Ann.Bull.RILP(1984)No.18, 127-134

PERCEPTION OF [L] AND [R] BY NATIVE SPEAKERS OF JAPANESE: A DISTINCTION BETWEEN ARTICULATORY TRACKING AND PHONETIC CATEGORIZATION

Virginia A. Mann

Abstract

Native speakers of Japanese may be unable to identify the phonemes [l] and [r] in spoken English. Nevertheless, in perceiving English utterances, they, like native speakers of English, unconsciously respond to differences in the vocal tract movements that convey [l] and [r]. One implication is that, preceding a language-specific level of speech perception where utterances are represented in terms of their constituent phonemes, there may exist a universally-shared level of speech perception where listeners use the speech signal to track movements of the vocal tract.

Introduction

What do native speakers of Japanese perceive as they listen to English utterances which contain [l] and [r]? In the absence of considerable experience with spoken English, many Japanese are unable to label, discriminate or produce [l] and [r] in a consistent fashion (Goto, 1971; Miyawaki, Strange, Verbrugge, Liberman, Jenkins and Fujimura, 1975; Mochizuki, 1981), which would seem to suggest that they hear these two speech sounds as one and the same. This study offers evidence that whether or not Japanese subjects can identify [l] and [r] phonetically, they tacitly respond to an articulatory difference between these speech sounds.

Such demonstration comes from a specific context effect in speech perception (for a general discussion of such effects, see Repp, 1982). That effect was observed in an earlier study in which the spoken syllables [al] and [ar] were placed in front of stimuli from along a continuum of synthetic speech syllables ranging from [da] to [ga]. The presence of the preceding syllables caused systematic shifts in the category boundary between [d] and [g]: When the preceding syllable was [al], the boundary was shifted towards more [g] percepts (less [d] percepts, relative to that obtained when the preceding syllable was [ar] (Mann, 1980).

One question to be asked is whether this context effect reflects some property of speech perception, as opposed to purely acoustic interactions. A relevant piece of information has come from use of the phenomenon known as duplex perception (Liberman, Isenberg and Rakerd, 1982; Rand, 1974). In duplex perception, one and the same stimulus is simultaneously heard as speech and as

nonspeech. This situation can be created by dividing synthetic speech syllables along a [da] to [ga] continuum into two parts: a constant base portion and a third formant transition, which, in isolation sounds like a "chirp", but when combined with the base provides the critical cue for the distinction between [da] and [ga]. When base and transition are presented dichotically, the third formant transition is simultaneously perceived in two ways: as speech and nonspeech. It fuses with the base to provides critical support for the perception of [da] or [ga], yet is also heard as a nonspeech "chirp". In a recent experiment, listeners were instructed to attend to one or the other of these percepts, as duplex stimuli from along a [da]-[ga] continuum were preceded by the syllables [al] and [ar]. Under instructions to ignore the speech percepts and attend to the nonspeech chirps, perception was continuous, and neither preceding syllable had any appreciable effect. In contrast, under instructions to label or discriminate stimuli on the basis of the speech percepts [da] and [ga], perception was categorical and the location of the category boundary depended on the nature of the preceding syllable [al] or [ar] (Mann and Liberman, 1983).

Thus the context effect of [al] and [ar] is evident only when acoustic stimuli are perceived as speech. Explanation of the specific effects of [l] and [r] can be had from the view that speech perception mirrors speech production. Here, two related observations are compelling. First, there is the finding that the effect of a preceding consonant on the distinction between [da] and [ga] is not limited to [l] and [r], but extends to the fricatives, [s] and [ʃ] (Mann and Repp, 1981), and that similarities are best described in terms of articulatory properties. Specifically, the speech sounds [l] and [s], which are produced with the tongue relatively forward in the mouth, shift perception away from [da] toward the more backwards [ga], relative to [r] and [ʃ], which are produced with a more retracted tongue posture. Second, the perceptual effects of [l] and [r] find a parallel in speech production, where, owing to coarticulation, the acoustic structure of [da] and [ga] can vary as a function of whether they follow [l] or [r] (Mann, 1980). Both observations support the view that the context effects of [l] and [r], along with many other context effects and trading relations (see, for example, Repp, 1982, Repp, Liberman, Eccardt and Pesetsky, 1978), represent a perceptual sensitivity to the consequences of coarticulation in the speech signal. Human listeners appear to possess some tacit knowledge about articulation and its consequences on the speech signal, and application of that knowledge may be part of what makes speech perception "special" (see, for example: Best, Morrongiello and Robson, 1981; Liberman, 1982; Mann and Liberman, 1983; Repp et al, 1978).

