Cognition, 24 (1986) 169–196

Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r"*

VIRGINIA A. MANN

University of California, Irvine

Abstract

Native speakers of Japanese may be unable to correctly identify the phonemes /l/ and /r/ in spoken English. Nevertheless, in perceiving English utterances, they, like native speakers of English, respond to the different acoustic patterns which convey /l/ and /r/ as if they are sensitive to differences in the vocal tract movements that convey /l/ and /r/. Support for this conclusion is provided by a study in which native speakers of Japanese and native speakers of English labelled stimuli along a synthetic /da/-/ga/ continuum when the stimuli were preceded by natural tokens of /s/ or /ʃ/, /al/ or /ar/. Each pair of precursors had contrasting effects on the location of the category boundary between /da/ and /ga/, and neither the direction nor the extent of contrast depended on native language experience. Significantly, /al/ gave rise to more /ga/ percepts than /ar/ for Japanese and English speakers alike, regardless of their ability to identify /al/ and /ar/, as such. Interpretation of these results rests on previous observations that the contrasting perceptual effects of /al/ vs. /ar/ and /s/ vs. /ʃ/ find parallels in the acoustic structure of natural utterances of /al-da/, /ar-da/ etc., due to coarticulation of the vocal tract movements that convey the preceding consonant and those that convey the following /da/ or /ga/. Apparently, native speakers of Japanese can be sensitive to the acoustic consequences of coarticulating /l/ or /r/ with /d/ or /g/ while being unable to categorize /l/ and /r/ as different phonemes. Preceding a language-specific level of perception

* This study was conducted at the Komaba Campus of the University of Tokyo, while the author was a Fulbright Fellow at the Research Institute of Logopedics and Phoniatrics, Faculty of Medicine, University of Tokyo. It was partially supported by NIH Grant HDO1994 and BRS Grant 05596 to Haskins Laboratories, Inc. and by NICHD Grant HD21182 to Bryn Mawr College, and by the Center for Cognitive Sciences at MIT. Recognition is due to Dr. M. Sawashima, 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. Drs. Michael Studdert-Kennedy, Alvin Liberman and Howard Hoffman are acknowledged for their helpful comments on earlier drafts. Reprint requests should be addressed to Virginia A. Mann, Program in Cognitive Science, School of Social Sciences, University of California, Irvine, CA 92717, U.S.A.

where speech sounds are represented in accordance with the constraints of a given phonological system, there may exist a universally-shared level where the representation of speech sounds more closely corresponds to the articulatory gestures that give rise to the speech signal.

Introduction

What do native speakers of Japanese perceive as they listen to English utterances that 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, & Fujimura, 1975; Mochizuki, 1981). Their behavior would seem to suggest that they hear these two speech sounds as one and the same. Yet this study offers evidence that Japanese subjects are perceptually sensitive to the acoustic consequences of an articulatory difference between utterances containing /l/ and /r/, whether or not they can explicitly identify these speech sounds.

The demonstration of this ability is provided by a certain context effect in speech perception. That effect was first observed in a previous study (Mann, 1980) in which natural utterances of /al/ and /ar/ were placed in front of synthetic speech stimuli from along an acoustic continuum that ranged from /da/ to /ga/. The resulting stimuli sounded like /al-da/, /al-ga/, /ar-da/ or /ar-ga/, and when native speakers of English were asked to decide whether "da" or "ga" was heard, their responses revealed that the location of the category boundary between /d/ and /g/ was influenced by certain attributes of the preceding utterance. In particular, when the preceding utterance ended in /l/, the boundary was shifted towards more /g/ percepts (fewer /d/ percepts), relative to that obtained when the preceding utterance ended in /r/ (Mann, 1980).

An explanation of the contrasting effects of /l/ and /r/ is offered by the view that speech signals are perceived as if by reference to the articulatory gestures by which they are produced. Many different forms of evidence support the view that sensitivity to the lawful relationship between acoustic speech signals and the articulatory gestures which they reflect is a highly specific, built-in property of the speech perception module (for a recent review, see Liberman & Mattingly, 1985). Here, two related observations are offered as evidence that the context effect of /l/ vs. /r/ may reflect listeners' sensitivity to the relationship between speech signals and articulatory gestures. First, there is the finding that, for native speakers of English, the effect of a preceding utterance on the distinction between /da/ and /ga/ is not limited to utterances

containing /l/ and /r/, but extends to utterances containing /s/ and /ʃ/ (Mann & Repp, 1981), and that similarities are best described in terms of certain articulatory properties of the preceding utterance. Specifically, utterances of /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 utterances of /r/ and /ʃ/, which are produced with a more retracted tongue posture. Second, the perceptual context effects of /l/ and /r/, /s/ and /ʃ/ find parallels in the acoustic structure of natural utterances of /al-da/, /ar-da/, /as-da/, /aʃ-da/ etc., and articulatory factors once again offer an explanation. Specifically, when a sequence of phonemes is produced, the gestures which convey the individual phonemes tend to overlap and become interwoven, and owing to this coarticulation, the acoustic structure of /da/ and /ga/ can systematically vary as a function of whether they follow /al/ or /ar/ (Mann, 1980), /s/ or /ʃ/ (Mann & Repp, 1981; Repp & Mann, 1981, 1982). Together, then, these two observations suggest that the context effect of utterances ending in /l/ and /r/, like many other context effects and trading relations (see, Repp, 1982, for a review) represents a perceptual sensitivity to the acoustic consequences of articulatory interactions.

One might nevertheless ask whether the effect of /al/ vs. /ar/ could be accounted for by reference to purely acoustic interactions instead of articulatory ones. To appeal to auditory interaction requires that we either put aside the coarticulatory facts entirely, or that we explain those facts with the assumption that speakers adjust the behavior of their articulators so as to produce in each context just those acoustic effects that will fit with preexisting auditory interactions. Neither alternative is very appealing, yet an account in terms of auditory interactions cannot, in principle, be ruled out and therefore merits discussion.

As a preface to considering an acoustic explanation of the contrasting effects of /l/ and /r/, the acoustic structure of the stimuli employed in Mann (1980) must first be described in some detail. In that experiment, the /da/-/ga/ stimuli varied in the onset of the third formant from 2690 to 2104 Hz in six approximately equal steps, but the stimuli were constant in all other respects, with onset frequencies of the second and first formant set at 1588 and 649 Hz, respectively. The context effect had emerged when the /da/-/ga/ stimuli were preceded by a 50 ms silent gap which was in turn preceded by natural tokens of /al/ or /ar/. In those tokens, the average frequencies for the offset of the second and third formants may be considered the most relevant to any consideration of acoustic interactions, as these were closest to the frequency region which contained the cues at stake in the /da/-/ga/ distinction. Averaged across the twelve tokens of each syllable, the offset for the second and third formants was 972 and 2711 Hz for /al/ and 1390 and 1733 for /ar/ (the standard

deviation in each case is less than 100 Hz: for details, see Mann, 1980).

Turning now to the possibility of an acoustic explanation of the context effect, perhaps the most appealing candidate is suggested by the phenomenon of post-stimulus masking (i.e. forward masking): Post-stimulus masking could cause the auditory response to the onset of the acoustic patterns that convey /da/ and /ga/ to be influenced by the acoustic patterns that convey a preceding /al/ or /ar/. The question to be asked is whether this change in auditory response can account for listeners' tendency to give more /ga/ responses in the context of /l/ relative to /r/. The possibility that different preceding utterances might have different influences on the auditory response to the onset of the /da/-/ga/ stimuli is suggested by studies of the physiological characteristics of auditory nerve fibers (see, for example Delgutte & Kiang, 1984; Harris & Dallos, 1979; Smith, 1977) and by psychophysical studies of human audition (see, for example Elliot, 1971; Moore, 1978). Both avenues of research offer evidence that, at a gap size of 50–100 ms, post-stimulus masking leads to a depression of the onset response to stimulation which falls in the same frequency range as the preceding stimulus. Thus, the preceding /al/ and /ar/ might be expected to have different effects on the auditory response to the onset of certain stimuli along the /da/-/ga/ continuum, in light of their different spectral characteristics.

