Cross-Language Study of the Perception of the F3 Cue for [r] versus [l] in Speech- and Nonspeech-Like Patterns*

Kuniko Miyawaki, + A. M. Liberman, ++ O. Fujimura, +++ Winifred Strange, ++++ and J. J. Jenkins

Our primary purpose in the experiments reported here was to see how linguistic experience affects the way we perceive the sounds of speech. To do that, we compare the discrimination of a particular phonetic contrast by two groups of people. The one group speaks a language which makes use of the distinction under study; the other group does not. If we find a difference in ability to discriminate, we attribute it to familiarity with the distinction, in the one case, and unfamiliarity in the other.

Several experiments employing this paradigm had been carried out before we undertook the one under discussion. Perhaps the first was a study on vowel discrimination by Stevens, Liberman, Studdert-Kennedy, and Öhman (1969). Using controlled synthetic patterns, these investigators found essentially no effect of linguistic experience: listeners who were unfamiliar with the vowels discriminated among them quite as well as those who used the vowels frequently in their native language. Studies on the discrimination of the voicing distinction in stop consonants have, however, yielded a different result. In that case, Abramson and Lisker (1970) found that listeners who were familiar with the distinction showed an increase in discrimination of the phonetic quality at the phonemic boundary; listeners who were unfamiliar with it did not.

That linguistic experience affects the perception of vowels and stop consonants differently is, perhaps, to be related to other differences between these two classes of sounds. Vowels are represented rather directly in the sound stream; they are, like most nonspeech sounds, perceived in continuous fashion; and, according to the results of some other experiments, they can apparently be processed in either cerebral hemisphere. Stop consonants, on the other hand, are complexly encoded in the sound stream; they are perceived in nearly categorical fashion; and they apparently need to be processed in the left (or language) hemisphere.

*Presented at the XXth International Congress of Psychology, Tokyo, 1972.

+ Department of Linguistics, University of Tokyo, Japan.

++ Haskins Laboratories, New Haven, Conn., and University of Connecticut, Storrs.

+++ Faculty of Medicine, University of Tokyo, Japan.

++++ Center for Research in Human Learning, University of Minnesota, Minneapolis.

[HASKINS LABORATORIES: Status Report on Speech Research SR-33 (1973)]
The experiment reported here deals with the effect of experience on the discrimination of [r] and [l]. There are at least four reasons for studying this distinction. First, these two segments constitute a class of speech sounds that appear to be intermediate between the vowels and stop consonants in regard to the differences just described. Second, the distinction is a natural one for cooperative research by Japanese and American colleagues, since it is familiar to native speakers of English but notoriously unfamiliar to the Japanese. Third, this distinction has an advantage over those studied earlier because it permits us not only to compare our two groups of listeners on discrimination of speech, but also to see how they discriminate the essential acoustic cue when it is presented in isolation and not heard as speech. And, fourth, the data we obtain will help to fill a gap in our knowledge of speech perception, since relatively little experimental work has been done on the discrimination of [r] and [l].

Figure 1 shows the stimuli, the experimental variable, and the two conditions of the experiment. The speech stimuli are shown at the left. There you see spectrograms of synthetic syllables. The one at the bottom is a reasonable approximation to [ra]; the one at the top approximates [la]. These two patterns are identical except for the starting point and transition of the third formant. For [r], at the bottom, the third formant starts low and rises; for [l], at the top, it starts high and falls. In the other stimuli in the series, the third formant started at levels that sampled the range between the extreme [r] and [l] values shown in the figure. There were 13 stimuli in all.

Corresponding samples of the nonspeech stimuli are shown at the right. These nonspeech stimuli are the formants of the speech patterns presented now in isolation. As these third formants constitute the only differences among the speech stimuli, they are alone responsible for the perceived differences among the speech stimuli. But when they are presented alone in the nonspeech condition, they do not sound at all like speech.