Given this effect and its explanation, we may ask whether preceding [l] and [r] will alter perception of the [da]-[ga] distinction among Japanese listeners who do not have the [l]-[r]

distinction in their native language. English and Japanese share many phonemes, including [d] and [g], but Japanese does not distinguish the liquids [l] and [r]. Its single liquid, [r], more clearly resembles an alveolar flap than English [r]. Absence of early experience with this phonetic contrast renders many native speakers of Japanese unable to distinguish English utterances which contain [l] and [r] in phonetic labeling tasks, discrimination tasks, and in their own productions (Goto, 1971; Miyawaki et al, 1975; Mochizuki, 1981). Yet two- to three-month-old American infants, have been found capable of making some discrimination between utterances that contain [l] and [r] (Eimas, 1975), and this raises certain questions about the role of experience in the development of speech perception.

One explanation of the speech perception abilities of infants vis-a-vis the phonetic difficulties of native speakers of Japanese is that a lack of specific experience has led to a loss of ability to perceive a difference between [l] and [r] (Eimas, 1975). Another, slightly different possibility is that infants may not perceive [l] and [r] as different phonemes, so much as they respond to differences in the vocal tract movements that convey [l] and [r]. That is, they may behave as if they can track articulatory gestures, without being able to categorize them phonetically. If so, lack of experience with the [l]-[r] distinction might lead to an inability to distinguish [l] and [r] phonetically, but not necessarily to a desensitization of the basic ability to track the gestures that transmit [l] and [r]. This possibility can be tested by using the present context effect to ask whether Japanese subjects who cannot phonetically categorize [l] and [r] can nonetheless respond to the underlying vocal tract gestures.

Methods

Subjects

Sixteen college freshmen enrolled in the first semester of a spoken English course at the University of Tokyo participated in the study. All were native speakers of Japanese who had never lived in an English-speaking society. They were selected by their English professor from a population of 150 students, on the basis of either superior (N=8) or inferior (N=8) performance on two standardized tests of spoken English perception and comprehension. In addition to these native speakers of Japanese, the experiment further included a control group of ten native speakers of English. They were undergraduates attending Bryn Mawr and Haverford Colleges.

Procedure

The experiment was divided into three stages and employed materials that have been described in detail elsewhere (Mann,

1980): a seven-member synthetic [da]-[ga] continuum and 12 natural tokens of [al] and [ar]. Stimuli along the [da]-[ga] continuum comprised three-formant syllables in which systematic variations in the onset of the third formant provided critical support for the [d]-[g] distinction. They were constructed so as to be compatible with the natural tokens of [al] and [ar]. Those tokens had been extracted from natural productions of [al-da], [al-ga], [ar-da] and [ar-ga] by a male native speaker of English in which the first syllable had been stressed. To control for the possibility of material-specific effects, three tokens of each of the four productions were employed.

In the first stage of the experiment, isolated stimuli from along the [da]-[ga] continuum were presented 12 times each, according to a randomized sequence. In the second, the [da]-[ga] stimuli were preceded by the tokens of [al] and [ar] and again presented 12 times in each context, according to an unblocked randomized sequence. In each stage, a 28-item practice sequence of the test items preceded the test sequence itself, and the task was to mark (on a response sheet containing both alphabetic script and Japanese Kana) whether a given stimulus contained [da] or [ga]. The third and final stage assessed subject's ability to identify [l] and [r] in the stimuli previously employed in the second stage of testing, by marking (on a response sheet written in alphabetic script) whether a given stimulus contained [al] or [ar]. In light of the potential difficulty of this task, listeners were first pretrained in the appropriate response categories for 28 items, and then given a practice sequence of 28 items in which they were told the correct response before listening to each stimulus. The test sequence then followed, randomized into a different order than that employed in the second stage of testing.