Specifically, the effects of post-stimulus masking should be as follows. A higher-frequency third formant offset is one attribute that distinguishes /al/ from /ar/, and the energy in that region can be expected to suppress the auditory response to /da/-/ga/ stimuli whose onset spectra contains energy in the same critical band (at 2700 Hz, the bandwidth for human listeners is approximately 400 Hz: Scharf, 1970). Thus for the first three to five stimuli along the continuum, a preceding /al/ should cause a relative weakening of the contribution of the third formant onset frequency to perception of the /d/-/g/ distinction. In contrast /ar/, whose third formant offset is considerably lower in frequency, would cause a relative weakening of the contribution of the second formant onset frequency, but little effect on the third formant onset frequency (given that critical bandwidth at 1700 Hz is approximately 200 Hz). Can these differences explain listeners' tendency to give more /ga/ responses to stimuli preceded by /al/ as opposed to /ar/?

In answering this question, it should first be noted that, in Mann's original experiment, the context effect of /l/ vs. /r/ was due to /l/ favoring /ga/ responses relative to stimuli preceded by /r/ and to stimuli presented in isolation. As the pattern of responses to stimuli preceded by /r/ did not significantly differ from that to isolated stimuli, it is appropriate that an explanation of the context effect be focused on the post-stimulus masking influences of /al/, and the postulated weakening of the response to the onset of the third for-

mant does precisely that. The masking account further predicts that post-stimulus masking should be most clearly evident in the case of the first three to five stimuli along the continuum, and for the most part, the data uphold this prediction as well. Although the effect of /al/ was not evident in the case of the first stimulus along the continuum, it was at its strongest for stimuli two through five (i.e. the mid-range stimuli). Thus post-stimulus masking offers a plausible explanation of the contrasting effects of /al/ vs. /ar/.

A more recent study, however, provides direct evidence against post-stimulus masking and other purely auditory explanations of the effect of /l/ vs. /r/ (Mann & Liberman, 1983). That study made use of the phenomenon known as duplex perception (Liberman, Isenberg, & Rakerd, 1982; Rand, 1974) to reveal that the effects of /l/ vs. /r/ reflect a special property of perception in the speech mode. In duplex perception, one and the same stimulus is simultaneously heard as speech and as nonspeech. This situation can be created by dividing synthetic stimuli from along a /da/ to /ga/ continuum into two parts; a constant base portion, which in isolation sounds like /da/, 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: It fuses with the base to provide critical support for the perception of /da/ or /ga/, yet is also heard as a nonspeech "chirp". The speech percept is not noticeably different from what listeners perceive when the transition and base are electronically mixed, and the "chirp" percept is not noticeably different from what subjects perceive when the transition is presented in isolation.

The advantage of the duplex phenomenon is that one stimulus is simultaneously perceived as speech and as nonspeech. By manipulating listeners' attention to one percept or the other we can discover the commonalties and distinctions between processing in the two modes, reasoning that any context effects which hold for both percepts merit an account in terms of auditory interactions, whereas effects that are restricted to one percept or the other warrant a mode-specific explanation. This approach was used to provide evidence that the contrasting effects of /al/ and /ar/ on perception of the /da/-/ga/ distinction reflect special properties of perception in the speech mode (Mann & Liberman, 1983), in an experiment that involved stimuli along a /da/-/ga/ continuum, divided into base and transition so as to form duplex percepts, and presented to listeners who were instructed to attend to one percept or the other as they attempted to discriminate stimuli in an AXB paradigm.

Under instructions to attend to the speech percepts of /da/ and /ga/, perception was categorical, and when the stimuli were preceded by natural tokens of /al/ and /ar/ the location of the category boundary shifted, with /al/ having

the expected effect of favoring more /ga/ percepts than /ar/. In contrast, under instructions to ignore the speech percepts and attend to the nonspeech chirps, perception was continuous and was not systematically altered by the presence of the preceding /al/ or /ar/. As regards the post-stimulus masking account of the effects of /al/ vs. /ar/, it should be noted that, since the effect on speech perception of the /da/-/ga/ stimuli obtained when the preceding /al/ and /ar/ and the third formant transition were presented to opposite ears, a cochlear-level explanation is ruled out. This argues against post-stimulus masking of the sort identified in studies of auditory nerve (for example Delgutte & Kiang, 1984; Harris & Dallos, 1979; Smith, 1977). A more direct argument against an acoustic explanation of the effects of /al/ and /ar/ is provided by the finding that the presence of /l/ vs. /r/ did not alter the pattern of listeners' responses under instructions to attend to their nonspeech percepts of the third formant transition.

Thus the contrasting influence of /al/ and /ar/ is evident only when acoustic stimuli are perceived as speech, in accordance with the view that this context effect reflects a special property of perception in the speech mode. It is as if in that mode of perception a stimulus which conveys /l/ or /r/ somehow serves as an anchor against which a following stimulus may be judged as /d/ or /g/. I return, then, to evidence that the specific effects of the preceding /l/ or /r/ operate in parallel to certain articulatory interactions, and to evidence that a similar context effect obtains when /da/-/ga/ stimuli are preceded by other speech sounds (e.g. /s/ and /ʃ/) which share certain articulatory properties with /l/ and /r/ but not the same spectral properties. Such evidence favors an explanation that the context effect reflects listeners' sensitivity to the lawful relationship between acoustic speech signals and the articulatory gestures which they represent, a sensitivity which is a highly specific, built-in property of the speech perception module (Liberman & Mattingly, 1985).

With this effect and its explanation in mind, we may now return to the question of what Japanese listeners perceive as they listen to utterances which contain /l/ and /r/. English and Japanese distinguish many of the same speech sounds, including /d/ and /g/. Thus deciding whether utterances contain /da/ or /ga/ should pose no difficulty for speakers of either language. However, Japanese does not distinguish the English liquids /l/ and /r/. There is no /l/ in Japanese, and while there is a Japanese /r/, it more clearly resembles the English alveolar flap (i.e., the medial consonant in "ladder") than English /r/. As noted previously, in the absence of early experience with a language in which /l/ and /r/ are contrastive, many native speakers of Japanese are unable to distinguish utterances which contain English /l/ and /r/ in either labelling or discrimination tasks which focus on the /l/-/r/ distinction (Goto, 1971; Miyawaki et al., 1975; Mochizuki, 1981). Are we to conclude that

Japanese listeners hear /l/ and /r/ as one and the same? Here it is asked whether the technique of using a context effect to indirectly assess sensitivity to articulatory differences between /l/ and /r/ will reveal perceptual sensitivities that have heretofore been undisclosed by tasks requiring explicit, categorical responses to /l/ and /r/.

A demonstration that utterances which end in /l/ and /r/ are able to differentially influence perception of the /d/-/g/ distinction by Japanese listeners would imply that native language experience has less of an influence on one aspect of speech perception, namely, the ability to untangle the acoustic consequences of articulatory interactions, than on another, namely, the ability to make phonemic categorizations of speech sounds. Evidence that native language experience influences speech categorization is provided by findings that adult subjects have difficulty identifying or discriminating phonetic contrasts that are not used distinctively in their native language (see, for example Lisker & Abramson, 1970; Werker, Gilbert, Humphrey, & Tees, 1981; Trehub, 1976). However, the possibility that there are also certain universal speech perception abilities which depend less strongly on native language experience is suggested by findings that infants behave as if they are able to perceive many phonetically-relevant properties of speech which are not necessarily exploited by their language community (see, for a review, Eilers, 1980; Werker et al., 1981). Further support for this contention is provided by a finding that, under certain circumstances, adults may discriminate speech sounds according to categories used in another, but not their own, native language (Werker & Tees, 1984b). At present, problems of interpretation arise as to whether the "universal" abilities shown by infants and adults are of a general acoustic sort or are somehow specific to speech perception (for a discussion, see Werker & Tees, 1984a). One virtue of the present context effect as a probe to universal processes in speech perception is that acoustic explanations may be discounted, as noted previously, by a consideration of known acoustic interactions and by reference to the results obtained with the duplex paradigm (Mann & Liberman, 1983).

