every stimulus and were limited to the two response alternatives. They completed two repetitions of the identification test for a total of 300 trials, comprising 20 judgments for each of 15 stimuli.

RESULTS

Comparison of American and Japanese Discrimination of the [r-l] Contrast

Most relevant to the purposes of this study are the data, shown in the lower half of Figure 2, on the discrimination of [r-l] by the two language groups. But before comparing those data, we should note, in the upper half of the figure, the results of the identification test which was given only to the Americans. There, where the percent of "r" responses is plotted for each of the 13 stimuli[2] of the "speech" series, we see that the American subjects did, in fact, divide the stimuli rather neatly into the two phoneme categories that our synthetic patterns were designed to embrace and, further, that the boundary between the categories is in the neighborhood of Stimulus 7.

Looking now at the lower graph, where percent correct in the discrimination task is plotted against the stimulus pair being tested, we see immediately that the performance by the two groups was markedly different. The American subjects discriminated well between those stimuli that were drawn from different phoneme categories, that is, those that straddle or include the one (Stimulus 7) closest to the boundary between [r] and [l]. However, they discriminated rather poorly those that were given the same category assignment in the identification test. The Japanese, on the other hand, showed no such increase in discrimination at the phoneme boundary; for the stimuli that lay within a phoneme class, their discrimination was close to that of the Americans.

Examination of the discrimination functions for individual subjects revealed that 34 of the 39 American subjects were highly accurate in discriminating pairs whose members were labeled as different phonemes (especially Pairs 5-8 and 6-9). Discrimination of pairs whose members were labeled as the same phoneme was considerably less accurate, although still above the 33% chance level. (A discussion of differences in discrimination data for the Americans I and Americans II groups is included in Appendix A.)

Examination of the data for the Japanese subjects, however, found little evidence of such accurate discrimination. Only three subjects showed distinct peaks in discrimination in the vicinity of the phoneme boundary indicated in the American identification data. One of these subjects, S.A. (23 years) lived in Germany between the ages of 12 and 16 and is a fluent speaker of German. Subject K.M. (23 years) lived in Ceylon between the ages of 12 and 16 and is a fluent speaker of English. Subject M.S. (43 years) received regular English training in Japan with an emphasis on reading and writing, starting at the age of 12. (Discrimination data for each of these subjects and for the remainder of the sample are given in Appendix B.)

Discrimination of the [r-l] Cue in Isolation [Nonspeech]—Americans and Japanese

As we pointed out in the introduction, it was possible in this experiment to compare the discrimination of the relevant acoustic cue (the F3 transition in this case) under two conditions: when it is the only basis for the perceived distinction (if any) between the speech sounds, and when it is presented in isolation and does not sound like speech at all. This comparison is of some interest even in the study of speech-sound discrimination that does not make a cross-language comparison. Thus, given an increase in the speech-sound discrimination at the phoneme boundary, as there was for the American subjects in our experiment, the nonspeech discrimination function helps us to know whether the discrimination peak is part of our general auditory perception or whether, alternatively, it is somehow peculiar to the speech context—that is, to perception in the speech mode. In the case of a cross-language comparison, the nonspeech discrimination data are potentially even more interesting. Having found a difference in speech-sound discrimination between the two language groups, as we did in our experiment, we can see in the nonspeech data where the difference might lie. If we assume, as we do, that the difference between the language groups reflects an effect of linguistic experience, then we can look to the nonspeech functions to help us decide whether that effect was at the auditory level or whether, alternatively, it was somehow specific to perception in the speech mode. If the effect were on auditory perception quite generally, we should expect the two groups to differ similarly on both the speech and nonspeech discrimination. Alternatively, if the effect is specific to the speech mode, we should expect the two groups to discriminate the nonspeech stimuli in similar fashion, however much they might differ in discrimination of the speech sounds. In all cases, a result that tends to put the effect in the speech mode could, of course, be interpreted alternatively as a purely auditory interaction between the cue and the constant acoustic context to which it is always added in the speech patterns. But such an interpretation is empty unless one can make sense of it in terms of what is known, on other grounds, about auditory perception.