Results

The three panels of Figure 1 summarize the results obtained from the the native speakers of English, and the Japanese students who were superior and inferior students of spoken English. For convenience, the results obtained in the first stage of testing with isolated [da]-[ga] stimuli are not included in this preliminary report, as the various groups did not differ in their perception of these sounds, and as the main interest is in the contrasting effects of [al] and [ar].

The native speakers of English (1a) were 100% correct in identifying [al] and [ar], and showed the anticipated context effect of [l] vs. [r]. The Japanese speakers who were superior students of spoken English (1b) were 99% correct in identifying [al] and [ar], which confirms previous indications (MacKain, Best and Strange, 1981) that at least some native speakers of Japanese can master the [l]-[r] distinction. Like the native speakers of English, these subjects showed the contrasting effects of [l] and

[r] on perception of [da] and [ga]. In contrast to the other two groups of subjects, those Japanese subjects who were inferior students of spoken English (lc) averaged only 58% correct identification of [al] and [ar], which is not significantly better than chance. Nonetheless, they showed the contrasting effects of [l] and [r] on perception of [da] and [ga]. Analysis of variance reveals significant main effects of stimulus number, $F(6,138)=905.79$; $p<.00001$, and context, $F(1,23)= 130.19$, $p<.00001$, and an interaction of these two variables, $F(6,128)= 31.95$, $P<.00001$, but no interaction between subject group and context. There was also a main effect of subject group, $F(2,23)=9.58$; $p<.001$ and an interaction involving subject group with stimulus number, $F(12,138)= 4.91$; $p<.00001$, and a small three-way interaction, $F(12,138)=2.19$; $p<.02$. Each of these reflects the slightly aberrant behavior of the superior students of English in labeling the endpoints of the continuum.

Discussion

Thus, all subjects responded to some difference between [al] and [ar], and adjusted their perception of the [da]-[ga] distinction accordingly. That is, all listeners, native speakers of English and Japanese alike, heard some difference between utterances which contained [l] and [r]. If it is accepted that the context effect of [l] and [r] is specific to speech perception (Mann and Liberman, 1983) and reflects listeners' sensitivity to the acoustic consequences of articulation, one implication of the fact that inferior students of spoken English were sensitive to the context effect yet unable to identify [l] and [r], is that perception of speech comprises at least two levels. At one level, speech signals are used to track the nearly continuously changing movements of the vocal tract. Beyond that, there is a level at which the continuous movements are categorized into strings of phonemes. It is the tracking of vocal tract movements that is directly responsible for those context effects and trading relations in speech perception which rest on the integration, interpretation and abstract representation of incoming sensation as the product of human vocalization. The ability to respond to speech sounds in this way is independent of native language experience; hence speakers are sensitive to the articulatory properties of [l] and [r] whether or not those phonemes are part of their native inventory. Moreover, the tracking of vocal tract movement may precede phonetic representation, as listeners may respond to movements which they cannot categorize phonetically. Phonetic categorization would appear to depend upon language experience, hence listeners may encounter difficulty when they are required to phonetically categorize consonants that are not in their native inventory.

The distinction between using the speech signal to track vocal tract movements and using it to abstract the phonetic

segments that such movements convey accounts for the present findings. It accords with evidence that subjects are sensitive to the articulatory properties of vowels that are not part of their native language (Whalen, 1981). Finally, it can offer a perspective on the interpretation of findings about the speech perception capabilities of infants. Infants have given evidence of perceiving many phonetically-relevant properties of utterances (see, for a review, Eilers, 1980, see also Kuhl, 1980, and Kuhl and Meltzoff, 1982), as well as evidence of trading relations (Miller and Eimas, 1983). They may perceive human speech in a special way, perhaps owing to proclivities of the left or dominant hemisphere (MacKain, Studdert-Kennedy, Spieker and Stern, 1983) which mediates speech perception in adults (Studdert-Kennedy and Shankweiler, 1970). At present, in the absence of any means of verifying that infants perceive phonemes, as such, it is premature to accept a conclusion that they are capable of phonetic perception. Yet the data surely imply that they possess some perceptual abilities that are the basis of adult phonetic perception (Miller and Eimas, 1983). One of these could well be the ability to track the vocal tract movements that give rise to incoming speech stimuli, regardless of specific language experience.