Two experiments are reported in the sections which follow. Experiment II, the major focus of this study, determined how Japanese subjects label synthetic /da/-/ga/ stimuli preceded by natural tokens of /al/ and /ar/ and how their performance relates to their ability to phonemically categorize /al/ and /ar/. Experiment I is a preliminary study which established that native speakers of Japanese are capable of showing context effects when the task involves labelling /da/ and /ga/ in English utterances which violate a general principle of the syllable structure of Japanese. It was prompted by the realization that context effects in utterances like /al-da/, /ar-da/, etc. could be doubly problematic for the native speaker of Japanese because such utterances not only

contain /l/ and /r/, but also contain a consonant sequence, and consonant sequences, in general, are not permissible in Japanese.

Experiment I

The phonology of Japanese does not permit consonant sequences to occur either within a syllable or across a syllable boundary; the most comparable situation involves two voiceless consonants separated by a devoiced vowel, in which case, the vowel still appears to be articulated (see Beckman & Shoji, 1984). To determine whether sequences of consonants are processed differently by native speakers of Japanese and English, this study assessed the effects of preceding /s/ and /ʃ/ on perception of the /da/-/ga/ distinction (Mann & Repp, 1981; Repp & Mann, 1981) by speakers of the two languages. Two different groups of Japanese subjects are included: those who are superior beginning students of spoken English, and those who are inferior ones. As noted in the Introduction, the effects of /s/ and /ʃ/ parallel those of /l/ and /r/, however, unlike /l/ and /r/, /s/ and /ʃ/ are contrasting phonemes in Japanese thus a lack of experience with this distinction is not at issue. Should the Japanese subjects prove sensitive to the context effect of /s/ vs. /ʃ/, it would demonstrate that a lack of familiarity with sequences of consonants is not a limiting factor. This finding is anticipated by a previous observation that, for native speakers of English, phonological restrictions against sequences of /ʃt/ and /ʃk/ in syllable initial position do not prevent native speakers of English from showing context effects in utterances such as /ʃta/ and /ʃka/ (Mann & Repp, 1981).

Methods

Subjects

Native speakers of Japanese

The Japanese subjects who participated in the study were 38 college freshmen enrolled in the first semester of a spoken English course at the University of Tokyo. 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 ($N = 19$) or inferior ($N = 19$) performance on two standardized tests: a test of the perception of spoken English words which had been developed at Keio University (Koike, 1978), and the Jacet-Coltd Listening Comprehension Test

Table 1. *Oral English profile: Japanese subjects participating in Experiment I*

	College testing		Experience prior to college[a] (0–5 pt. scale, 5 = extensive experience)				
	Jacet-Coltd (max. 120)	Koike (max. 50)	Before Jr. high	Jr. high		Sr. high	
				School	Home	School	Home
Superior students	97.5	48.4	1.95	3.63	3.32	2.74	1.42
Inferior students	42.6	39.3	0.77	3.09	2.23	2.32	0.45

[a] Ratings computed by the students' English professor on the basis of responses to a questionnaire.

which involves the comprehension of spoken English sentences (Jacet-Coltd Form A, 1975). For the inferior and superior students, the cutoff scores for the Keiko test were 44/50 and 48/50, respectively; for the Jacet-Coltd test, 68/120 and 88/120, respectively. Mean scores on each test appear in Table 1, together with a summary of prior exposure to oral English.

Native speakers of English

In addition to the native speakers of Japanese, the experiment further included a control group of ten native speakers of English who were freshmen attending Bryn Mawr and Haverford Colleges.

Materials and procedure

The experiment employed materials that have been described in detail elsewhere (Repp & Mann, 1981): a seven-member synthetic /da/-/ga/ continuum and natural tokens of /s/ and /ʃ/. 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. There were seven CV stimuli, distinguished solely by the onset of the third formant frequency, which decreased from 3222 Hz to 1902 Hz in steps of approximately 215 Hz. All stimuli had stepwise-linear 50-ms formant transitions (from 285 to 771 Hz for the first formant, from 1770 to 1223 for the second and to 2520 Hz for the third) followed by 200-ms steady-state portions. Fundamental frequency fell linearly from 110 to 80 Hz and the amplitude contour was flat with 50 ms onset ramp and 30 ms offset ramp.

The tokens of /s/ and /ʃ/ had been extracted from natural productions by a male, phonetically-trained native speaker of English of /sta/, /ska/, /ʃta/ and /ʃka/. The spectral composition of these stimuli is summarized in Figure 5 of Repp and Mann (1981); the major prominence for tokens of /s/ is between 3500 and 4000 Hz, whereas that for tokens of /ʃ/ is between 1800 and 3000 Hz. The utterances had been produced in a random order and digitized at 10,000 Hz. For three different tokens of each of the possible productions, the fricative noises were excerpted and stored separately for three different tokens of each of the four possible productions. As in Repp and Mann's study, the use of multiple tokens served two functions. First, it offered a control for the possibility of material-specific effects. Second, the systematic variations in the utterances from which the fricative noises were drawn (i.e. whether they contained /t/ or /k/) provided a means of assessing trading relations as well as context effects. There was interest in the question of whether Japanese subjects might show trading relations and/or context effects, given that Repp and Mann found native speakers of English to show both. In their study, /s/ had favored more "g"/"k" responses than /ʃ/ (a context effect), but fricative noises which had been extracted from utterances that contained /k/ also favored more "g"/"k" responses than those produced before /t/ (a trading relationship). We will return to the significance of these findings in the presentation and discussion of the results.

There were a total of three stages to the experiment. In the first, isolated stimuli from along the /da/-/ga/ continuum were presented 10 times each, according to a randomized sequence. A practice sequence of 20 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/. In the second, the /da/-/ga/ stimuli were preceded by the tokens of /s/ and /ʃ/ with a 75 ms gap between the offset of frication and the onset of the CV portion. Each CV stimulus was presented 12 times in each context (i.e. 4 times per token), according to a completely randomized sequence, preceded by a completely randomized 28-item practice sequence of the test items. In this stage, as previous studies employing /da/-/ga/ stimuli preceded by /s/ and /ʃ/ had revealed that native speakers of English hear the resulting stimuli as /sta/, /ska/, etc. (in keeping with the phonological constraints of English), subjects were given the option of choosing to label the stimuli as /da/-/ga/ or /ta/-/ka/.

The third and final stage of the experiment assessed subjects' ability to identify /s/ and /ʃ/ in the test materials. Perception of /s/ vs. /ʃ/ was assessed at the conclusion of the experiment in order to minimize the possibility that a lack of familiarity with the test utterances might limit the Japanese listeners' ability to categorize these speech sounds. The response sheet contained both

alphabetic transcription ("s" vs. "sh") and Japanese Hiragana (for the syllables /su/ and /ʃi/ as these most closely resemble isolated fricatives insofar as vowel devoicing is permissible in certain contexts). Subjects were asked to mark which orthographic convention they preferred. They received a practice sequence of 28 items, followed by a test sequence, where in each case, the items were the same as those employed in the second stage of testing, presented in a completely randomized sequence.