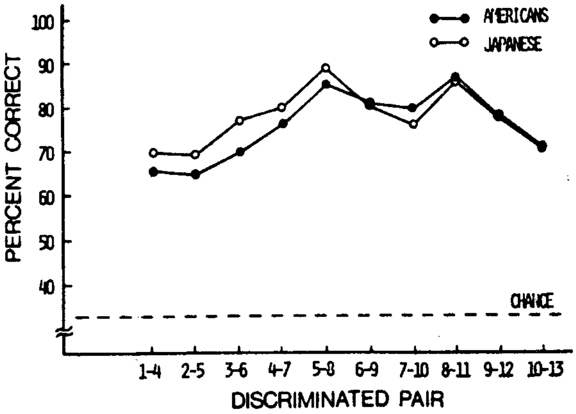

Figure 3. Pooled discrimination of nonspeech stimuli by Americans (closed circles) and Japanese (open circles).

In the case of cross-language comparisons, the results of the relevant nonspeech discrimination provide a useful check on the procedures as well. If there are no differences between groups for the nonspeech stimuli, we can be more confident that the differences in discrimination of the speech sounds were not due to some uncontrolled methodological factors in the conduct of the experiment.

The pooled data for discrimination of the F3 cue in isolation are shown in Figure 3 for both the American and Japanese subjects. Looking first at the results for the American subjects, we see that the shape of the function is quite different from that obtained when the same acoustic variable was perceived in a speech context where it cued the distinction between [r] and [l]. The difference between speech and nonspeech discrimination functions is similar to the finding of an earlier experiment on place distinctions in voiced stops (Mattingly, Liberman, Syrdal, & Halwes, 1971), where the relevant cue was tested in and out of speech context. In both experiments, it is apparent that the discrimination peak obtained in the speech context is peculiar to that context and is not, more generally, characteristic of the way we perceive the relevant acoustic variable.

But it is the nonspeech discrimination function obtained with the Japanese subjects that is of particular interest. We see very clearly that the Japanese do not differ from the Americans on any of the comparison pairs. The nonspeech discrimination functions are virtually identical for the two groups of subjects. We conclude, then, that the differences between the groups on the speech stimuli are a function of processes specific to the perception of speech, or at least speech-like stimuli such as ours, as opposed to stimuli that cannot be identified as phonological units. Also, the results suggest that the procedures for testing the two groups were comparable, and that the differences on speech discrimination cannot be attributed to uncontrolled methodological factors.

It is interesting to note that, for both groups, discrimination for all nonspeech comparison pairs is quite accurate (ranging from 66% to 89%). That is, both Japanese and Americans were able to discriminate differences in F3 patterns when they were presented in isolation. This suggests that the poor discrimination by Japanese for all speech comparison pairs and by Americans for within-category pairs is not due to the acoustic differences per se being indiscriminable, but rather has something to do with the phonemic identity of the speech patterns which contain these F3 patterns. However, two factors may have contributed to the relatively better discrimination of the isolated formants: the F3 patterns were presented at a much higher amplitude than the F3 components within the speech patterns, and it is possible that the lower formants in the speech patterns masked the F3 component to some extent. More research that measures the effects of intensity and masking on the perception of nonspeech is needed to explore these factors.

In both the Japanese and American nonspeech functions, two comparison pairs appear to be discriminated slightly better than the others. It is interesting to note that each of these pairs contains Stimulus 8 (5-8 and 8-11). Stimulus 8 is unique in that its F3 does not contain a frequency transition. In other words, this pattern is a steady state in contrast to Stimuli 1 through 7, which contain rising transitions, and Stimuli 9 through 13, which have falling transitions. It appears that subjects were able to distinguish between "no transition" vs. "some transition" slightly better than between transitions with different slopes.