Figure 2 shows the results obtained with a group of 12 American listeners, and sets the stage for the comparison with the Japanese. The graph at the top shows what happened when the 13 synthetic speech stimuli were presented singly, in random order, for absolute identification as [ra] or [la]. The graph is plotted as the percent of [r] responses against the position of the third formant, indicated simply by the numbers 1 to 13. We see that the American listeners did sort the stimuli in a fairly consistent way.

The graph at the bottom shows how the same subjects discriminated the speech stimuli. To measure discrimination, we arranged the synthetic stimuli in sequences of three, such that two members of each sequence were identical and one was different. The listeners' task was to say which one was different. Each listener was instructed, for that purpose, to use any differences he could hear. Let me illustrate a result. The first datum point at the left shows what happened when the stimuli in the triplet were, as indicated on the abscissa, Numbers 1 and 4. The point we see on the graph shows that our subjects correctly discriminated these two stimuli 40% of the time. All stimulus combinations were, like those, three steps apart on the stimulus scale.

Looking at the entire graph, we see that discrimination rose to a fairly high peak at a point corresponding approximately to the phonemic boundary, as shown in the identification function at the top. Thus, we have here an approximation to the kind of quantal or categorical perception that has been found
Figure 1: Spectrograms of speech and nonspeech stimuli—[la] upper and [ra] lower.
Figure 2: American results - identification vs. discrimination.
previously for the stops but not the vowels. We must emphasize, however, that a proper comparison with the categorical perception of the stops must await further research and analysis.

Figure 3 permits us to compare discrimination of the speech and nonspeech patterns. On the left we have reproduced the discrimination result, already shown in Figure 2, that was obtained with the speech-like stimuli. On the right are results obtained with the nonspeech stimuli—that is, with the third-formant pattern in isolation. We see that the nonspeech function is quite different from the speech function: the nonspeech function does not show the marked peak at the phoneme boundary and tends to be high throughout. We might infer, then, that the peaking at the phoneme boundary is not so much in the psychoacoustic response to the stimulus as it is a property of perception in the speech mode.

Figure 4 shows the results obtained with 21 Japanese listeners. Looking at the speech discrimination function at the left, we see that it is not peaked as the American function was, and is rather low throughout. The nonspeech function at the right shows that the Japanese subjects discriminated the third formant transition in isolation better than in the speech patterns with lower formants.

I should note that two of the Japanese subjects had spoken English as children. Their results were nevertheless included in the graph of Figure 4. In Figure 5 the data obtained with these two subjects are presented separately. We see that their discrimination functions look very much like those of the Americans.

In Figure 6 we compare the overall results of the American and Japanese subjects. Looking first at the results with the speech-like syllables, at the left, we see that the Americans do indeed discriminate better than the Japanese, especially in the vicinity of the phoneme boundary. But the nonspeech functions at the right are remarkably alike. Thus, the two groups do not differ in their ability to discriminate the essential acoustic pattern when it is presented in isolation, but they do differ when the same pattern serves as a phonetic cue.

I will summarize. The acoustic cue responsible for the perceived distinction between [r] and [l] was, in one set of stimuli, varied in equal steps in synthetic syllables that could be perceived as [r] or [l] plus a vowel. In another set this acoustic characteristic was presented in isolation and did not sound like speech. American listeners, to whom the phonemic distinction is familiar, showed a peak in discrimination similar to those found in the discrimination of the stops. Discrimination of the crucial acoustic pattern in isolation was clearly different. Japanese listeners, who are generally quite unfamiliar with the [r-l] distinction, discriminated it very poorly by comparison with the Americans. Their relative inability to discriminate [r] and [l] cannot be attributed to an inability to discriminate the acoustic patterns, since their discrimination performance was at least equal to that of the Americans when the cue pattern was presented in isolation. Thus, we have here a rather large effect of linguistic experience on perception in the speech mode. In this respect, [r] and [l] are more like the stops than like the vowels. But whether the effect we found here is as large as that obtained with the stops can only be determined by further research.
Figure 3: Overall American (13 subjects).
Figure 4: Overall Japanese (21 subjects).
Figure 5: Two special Japanese subjects.
Figure 6: Overall American vs. overall Japanese.
REFERENCES