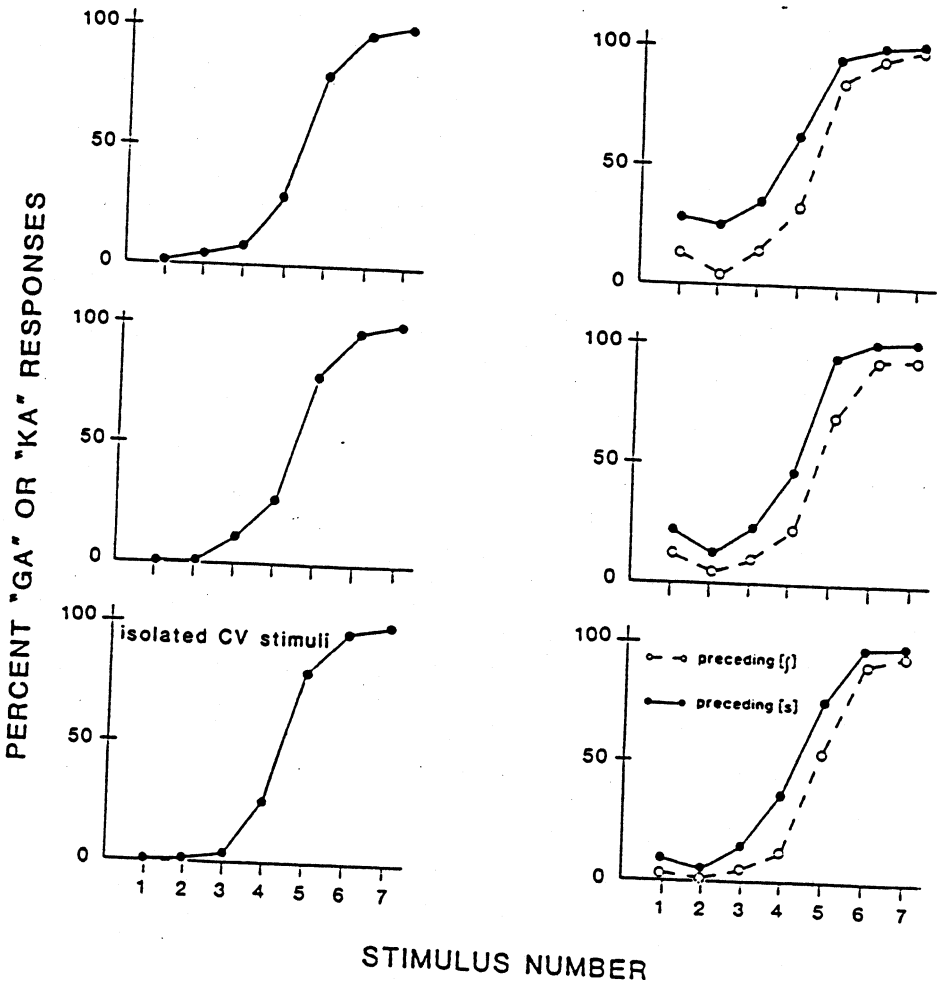
Results

Analysis of the data began with the third stage of testing, which revealed that all subjects were able to label /s/ and /ʃ/ correctly. There were no appreciable differences in the accuracy of the two groups of Japanese subjects or between them and the native speakers of English: All speakers of English had been 100% correct, the superior students of spoken English had averaged 99.8% correct, and the inferior students, 99.4%. Only three of the nineteen superior students had made any errors, and the maximum error rate was 2.5%; only six of the nineteen inferior students had made errors and the maximum rate was 3.6%.

The data from the first two stages of testing comprise the labelling of the /da/-/ga/ stimuli in isolation and in the context of a preceding /s/ or /ʃ/. Figure 1 summarizes the average responses of each group of subjects in terms of the mean percentage of "ga" (or "ka") responses given to stimuli at each of the seven positions along the synthetic /da/-/ga/ continuum and also summarizes the context effect of /s/ vs. /ʃ/. From top to bottom, the panels contain the results for the three subject groups: the native speakers of English, the native speakers of Japanese who were superior students of spoken English, and the native speakers of Japanese who were inferior students of spoken English. The panels on the left concern labelling of the isolated /da/-/ga/ stimuli, collapsed across subjects within each group. Those on the right concern labelling of stimuli preceded by /s/ and /ʃ/, with each curve representing the average percentage of "ga"/"ka" responses collapsed across the six tokens of each fricative and across subjects.

It can be seen quite clearly that all subjects labelled the stimuli in a comparable fashion. In the left-hand panels, the percentage of /ga/ responses varied according to stimulus position ($F(6,270) = 445.15, p < .0001$), showing the ogive curve which is typical of labelling functions for synthetic CV continua. When, as shown in the right-hand panels, the /da/-/ga/ stimuli were preceded by /s/ and /ʃ/ the anticipated context effect emerged. As in Mann and Repp (1981), there is a context effect due to /s/, and most importantly, this result held for all three groups of subjects. In interpreting the results in

Figure 1. *The pattern of "ga" responses given to stimuli along an acoustic /da/-/ga/ continuum when the stimuli were presented in isolation (left panels) and when they were preceded by /s/ and /ʃ/ (right panels). From top to bottom, subjects include: (1) native speakers of English, (2) native speakers of Japanese who are superior students of spoken English, and (3) native speakers of Japanese who are inferior students of spoken English.*



the right-hand panel of Figure 1, an ANOVA was computed on the total number of "ga" responses given to stimuli across the continuum, considering the context effect of /s/ vs. /ʃ/, and whether that fricative had been extracted from a syllable containing /t/ or /k/ (i.e. the trading relation involving cues to /t/ vs. /k/). There was a main effect of subject group ($F(2,45) = 6.19$, $p < .001$), and a main effect of /s/ vs. /ʃ/ ($F(1,45) = 70.65$, $p < .0001$), but no interaction involving these variables ($p > .1$). In general, the effect of subject group was due to the tendency of the inferior students of spoken English to give fewer "ga"/"ka" responses than subjects in the other two groups, but this tendency did not appear to limit their sensitivity to the influence of /s/ vs. /ʃ/.

Aside from the context effect due to /s/ vs. /ʃ/ there was also a trading relation which reflected whether /t/ or /k/ had been present in the original utterances from which the tokens of /s/ and /ʃ/ had been extracted ($F(1,45) = 70.13$, $p < .0001$), this being particularly strong for tokens of /s/ ($F(1,45) = 18.75$, $p < .001$). All three groups showed this effect to the same extent, as evidenced by a lack of two- or three-way interactions involving subject group. The effect of tokens extracted from utterances containing /t/ vs. /k/ is not evident in Figure 1, where the data are collapsed across all token variations, but when responses were considered in terms of whether each fricative noise had been extracted from a syllable containing /t/ vs. /k/ it was evident, as had been observed by Mann and Repp (1981), that tokens extracted from /ska/ and /ʃka/ gave rise to more "ga"("ka") responses than those extracted from /sta/ and /ʃta/ (an average of 65% vs. 41% "ga"/"ka" responses). This contrast has been taken as evidence of listeners' ability to recover certain acoustic properties of the fricative noise which directly reflect the place of articulation of the following stop consonant. The effect is termed a trading relation as opposed to a context effect, since it is assimilatory (i.e. the original presence of /k/ favors perception of "g"/"k") whereas context effects are contrastive (i.e. the more forward-produced fricative, /s/ favors perception of the less forward-produced consonants "g" and "k"). (A further discussion of trading relations and context effects is available in Repp, 1982.)

The most notable difference between subjects concerned the orthographic convention used to label the utterances in this Experiment. In keeping with the phonological constraints of English, all of the native speakers of English labelled the /da/-/ga/ stimuli as "ta" and "ka" in the context of the preceding /s/ and /ʃ/, whereas only 12 of the nineteen superior students and only four of the nineteen inferior students of spoken English did so. In labelling the English /s/ and /ʃ/, all of the native speakers of English, and sixteen of the superior students of spoken English chose alphabetic notation, whereas only six of the inferior students did so.

Discussion

The results of the first experiment reveal that native speakers of Japanese label stimuli along an English /da/-/ga/ continuum much the same as do native speakers of English, regardless of their competence in spoken English. Like native speakers of English, they also exhibit a context effect in which they appear to take account of the place of articulation of a preceding /s/ or /ʃ/ when they determine the place of articulation of a following /d/ or /g/. For superior and inferior students of spoken English, for Japanese and English speakers, alike, the perceived place of articulation tends to be shifted away from that of the preceding /s/ or /ʃ/. This contrast between properties of successive speech sounds is termed a context effect, and has been presumed to reflect a perceptual adjustment for assimilative coarticulatory interactions in fricative-stop consonant sequences (Mann & Repp, 1981; Repp & Mann, 1981).