Categorical Perception of [r] and [l]

We may now turn to a consideration of the relation between the identification and discrimination functions obtained for the stimuli presented to the American subjects. A reexamination of Figure 2

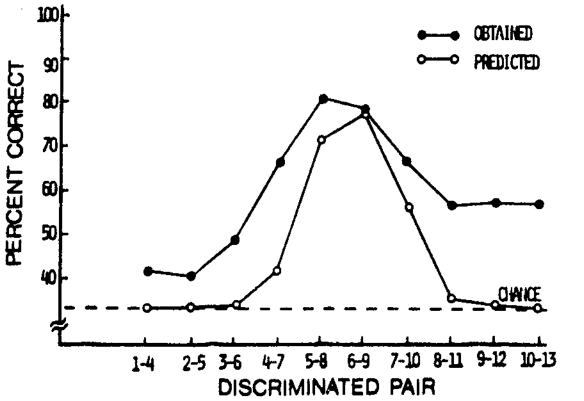

Figure 4. Obtained (closed circles) and predicted (open circles) functions for pooled discrimination by Americans. (See text for explanation of predicted function.)

shows a striking correspondence between the sharp change in identification of the stimuli as [r] and [l] and the peak in the discrimination function. The close relation between identification and discrimination is similar to that found for stop consonants, and has been referred to as "categorical perception" (Liberman et al., 1957). In contrast, the correlation between identification and discrimination does not always hold for other speech sounds, such as steady-state vowels (Fry et al., 1962).

A strong test for the presence of categorical perception may be made by predicting the shape of the discrimination function. If one makes the extreme assumption that subjects are able to discriminate speech stimuli only when they label them differently, it is possible to predict their discrimination functions from their identification performance. Each of the two stimuli in an oddity triad has a probability of being labeled as "r" or as "l," as determined in the identification test. From these data, it is possible to calculate the probability of the triad being heard as each of the possible sequences of the two phonemes. Only some of these perceived sequences will result in correct choices of the odd member, and those probabilities may be summed for each stimulus order. The probability of correct discrimination for a stimulus pair will be an average of the probabilities for the six possible orders. If Pr is the probability of one member being heard as "r" and Pr′ is the probability of the other member being heard as "r," then the average probability of correct discrimination is found to be $P_{corr} = [1 + 2(Pr - Pr')^2]/3$.

The predicted discrimination function for the pooled data is shown in Figure 4. As is typical of such functions, the location and extent of the discrimination peak is fairly accurately predicted, while within-category discrimination is underestimated. This suggests that even though subjects labeled the stimuli as the same phoneme, they were able to discriminate intraphonemic variants to some extent. This point also conforms with the observation that the discrimination by Japanese subjects, even though poor, was better than chance.

DISCUSSION

Returning now to the questions that prompted this study, we may conclude that rather clear answers have been obtained. We note, first, that familiarity with the [r-l] distinction obviously has a major impact on the ability to make correct discriminations in an oddity test. In this respect, the findings are overwhelming. American subjects show a peak of highly accurate discrimination at the point where stimuli from different phonetic classes are being contrasted. Japanese subjects show no such accurate discrimination at any point along the stimulus dimension. Moreover, the results are consistent for individuals, not merely characteristic of group averages. Of the 39 American subjects, 34 showed clear discrimination peaks in their individual protocols, while only 3 of the 21 Japanese subjects did. Furthermore, two of the three Japanese subjects who did show discrimination peaks learned languages with the relevant liquid contrast as early adolescents. It is reasonable to conclude, therefore, that considerable experience with the linguistic distinction is prerequisite to successful performance on the discrimination test with synthetic speech stimuli such as we have employed. Also, since all Japanese subjects had studied English extensively, it is tempting to hazard the hypothesis that discrimination requires effective *phonetic* experience at a relatively early age, say early adolescence.