Acknowledgments

This study was completed at the Research Institute of Logopedics and at the Komaba Campus of the University of Tokyo, while the author was a Fulbright Fellow. Recognition is due to Dr. Shigeru Kiritani and Dr. Hiroshi Suzuki for their advice and for their help in procuring subjects and a testing site. Ms. Michiko Mochizuki-Sudo is to be thanked for translating the instructions to the subjects.

References

1. Best, C. T., B. Morrongiello and R. Robson (1982); Perceptual equivalence of acoustic cues in speech and nonspeech perception. Perception & Psychophysics, 29, 191-211.
2. Eilers, R. (1980); Infant perception: History and Mystery. In G. H. Yeni-Komshian, J. F. Kavanaugh and C. A. Ferguson (Eds.) Child Phonology: Perception and Production, Volume 2. New York: Academic Press.
3. Eimas, P. (1975); Auditory and phonetic coding of the cues for speech: Discrimination of the [r]-[l] distinction by young infants. Perception & Psychophysics, 18, 341-347.
4. Goto, H. (1971); Auditory perception by normal Japanese adults of the sounds "R" and "L". Neuropsychologia, 9, 317-323.
5. Kuhl, P. K. (1980); Perceptual Constancy for Speech Sound Categories in Early Infancy. In G. H. Yeni-Komshian, J. F. Kavanaugh, and C. A. Ferguson (Eds.) Child Phonology:

Perception and Production, Volume 2. New York: Academic Press.
6. Kuhl, P. K. and A. N. Meltzoff (1982); The bimodal perception of speech in infancy. Science, 218, 1138-1144.
7. Liberman, A. M. (1982); On finding that speech is special. American Psychologist, 37, 148-167.
8. Liberman, A. M., D. Isenberg and B. Rakerd (1982); Duplex perception of cues for stop consonants: Evidence for a phonetic mode. Perception and Psychophysics, 30, 133-143.
9. MacKain, K. S., C. T. Best and W. Strange (1981); Categorical perception of English [r] and [l] by Japanese bilinguals. Applied Psycholinguistics, 2, 369-390.
10. MacKain, K., M. Studdert-Kennedy, S. Spieker and S. Stern (1983); Infant Intermodal Speech Perception is a Left Hemisphere Function. Science, 219, 1347-1349.
11. Mann, V. A. (1980); Influence of preceding liquid on stop consonant perception. Perception & Psychophysics, 28, 407-412.
13. Mann, V. A. and A. M. Liberman (1983); Some differences between phonetic and auditory modes of perception. Cognition, 14, 211-235.
14. Mann, V. A. and B. H. Repp (1981); Influence of preceding fricative on stop consonant perception. Journal of the Acoustical Society of America, 69, 548-558.
15. Miller, J. L. and P. D. Eimas (1983); Studies on the categorization of speech by infants. Cognition, 13, 135-166.
16. Miyawaki, K., W. Strange, R. Verbrugge, A. M. Liberman, J. J. Jenkins and O. Fujimura (1975); An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Perception & Psychophysics, 18, 331-340.
17. Mochizuki, M. (1981); The identification of /r/ and /l/ in natural and synthesized speech.Journal of Phonetics, 9, 283-303.
18. Rand, T. C. (1974); Dichotic release from masking for speech. Journal of the Acoustical Society of America, 55, 678-680.
19. Repp, B. H. (1982); Phonetic trading relations and context effects: New evidence for a phonetic mode of perception. Psychological Bulletin, 92, 81-110.
20. Repp, B. H., A. M. Liberman, T. Eccardt and D. Pesetsky (1978); Journal of Experimental Psychology: Human Perception and Performance, 4, 621-637.
21. Studdert-Kennedy, M. and D. Shankweiler (1970); Hemispheric specialization for speech perception. Journal of the Acoustical Society of America, 48, 579-594.
22. Whalen, D. H. (1981); Effects of vocalic formant transitions and vowel quality on the English [š]-[s] boundary. Journal of the Acoustical Society of America, 69, 275-282.