Japanese subjects are further like native speakers of English in that they prove sensitive to whether tokens of /s/ or /ʃ/ were extracted from utterances in which they were originally produced before /t/ or /k/. This sensitivity involves an assimilation of temporally and spectrally diverse cues to a given speech sound, and is termed a trading relation. Presumably, the trading relation occurs because listeners are able to perceptually recover those cues in the fricative noise which are a direct consequence of anticipatory articulation of a following stop consonant (Repp & Mann, 1981).

The only noteworthy difference which related to native language experience and the level of spoken English competence concerns the labelling responses of the less skilled students of English. These subjects gave fewer "ga"/"ka" responses, were less likely to comply with certain conventions of English phonology/orthography as to the transcription of voicing, and also tended to employ Japanese Kana instead of the English alphabet. Despite this tendency, however, all three groups of subjects showed context effects and trading relations to the same extent. It can be concluded that, despite a lack of familiarity with consonant sequences, native speakers of Japanese are equivalent to native speakers of English in their perception of /da/-/ga/ stimuli preceded by /s/ and /ʃ/. Experiment II now turns to the focal issue of this study: whether Japanese subjects who are unable to identify English /l/ and /r/ are sensitive to the influences of a preceding /l/ or /r/ on perception of /da/-/ga/ stimuli.

Experiment II

Experiment II used the same three groups of subjects who participated in Experiment I to investigate the contribution of language experience to the influence of /al/ and /ar/ on perception of /da/-/ga/ stimuli. Several different outcomes are conceivable. First, all subjects might exhibit the same context effect as native speakers of English, whether or not they perceive /l/ and /r/ as different phonemes. Such a finding would reveal that all subjects are perceptually sensitive to some difference between utterances which contain /l/ and /r/. The context effect would occur because listeners would be able to take account of the acoustic consequences of certain articulatory differences between /l/ and /r/ when required to determine whether a subsequent phoneme is /d/ or /g/, even though they might not be able to identify /l/ and /r/ as such. This would imply the existence of some universal level of speech perception where utterances of any language are represented in terms which correspond rather objectively to their underlying articulatory properties, as opposed to a language-dependent level where utterances are represented according to the phonetic inventory of a given language.

Other possible outcomes involve a difference between native speakers of English and Japanese speakers who cannot distinguish /l/ and /r/. Here, two possibilities are considered. On the one hand, those subjects who cannot label /l/ vs. /r/ correctly may fail to exhibit any contrasting effects at all. They may perform the same whether listening to isolated stimuli or to stimuli preceded by /al/ and /ar/—as if they ignored the context altogether; they may treat both /al/ and /ar/ as native speakers of English treat /al/—implying some appreciation of the context but a failure to differentiate /l/ vs. /r/. In either case, it would reveal that the context effect exhibited by native speakers of English operates at a language-dependent level of processing which is unavailable to speakers who cannot categorize /l/ and /r/, and would constitute yet another finding that many Japanese listeners hear no difference between English utterances which contain /l/ and /r/.

Another possible difference between native speakers of Japanese and native speakers of English is that /al/ and /ar/ will have contrasting effects on the perception of /da/ and /ga/ by the Japanese listeners, but in opposite directions for subjects who can, and cannot label /l/ and /r/. That is, subjects who can label /l/ and /r/ correctly might give the same result as native speakers of English—a context effect involving more "ga" responses in the context of /al/—whereas those who label /l/ and /r/ indiscriminately might give more "da" responses instead. This could occur if, in utterances such as /al-da/ etc., Japanese listeners hear /a-da/ instead of /al-da/ not because they ignore the cues to /l/, but because they incorrectly assimilate cues to the /l/ and the /d/

into a single consonant. That is, the finding that a preceding /al/ favors more /da/ percepts than /ar/ could suggest that the Japanese subjects erroneously assimilate information about the relatively forward-articulated liquid, /l/ with information about the following stop consonant, as if they correctly perceive the place of articulation of /l/ vs. /r/ but assimilate that information into perception of the /d/-/g/ stimuli rather than contrasting information about the two speech segments. Such a segmentation problem might seem unlikely, given some observations that the problems of Japanese listeners tend to involve confusing /l/ and /r/ rather than a failure to recognize that a liquid has occurred. However there is evidence that Japanese listeners have inordinate difficulty in labelling /l/ and /r/ when they form part of a consonant sequence (Mochizuki, 1981; Sheldon & Strange, 1982). Also, on the Koike test (Koike, 1978) a common misperception involves the omission of a liquid from a consonant sequence (i.e. confusing "barn" and "bon", or "bird" and "bud").

Methods

Subjects

The subjects were the same Japanese and American college students who participated in Experiment I.

Materials and procedure

The experiment employed materials that have been described in detail elsewhere (Mann, 1980): a seven-member /da/-/ga/ continuum and 12 natural tokens of /al/ and /ar/. The tokens of /al/ and /ar/ had been produced by a male, trained phonetician who was a native speaker of English. They occurred as part of the disyllables /al-da/, /al-ga/, /ar-da/ and /ar-ga/ spoken with primary stress on the first syllable: three tokens of each disyllable had been digitized at Haskins Laboratories and the VC and CV portions had been separated (i.e. the portions of each disyllable preceding and following the stop-closure interval) and stored for later use. The use of multiple tokens of each VC replicated Mann (1980) and was designed to control for the possibility of material-specific effects, and to probe for the possibility of a trading relation orthogonal to the context effect of /l/ vs. /r/. In Mann's original study, /l/ had favored more "ga" percepts than /r/ (a context effect), but tokens of /al/ and /ar/ which had been extracted from utterances containing /ga/ gave rise to more "ga" percepts (a trading relation).

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 tokens of /da/ and /ga/ which had followed the natural tokens of /al/ and /ar/. The stimuli differed only in the onset of the third formant, which ranged from 2690 to 2104 Hz in six approximately equal steps. Onset values for the first and second formants were fixed at 310 and 1588 Hz, respectively; steady-state values for the first three formants were 649, 1131 and 2448 Hz, respectively and all formant transitions were step-wise linear and 100 ms in duration. Total duration was 180 ms and amplitude and pitch contour reflected those in the natural tokens of /da/ and /ga/.

As in Experiment I, there were three stages. In the first, 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/ with a 50 ms silent interval separating the VC and CV portions. Each CV stimulus was presented 12 times in each context (4 times per token), according to a completely 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 a subject's ability to identify /l/ and /r/ in the same test stimuli that had previously been employed in the second stage of testing; it followed the study of the context effect to assure that the ability of the Japanese subjects was not underestimated due to inadequate exposure to the test materials. In light of the potential difficulty of this task, which was to mark on a response sheet written in alphabetic script whether a given stimulus contained /al/ or /ar/, listeners were presented with 28 items in which /l/ and /r/ alternated, followed by a randomized practice sequence of 28 items for which they were told the correct responses. The session concluded with their labelling of the randomized test sequence without response feedback.

Results

An assessment of subjects' ability to identify /l/ and /r/ is critical to interpretation of the results of Experiment II, therefore the analysis of the data began with those responses obtained in the third stage of the experiment in which subjects labelled the test stimuli as containing /al/ or /ar/. All of the native speakers of English were 100% correct, however, there was considerable individual variation in the performance of the Japanese subjects: It was apparent, as others have noted (MacKain, Best, & Strange, 1981), that at least

some native speakers of Japanese can master the /l/-/r/ distinction, but also, as noted previously (Mochizuki, 1981), that the identification of /l/ and /r/ often poses a problem when these form part of a consonant sequence. In order to assess the relation between the accuracy of phonemic labelling and sensitivity to the context effect of /l/ and /r/, two extreme groups of Japanese subjects were selected for further analysis. They included the eight inferior students of spoken English whose performance was within 15% of chance (i.e. 50%, $p > .1$) and the eight superior students whose labelling of /l/ and /r/ was 98% accurate or better. Average performance for the inferior students was 58%, which did not significantly differ from chance ($p > .1$), whereas average performance for the superior students was 98%. Relevant background information appears in Table 2, where it can be seen that one distinguishing trait of the superior students of spoken English was more intensive exposure to spoken English prior to and during junior high school.