The finding that Japanese subjects cannot for the most part discriminate [r] and [l] over this range of synthetic stimuli confirms the observation of Goto (1971) that native Japanese speakers who are highly fluent in English cannot perceive the distinction between [r] and [l] produced by other speakers (both Japanese and American). Even more interesting, Goto reports that his subjects cannot distinguish reliably *their own* tokens of [r] and [l], even when American speakers judge the tokens to be appropriate instances of the two phones. Thus, the lack of discrimination of synthetic stimuli covering a range of variation is in harmony with what is known about the properties of perception of real speech in normal contexts. This does not mean, however, that training after adolescence does not help at all. In fact, some of our Japanese subjects may not fail in discriminating natural utterances of [rɑ] and [lɑ]. The stimuli compared on the discrimination test are undoubtedly much more similar to each other than optimal instances of the phonemes. Also, there may be other cues for the distinction in natural utterances which some Japanese subjects may depend on more heavily than do Americans.

Second, it is apparent that the difference in discrimination performance is limited to the speech-like condition. No difference appeared between the American and Japanese groups in the discrimination of the acoustic cue for [r] vs. [l] when it was presented in isolation. This finding is consistent with the argument that speech perception is a special mode of auditory perception that is accomplished in quite a different manner from general auditory perception. In all cases, such an argument must, as we said earlier, leave room for the fact that even though the acoustic cue being discriminated was always the only variable, it was presented by itself in the nonspeech case, while in the speech case, it was added to a fixed auditory pattern, thus creating the possibility of an auditory interaction. In this experiment, comparison of the speech and nonspeech discrimination functions must also take into account the differences in amplitude of the F3-transition cue in the two cases and the possibility that in the speech context the F3 cue was to some degree masked by the constant F1 and F2.

Finally, the study yielded results concerning the "categorical perception" of liquids in initial position in English. While American listeners make more correct discriminations of stimuli than would be predicted from a strict categorical perception hypothesis, the match between predicted and obtained discrimination functions resembles more closely that obtained for stop consonants than that obtained for vowels (Fry et al., 1962; Liberman et al., 1957).

Since the present study was performed, Eimas (in press) has studied how 2- and 3-month-old infants perceive the stimuli utilized in this study. Using a habituation paradigm, he tested the discrimination of speech stimuli both within and between the [l] and [r] categories. The infant discriminations were remarkably parallel to those we obtained with American adults. Infants who were habituated to stimuli from one side of the adult boundary and then switched to stimuli from the other side of the boundary showed impressive recovery from habituation. Within-class shifts of stimuli produced much less recovery. However, shifts within the [l] category produced greater recovery than shifts within the [r] category, reflecting the tendency shown by American subjects to discriminate within the [l] category better than within the [r] category. Infants tested with comparable shifts in the nonspeech stimuli (F3 alone) failed to show significantly different recovery from habituation in all conditions. Thus, the infant data are parallel in all respects to the American adult data that we have presented here. Obviously, it would be of great interest now to follow the course of habituation-discrimination in Japanese children.

APPENDIX A

Specific Procedures

Americans I. Subjects were assigned to one of two counterbalanced conditions according to convenience in scheduling test sessions. Ten subjects were tested in the speech-first condition, nine subjects in the nonspeech-first condition. During an initial session, all subjects were given familiarization on both stimulus series. Discrimination testing began the following day, after subjects were again familiarized with the task by listening to 10 triads of the first test without responding. The procedures in speech-first and nonspeech-first conditions were identical except for the order of presentation of the stimulus series for discrimination. For the speech-first subjects, the order for the first day was as follows: Speech Tests 1 and 2, Nonspeech Tests 1 and 2, Speech Tests 1 and 2, Nonspeech Tests 1 and 2. For nonspeech-first subjects, the order was reversed, i.e., Nonspeech Tests 1 and 2, Speech Tests 1 and 2, etc. The order of presentation on the second day of discrimination testing was the same as for Day 1 for each group. Both groups completed identification tests on the third day.