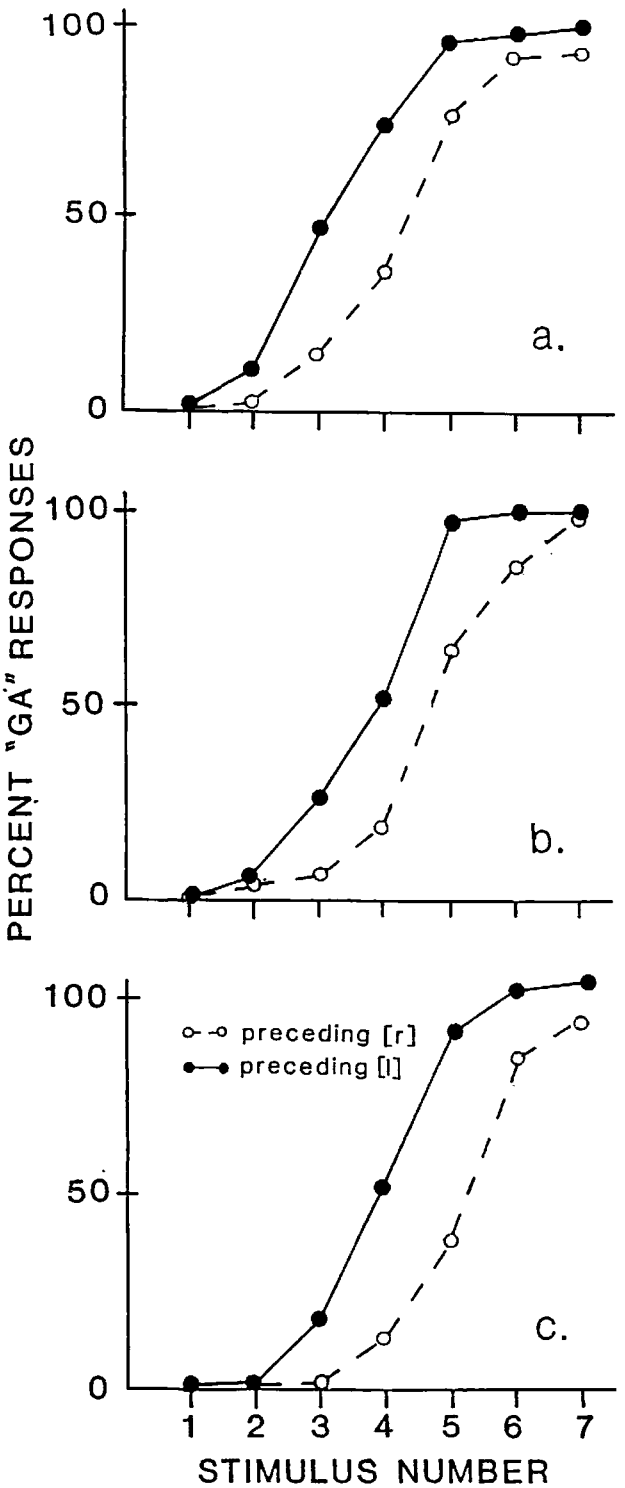

Figure 1: The contrasting effects of [l] and [r] on perception of the [d]-[g] distinction by: a) native speakers of English who are 100% correct in identifying [l] and [r], b) native speakers of Japanese who are 99% correct in labeling [l] and [r], and c) native speakers of Japanese who perform at chance level in labeling [l] and [r].