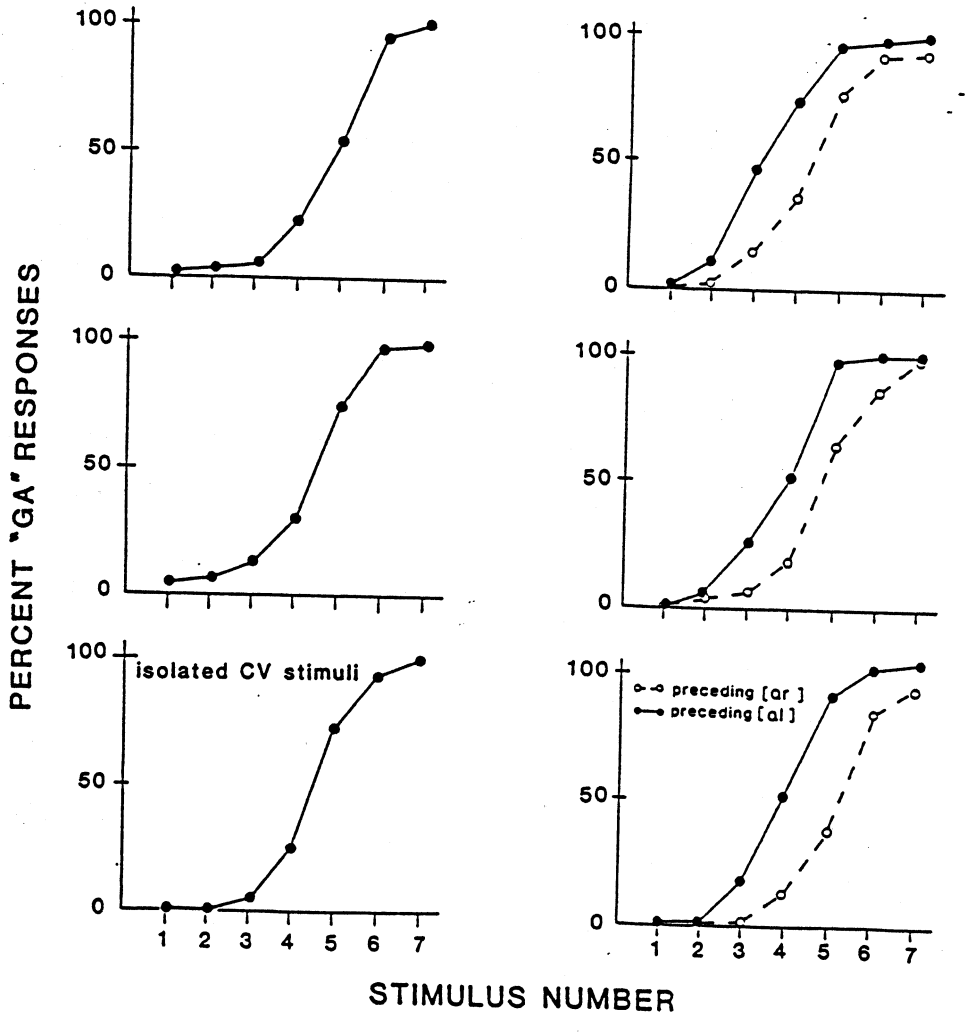
For these two groups of Japanese speakers, and for the native speakers of English, the context effect of /al/ vs. /ar/ can be seen in Figure 2. That figure summarizes the results obtained in the first and second stages of testing in terms of the mean percent of "ga" responses given to stimuli at each position along the continuum. From top to bottom, the panels represent the results of the native speakers of English, the native speakers of Japanese who could label /al/ and /ar/, and the native speakers of Japanese who could not label /al/ and /ar/ correctly. Panels on the left concern labelling of the isolated /da/-/ga/ stimuli, whereas those on the right concern the context effect of /al/ and /ar/, averaged across the six tokens of each syllable (i.e. ignoring the possible trading relation).

Table 2. *Oral English profile: Japanese subjects participating in Experiment II*

	College testing		Experience prior to college[a] (0–5 pt. scale, 5 = extensive experience)				
	Jacet-Coltd (max. 120)	Koike (max. 50)	Before Jr. high	Jr. high		Sr. high	
				School	Home	School	Home
Superior students	97.5	47.6	1.75	4.38	3.38	3.00	1.13
Inferior students	37.5	40.5	0.05	2.50	1.75	1.63	1.00

[a] Ratings computed by the students' English professor on the basis of responses to a questionnaire.

Figure 2. *The pattern of "ga" responses given to stimuli along an acoustic /da/-/ga/ continuum when the stimuli were presented in isolation (left panels), and when they were preceded by /al/ and /ar/ (right panels). From top to bottom, subjects include: (1) native speakers of English who are 100% correct in identifying /l/ and /r/, (2) native speakers of Japanese who are 99% correct in labeling /l/ and /r/, and (3) native speakers of Japanese who perform at chance level in labeling /l/ and /r/.*



Inspection of the left-hand panels of Figure 2 will reveal that, as in Experiment I, responses to the isolated /da/-/ga/ stimuli varied as a function of stimulus position along the continuum ($F(6.138) = 905.79$, $p < .0001$), and the native speakers of Japanese did not differ from each other, or from the native speakers of English ($p > .1$). Let us now turn to the context effects of /al/ vs. /ar/ as they appear in the right-hand panels. Note that the context effect is the same for all three subject groups. Most important are the data obtained from the inferior students of spoken English (lower panel), as they indicate that the context effect can occur in the absence of correct identification of /al/ and /ar/, as such. These subjects had been selected for their inability to phonetically identify /al/ and /ar/, and their average performance on that task was not significantly better than chance (i.e. 58% correct, $p > .1$). Nonetheless they showed a context effect of /al/ vs. /ar/ which was equivalent in direction and magnitude to that of the other subjects. An ANOVA was computed on the total number of "ga" responses across the continuum as a function of subject group, the identity of the preceding liquid (the context effect), and whether the preceding syllable had originally been produced before /da/ or /ga/ (the trading relation). There was a significant main effect of subject group, due to the inferior students of spoken English having given fewer "ga" responses than subjects in the other groups ($F(2.23) = 13.53$, $p < .0001$), and a main effect of /l/ vs. /r/ ($F(1.23) = 108.23$, $p < .0001$), but no interaction between these two factors.

In addition to the effects shown in Figure 2, there was also evidence of a trading relation, analogous to that observed in Experiment I, and like that observed by Mann (1980). That effect was not due to whether an utterance contained /l/ vs. /r/, but reflected the nature of the original utterances from which the tokens of /al/ and /ar/ had been drawn. In general, tokens drawn from /al-ga/ and /ar-ga/ gave rise to more "ga" percepts than those drawn from /al-da/ and /ar-da/ ($F(1.23) = 9.93$, $p < .004$), and this interacted with stimulus number ($F(6.138) = 2.38$, $p < .03$). Yet here, again, there was no interaction with subject group ($p > .1$). All three groups of subjects, then, exhibited both a context effect due to /l/ vs. /r/ and a trading relation due to whether the token of /al/ or /ar/ had originally been produced before /da/ vs. /ga/. The direction and extent of each effect was constant across subject groups.

General discussion

This study has asked whether limited experience with spoken English limits the sensitivity of native speakers of Japanese to certain effects which can

occur in the perception of English utterances. Like English, Japanese distinguishes the phonemes /d/ and /g/, /s/ and /ʃ/, but in contrast to English, Japanese does not permit sequences of consonants to occur. That a lack of familiarity with consonant sequences does not alter the perceptual interactions which occur when stimuli along a /da/-/ga/ continuum are preceded by /s/ and /ʃ/, was revealed by the results of Experiment I. Japanese further differs from English in that it does not distinguish the phonemes /l/ and /r/, and this has been observed to cause identification and discrimination of these phonemes to be difficult for many native speakers of Japanese. However, problems with the identification of /l/ and /r/ notwithstanding, Experiment II revealed that Japanese listeners are sensitive to the perceptual effects which occur when /da/-/ga/ stimuli are preceded by /al/ and /ar/.