Subjects were tested in small groups (from one to four) in sessions which lasted about 2 h. Testing was conducted in a quiet experimental room. Stimuli were reproduced on a Crown CS 822 tape recorder and presented to subjects binaurally over Koss Pro-600A earphones. Signal levels were monitored with a Heathkit IM21 AC VTVM at the output to the earphones. Both speech and nonspeech stimuli were presented at a sound level approximately

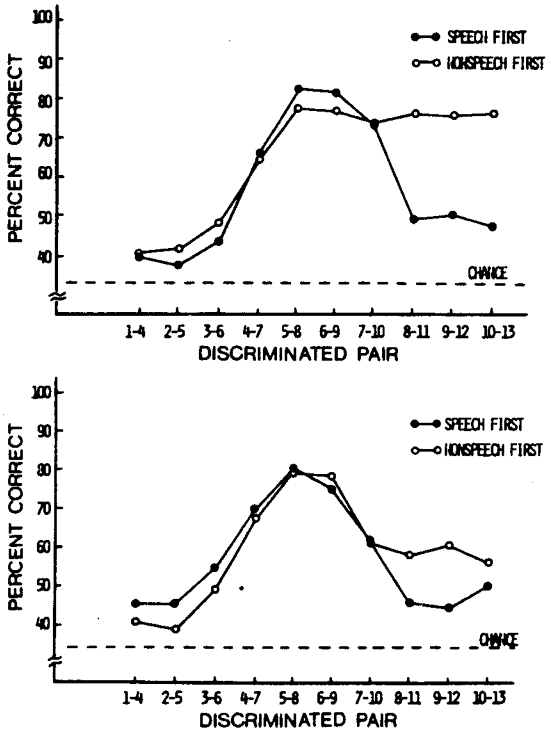

Figure 5. Pooled discrimination of speech stimuli for Americans I (upper graph) and Americans II (lower graph), speech-first and nonspeech-first conditions.

70 dB above threshold. The playback amplitude for the nonspeech stimuli was increased to make the isolated F3 patterns equal in peak amplitude to the three-formant speech patterns. Thus, the amplitude of the isolated F3 patterns was far greater than the amplitude of the comparable F3 components *within* the three-formant patterns. However, the *total* signal for each set of stimuli was equal in amplitude and duration.

Americans II. The 20 subjects were divided into two groups of 10 each according to convenience in scheduling test sessions. The groups were tested in two counterbalanced conditions, speech first and nonspeech first, in a manner similar to the Americans I, except for the following. While the Americans I groups were presented Speech and Nonspeech Tests 1 and 2 alternately within a single session of discrimination testing, the Americans II groups completed two repetitions of Tests 1 and 2 for the first stimulus series before proceeding to the other stimulus series. Thus, on Day 1 the speech-first group completed *two* repetitions of Speech Tests 1 and 2, then completed two repetitions of Nonspeech Tests 1 and 2. The order on Day 2 was identical to that of Day 1 for each group.

Another difference in procedure from the Americans I was in familiarization. For the Americans II groups, familiarization took place for each stimulus series just prior to the first discrimination test in that series. After discrimination tests were completed for the first series (i.e., Tests 1, 2, 1, 2), subjects were given familiarization on the other stimulus series and then proceeded with the tests. No familiarization was given on Day 2; subjects were merely reminded of the test procedure and told what series they would be listening to first.

Subjects were tested in a sound-attenuated experimental room using the same equipment and procedures as for the Americans I. Speech stimuli were presented at a sound level about 70 dB above threshold. Nonspeech stimuli were presented at -5 dB relative to the speech. (The absolute amplitude of the isolated F3 patterns was still far above that of the F3 component within the speech patterns.)

Japanese. All 21 subjects were tested using the Americans II nonspeech-first presentation order. That is, the order on Day 1 was: nonspeech familiarization, Nonspeech Tests 1, 2, 1, 2, speech familiarization, Speech Tests 1, 2, 1, 2. Day 2 was the same as Day 1, except that no familiarization was given.