Specifically, Experiment II showed that, in stimuli which native speakers of English label as /al-da/, /al-ga/, /ar-da/ and /ar-ga/, native speakers of Japanese respond to some difference between utterances that contain /l/ and /r/ which influences their perception of /d/ and /g/ whether or not they can identify the /l/ or /r/ as such. One implication is that Japanese listeners are sensitive to some difference between English utterances that contain /l/ and /r/. If it is accepted that the contrasting effect of syllables ending in /l/ and /r/ on the perception of a following syllable as /da/ or /ga/ is specific to speech perception (Mann & Liberman, 1983) and reflects listeners' sensitivity to the acoustic consequences of certain articulatory interactions (Mann, 1980), then a further implication is that speech perception comprises at least two levels: one universal, the other language-dependent.

At the first level, listeners process speech signals as if they are tracking articulatory movements in a more-or-less objective fashion. Beyond that, there is a level (or levels) at which the speech signal is somehow categorized into language-dependent segments which conform to the phonological rules governing permissible phonemes and sequences of phonemes in a given language. It is the first level that is most 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 acoustic consequences of the different articulatory properties of /l/ and /r/ whether or not they can directly categorize those acoustic consequences in terms of segments in a given phonological inventory. Apparently the representation of utterances in correspondence with their underlying articulatory gestures cannot intervene directly upon consciousness, otherwise all Japanese listeners would be able to draw upon that level and thus solve the /l/-/r/ problem in perception and production. Nonetheless, it is interesting

to note that some of the more successful methods of teaching Japanese speakers how to distinguish /l/ and /r/ do involve explicit reference to the articulation of /l/ and /r/ (cf. Sheldon & Strange, 1982).

The language-independent perception of articulatory gestures, which is accomplished at the universal level, may precede the language-dependent perception of an utterance and it is this which allows listeners to be sensitive to the acoustic consequences of movements which they fail to properly categorize as one phonological segment or another. Segmenting an utterance into phonemes or syllables would appear to depend upon language experience, hence listeners may encounter difficulty when they are required to identify speech sounds that are not in their native inventory. Such was the case when the Japanese listeners who were inferior students of spoken English were asked to label /al/ and /ar/ in Experiment II. However, in the same experiment, acoustic differences between /al/ and /ar/ which reflect their different articulatory origins nevertheless influenced both Japanese and English listeners' perception of a following phoneme as /d/ or /g/. Hence responses which are directly mediated by a language-dependent level can still be influenced by processing at a universal level of speech perception.

This distinction between universal and language-dependent levels of speech perception is reminiscent of Werker and Tees' (1984a and b) suggestion that there is a level of processing intermediate between what is traditionally called phonetic processing, and that which is traditionally called nonspeech acoustical processing. In the Werker and Tees (1984b) account, the intermediate, universal level is termed "phonetic" as it is postulated to correspond to natural phonetic boundaries, whereas that higher level which corresponds to native language boundaries is termed "phonemic". It is further postulated that the phonetic level of perception is not altered by native language experience, but native language experience is directly responsible for the development of the phonemic level. Finally, the universal phonetic level dominates the speech perception behavior of infants whereas the language-dependent phonemic level tends to dominate the speech perception behavior of adult subjects.

As regards the perception of /l/ and /r/, it can be reasoned that, since four-month-old infants distinguish utterances that contrast /l/ and /r/ (Eimas, 1975), and since infants at this age should be relying on the phonetic level of processing, perception of the /l/ and /r/ distinction most likely begins in ontogeny as a universal phonetic ability. The subsequent perceptual behavior of infants who go on to become adult speakers of Japanese would come to be dominated by a phonemic level of processing which accords with the phonemic inventory of Japanese, whereas that of infants who become adult speakers of English would be dominated by a phonemic level of processing

that is appropriate to English. For the speakers of Japanese, in particular, a lack of experience with utterances which contrast /l/ and /r/ would prevent the /l/-/r/ distinction from being represented at the phonemic level, but a lack of experience would not change the processing of this distinction at the phonetic level. Japanese adults would appear unable to distinguish /l/ and /r/ because their perceptual behavior is dominated by a phonemic level of processing where the /l/-/r/ distinction is not represented.

The present results might conform with the Werker and Tees account were it to be postulated that the context effect of /al/ and /ar/ operates at the universal "phonetic level" of processing, whereas the identification of /al/ and /ar/ is accomplished at the language-dependent "phonemic level". The question which arises, however, is whether "phonetic" and "phonemic" are appropriate characterizations of the universal and language-dependent levels of speech perception implicated herein. The use of these terms by Werker and Tees corresponds to a distinction between universal phones and language-dependent phonemes as theoretical constructs postulated in the fields of phonetics and phonology (see, for example, Sommerstein, 1977). Both are abstract units whereby the nearly continuous acoustic speech signal can be parsed into a linear string of discrete segments consisting of matrices of "phonetic features" defined in articulatory (e.g., velar place of articulation) and/or acoustic terms (e.g., sonorant). Universal phones are language-independent constructs which comprise the exhaustive set of all possible phonetic feature matrices that can be produced by human speakers, whereas phonemes are language-dependent constructs which comprise a reduced set of the possible matrices of phonetic features which distinguish the utterances of a given language, specifying only those features which serve to keep the utterances of that language apart.

Is it appropriate to refer to these constructs and the features they subsume in order to account for the present findings? Interpretation of many results, including the present ones, can be enhanced by reference to such features as place of articulation, yet this need not imply that features, as such, play a role in speech perception. Indeed, the role of phonetic features in speech perception remains debatable, at best (see Diehl, 1981; Soli & Araby, 1979; Soli, Araby & Carroll, in press). The results of the present study do not argue for the perception of phones or phonemes so much as for the existence of some universal level of speech perception which influences processing at the language-dependent level (or levels) that normally mediates speech labelling responses. At the language-dependent level, native speakers of Japanese might represent /l/ and /r/ as one and the same, perhaps as allophones of Japanese /r/. But it remains unclear as to whether the perceptual representation is in terms of phonemes, as opposed to some other language-dependent

unit. Following Sapir's (1933) suggestion, we might entertain the possibility that, in most speech perception experiments, subjects' responses appear to be guided by a phonemic level of processing; however, it is also plausible that their responses might reflect a syllabic (Savin & Bever, 1970) or even a lexical level (an influence of lexical knowledge in speech perception is apparent from the results of Ganong, 1980, for example). Owing to present uncertainty as to whether the perceptual process computes phonemes, syllables, etc., perhaps the language-dependent level of perception can only be regarded as corresponding to some phonological category or categories. In any event, we might eschew some of the problems which arise from linguistic theory's failure to identify a "phonemic" representation of the sort postulated by classical linguists (cf. Chomsky & Halle, 1968) by avoiding that term while leaving open the question of whether phonemes are perceived as such.

The main contribution of the present results to our knowledge of speech perception is the demonstration that Japanese listeners can perceive some difference between utterances of /l/ and /r/ which influences their perception of an adjacent /d/ or /g/. Even though such listeners cannot label /l/ and /r/ as such, they perceive a difference which relates to certain articulatory properties of the gestures that convey /l/ and /r/. The fact that the influence of /l/ vs. /r/ on perception of the /da/-/ga/ distinction is a contrastive context effect, and not an assimilative trading relation, suggests that the Japanese listeners are perceiving /l/ and /r/ as segments of the speech signal which are distinct from the adjacent /d/ or /g/, and this is at least consistent with the possibility that /l/ and /r/ are heard as phones. But there is no confirmation that perception involves phones, as such. Ultimate characterization of both the universal and language-dependent levels of speech perception may have to await further clarification of the nature of phonological segments much less the very existence of a discrete set of universal articulatory phonetic features (cf. Ladefoged & Bhaskararao, 1983).