Subjects were tested individually in a sound-attenuated experimental room. Stimuli were reproduced on a TEAC-type tape recorder and presented to subjects binaurally over Iwatsu DR-305 stereo earphones. Speech stimuli and nonspeech stimuli were output from the tape recorder at about 74 and 76 dB above threshold, respectively. However, each subject adjusted the signal level at his earphones by means of an Ando SAL-20 attenuator, which had a range of 20 dB in 2-dB steps. Attenuation levels that subjects selected as "most comfortable" varied from -2 to -16 dB. The average listening level for speech was approximately 68 dB above threshold; for the nonspeech stimuli, the average was approximately 70 dB above threshold. Thus, as was the case for the American subjects, the isolated F3 patterns were heard at a much higher absolute level than the F3 component within the speech patterns.

Comparison of Results for Speech-First and Nonspeech-First Groups

The upper panel of Figure 5 presents the results of speech discrimination tests for the Americans I speech-first and nonspeech-first groups. The major difference between the groups is their discrimination of comparison pairs within the "l" category. The nonspeech-first subjects were able to discriminate these pairs as accurately as they did the between-category pairs. This could not be predicted from their identification performance, which was very similar to that of the speech-first subjects. An inspection of individual subjects' functions showed that six of the nine subjects produced functions with the elevated within-"l" discrimination. The other three subjects produced functions similar to the speech-first results.

The lower panel of Figure 5 presents the comparable discrimination results for the Americans II speech-first and nonspeech groups. Again, the only difference between the groups is their performance on the within-"l" comparison pairs. However, the difference is much smaller than for the Americans I subjects. The Americans II nonspeech-first subjects showed more nearly "categorical" performance; i.e., in spite of better discrimination of the within-"l" pairs than the within-"r" pairs, performance within either category was still inferior to that for between-category pairs.

Two differences in procedure might have contributed to the different results for Americans I and Americans II nonspeech groups. First, recall that for the Americans I group, familiarization took place in a separate session the day before discrimination testing. Speech familiarization was always given before nonspeech familiarization. Thus, for these subjects, both nonspeech familiarization and testing (Tests 1 and 2) intervened between speech familiarization and the initial speech discrimination tests. In addition, both discrimination testing sessions began with the nonspeech stimuli. These factors apparently caused some "interference" in the speech discrimination task. Subjects may have adopted a "nonspeech" listening strategy, since they were told to use any criteria they could to discriminate the odd member of the triads. Once having established a strategy, the subjects seem to have maintained it throughout testing, since the data for the first and second halves of each day's testing, and the data for Day 1 and Day 2, are very consistent.

In contrast, the Americans II nonspeech-first group received their speech familiarization immediately prior to speech discrimination tests on the first day of testing. This might have helped to establish a "speech" listening strategy for these subjects. None of the 10 subjects produced speech functions like the Americans I nonspeech-group function, although most showed some elevation in discrimination of the within-"l" category pairs.

A second difference in procedure might have contributed in a related manner. The Americans I group switched from nonspeech to speech stimuli twice within a testing session, whereas the

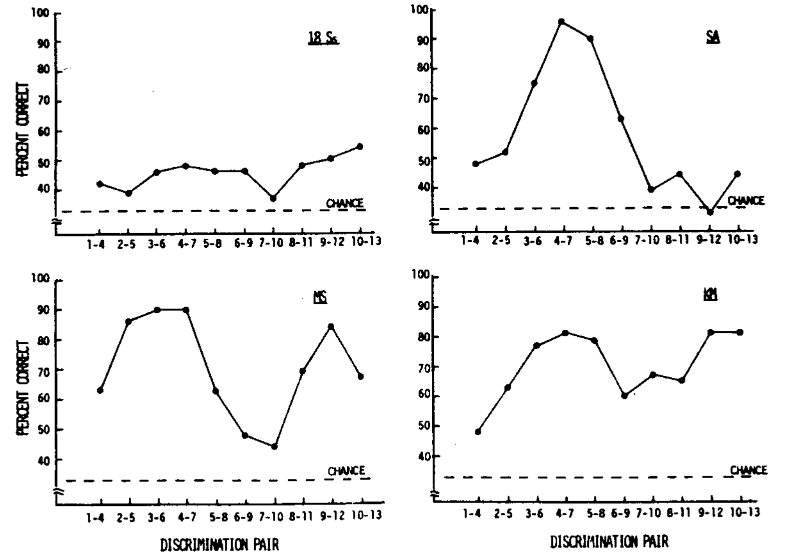

Figure 6. Pooled discrimination of speech stimuli by 18 Japanese subjects (upper left) and individual discrimination functions for three exceptional Japanese subjects. (See text for explanation of exceptional subjects.)

Americans II group completed all nonspeech tests before going on to speech tests. Again, the former procedure may have biased subjects toward a "nonspeech" listening strategy, whereas the latter procedure provided a clear distinction between the two series of stimuli. Additional support for the notion that the high discriminability of the within-"l" category pairs is a nonspeech phenomenon is given by the results of some of the Japanese subjects on the speech discrimination trials. The average curve for Japanese subjects climbs at the "l" end. (See Figure 2.) Most of this is accounted for by five subjects whose individual discrimination functions showed relatively more accurate discrimination of Pairs 8-11, 9-12, and 10-13.

The Japanese pooled data may be compared with the Americans II nonspeech-first group (compare Figures 2 and 5), since the order of presentation is identical for these groups. The difference in discrimination in the region of the Americans' category boundary is clearly present in this comparison.

APPENDIX B

The discrimination data for the three Japanese subjects who showed peaks of high discrimination are given in Figure 6. As adolescents, K.M. and S.A. learned languages employing the phonemic distinction between [r] and [l]. M.S. did not. The upper left panel shows the pooled speech discrimination data for the remaining 18 Japanese subjects.

REFERENCE NOTES

1. Fujisaki, H., & Kawashima, T. *On the modes and mechanisms of speech perception.* Research on Information Processing, Annual Report No. 2, University of Tokyo, Division of Electrical Engineering, Engineering Research Institute, 1969, 67-73.
2. Miyawaki, K. *A preliminary study of American English /r/ by use of dynamic palatography.* Annual Bulletin, Research Institute of Logopedics and Phoniatrics, Faculty of Medicine, University of Tokyo, 1972, **6**, 19-24.

REFERENCES

Abramson, A. S., & Lisker, L. Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the 6th International Congress of Phonetic Sciences* (Prague, 1967). Prague: Academia, 1970. Pp. 569-573.

DeLattre, P. C., Liberman, A. M., & Cooper, F. S. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 1955, **27**, 769-773.

Eimas, P. D. Developmental aspects of speech perception. In R. Held, H. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology*. New York: Springer-Verlag, in press.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. The identification and discrimination of synthetic vowels. *Language and Speech*, 1962, **5**, 171-189.

Fujimura, O. Syllable as a unit of speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1975, **23**, 82-87.

Goto, H. Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia*, 1971, **9**, 317-323.

Heffner, R.-M. S. *General phonetics.* Madison: University of Wisconsin Press, 1952.

Jones, D. *An outline of English phonetics.* Cambridge Mass: Heffer, 1956.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431-461.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, **54**, 358-368.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 1971, **2**, 131-157.

Miyawaki, K. *A study of lingual articulation by use of dynamic palatography.* Masters thesis, Department of Linguistics, University of Tokyo, March 1973.

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

Pisoni, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 1973, **13**, 253-260.

Pisoni, D. B. Auditory short-term memory and vowel perception. *Memory & Cognition*, 1975, **3**, 7-18.

Shankweiler, D. P., & Studdert-Kennedy, M. Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 59-63.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M.,

& Ohman, S. E. G. Cross-language study of vowel perception. *Language and Speech*, 1969, 12, 1-23.

Vinegrad, M. D. A direct magnitude scaling method to investigate categorical versus continuous modes of speech perception. *Language and Speech*, 1972, 15, 114-121.

NOTES

1. These two stimuli had such extreme values of F3 that some pilot subjects heard them as [ra], with a noisy glide superimposed on it.

2. Since Stimuli 14 and 15 were deleted from discrimination, the identification data for these stimuli were not included in the analysis.

(Received for publication April 11, 1975;
revision received August 11, 1975.)