In summary, the present results suggest that, in addition to a level or levels where speech perception corresponds to the phonological categories of a given native language, there exists a universal level where speech perception corresponds more objectively to the articulatory gestures that give rise to speech signals. It is encouraging to note that this view is consistent with what is known about the speech perception capabilities of prelingual infants. Infants tested within the first six months of life have given evidence of perceiving many phonetically-relevant properties of speech (see, for a review, Eilers, 1980) and have also given evidence of perceptual trading relations that may be based in articulatory interactions (Miller & Eimas, 1983). At present, in the absence of any means of verifying that infant listeners perceive phones or phonemes, as such (Best, in press; MacKain, 1982; Werker & Tees, 1984a)

it may be premature to regard them as capable of phonetic or phonemic perception. Yet the data surely imply the existence of some perceptual abilities that are the basis of adult speech perception (Miller & Eimas, 1983), and one of these could well be the ability to perceive speech in some manner that corresponds to the vocal tract movements that give rise to speech signals. This would be consistent with some recent evidence (Kuhl & Meltzoff, 1982) that infants as young as four months are predisposed to integrate the sight of a talker with the acoustic speech stream in a way that shows sophistication about the visible and audible consequences of articulation, perhaps owing to special proclivities of the left or dominant hemisphere (MacKain, Studdert-Kennedy, Spieker, & Stern, 1983) which mediates speech perception in adults (Studdert-Kennedy & Shankweiler, 1970).

The speech perception behavior of infants, then, could be dominated by the postulated "universal level" of processing in contrast to that of adults, which tends to be dominated by the "language-dependent" level, and this is in keeping with the spirit of Werker and Tees' (1984a and b) proposal. Yet, in adult subjects, a universal level of speech processing may still remain, as indicated by the Japanese listeners' perceptual sensitivity to the effects of /l/ vs. /r/ on the /da/-/ga/ distinction. From the perspective of a native speaker of Japanese, the context effect of /l/ vs. /r/ involves the influence of a non-native contrast on perception of a contrast that is part of the native phonemic inventory. The effect itself operates at a universal level of processing where the lawful relationship between acoustic speech stimuli and articulatory gestures is captured, and it can be measured through the pattern of responses that listeners give as they report which phonemes of their language a given utterance reflects.

References

Beckman, M., & Shoji, A. (1984). Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese. *Phonetica, 41*, 61–71.

Best, C.T. (in press). Discovering messages in the medium: Speech and the prelingual infant. To appear in H.E. Fitzgerald, M. Yogman & B. Lester (Eds.), *Advances in behavioral pediatrics: Theory and research, Volume 2*.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Delgutte, B., & Kiang, N.Y. (1984). Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. *Journal of the Acoustical Society of America, 75*, 897–907.

Diehl, R.L. (1981). Feature detectors for speech: A critical reappraisal. *Psychological Bulletin, 89*, 1–18.

Eilers, R. (1980). Infant perception: History and mystery. In G.H. Yeni-Komshian, J.F. Kavanaugh & C.A. Ferguson (Eds.), *Child phonology: Perception and production, Volume 2*. New York: Academic Press.

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.

Elliot, L.L. (1971). Backward and forward masking. *Audiology, 10*, 65–76.
Foss, D.J., & Blank, M.A. (1980). Identifying the speech codes. *Cognitive Psychology, 12*, 1–31.
Foss, D.J., & Gernsbacher, M.A. (1983). Tracking the dual code: Towards a unitary model of phone identification. *Journal of Verbal Learning and Verbal Behavior, 22*, 609–632.
Ganong, W.F., III (1980). Phonetic categorization in auditory word perception. *Journal of Experimer Psychology: Human Perception and Performance, 6*, 110–125.
Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "R" and "L". *Neu psychologia, 9*, 317–323.
Harris, D.M., & Dallos, P. (1979). Forward masking of speech by the auditory nerve system. *Journal Neurophysiology, 42*, 1083–1107.
Jacet-Coltd (1978). *Technical manual for JACET-COLTD Listening Comprehension Test, Form A*. Japa -Kaitakusha.
Koike, I. (1978). *Gaikokugo to shiteno Eigo no "hearing" nouryoku keisei youin no jishouteki kenku (I)*. Ke University: Department of Economics.
Kuhl, P.K., & Meltzoff, A.N. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138–1144.
Ladefoged, P., & Bhaskararao, P. (1983). Non-quantal aspects of consonant production: A study of retrofle consonants. *Journal of Phonetics, 11*, 291–302.
Liberman, A.M., Isenberg, D., & Rakerd, B. (1982). Duplex perception of cues for stop consonants: Evidenc for a phonetic mode. *Perception & Psychophysics, 30*, 133–143.
Liberman, A.M., & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition, 2* 1–36.
Lisker, L., & Abramson, A. (1970). The voicing dimensions: Some experiments in comparative phonetics. I *Proceedings of the 6th International Congress of Phonetic Sciences*. Prague: Academia.
MacKain, K.S. (1982). Assessing the role of experience on infants speech discrimination. *Journal of Chil Language, 9*, 527–542.
MacKain, K.S., Best, C.T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*, 369–390.
MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, S. (1983). Infant intermodal speech perception is a left hemisphere function. *Science, 219*, 1347–1349.
Mann, V.A. (1980). Influence of preceding liquid on stop consonant perception. *Perception & Psychophysics, 28*, 407–412.
Mann, V.A., & Liberman, A.M. (1983). Some differences between phonetic and auditory modes of percep- tion. *Cognition, 14*, 211–235.
Mann, V.A., & Repp, B.H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America, 69*, 548–558.
Miller, J.L., & Eimas, P.D. (1983). Studies on the categorization of speech by infants. *Cognition, 13*, 135–166.
Miller, G.A., & Nicely, P.E. (1959). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27*, 338–352.
Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J., & Fujimura, O. (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 Phonetics, 9*, 283–303.
Moore, B.C.J. (1978). Psychophysical tuning curves measured in simultaneous and forward masking. *Journal of the Acoustical Society of America, 63*, 524–532.
Rand, T.C. (1974). Dichotic release from masking for speech. *Journal of the Acoustical Society of America, 55*, 678–680.
Repp, B.H. (1982). Phonetic trading relations and context effects: New evidence for a phonetic mode of perception. *Psychological Bulletin, 92*, 81–110.

Repp, B.H., Liberman, A.M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 621–637.

Repp, B.H., & Mann, V.A. (1981). Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America, 69*, 548–558.

Repp, B.H., & Mann, V.A. (1982). Fricative-stop coarticulation: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 71*, 1562–1567.

Sapir, E. (1933). La réalité psychologique des phonèmes. *Journal de Psychologie Normale et Pathologique, 30*, 247–265. Translated as "The psychological reality of phonemes", in Mandelbaum, (Ed.) *Selected Writings of Edward Sapir*. Berkeley: University of Ca. Press, 1949.

Savin, H.B., & Bever, T.G. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior, 9*, 295–302.

Scharf, B. (1970). Critical bands. In J.V. Tobias & E.D. Schubert (Eds.) *Foundations of Modern Auditory Theory, Vol. 1*. New York: Academic Press.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede perception. *Applied Psycholinguistics, 3*, 243–261.

Smith, R.L. (1977). Short-term adaptation in single auditory-nerve fibers: Some post-stimulatory effects. *Journal of Neurophysiology, 40*, 1098–1112.

Soli, S.D., & Araby, P. (1979). Auditory vs. phonetic accounts of observed confusions between consonantal phonemes. *Journal of the Acoustical Society of America, 66*, 46–59.

Soli, S.D., Araby, P., & Carroll, J.D. (in press). Representation of discrete structure underlying observed confusions between consonantal phonemes. *Journal of the Acoustical Society of America*.

Sommerstein, A.H. (1977). *Modern Phonology*. Baltimore: University Park Press.

Studdert-Kennedy, M., & Shankweiler, D. (1970). Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America, 48*, 579–594.

Trehub, S. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development, 47*, 466–472.

Werker, J.F., Gilbert, J.H.V., Humphrey, K., & Tees, R.C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52*, 349–355.

Werker, J.F., & Tees, R.C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.

Werker, J.F., & Tees, R.C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75*, 1866–1878.

Résumé

Les locuteurs natifs du japonais sont incapables d'identifier correctement les phonèmes /l/ et /r/ de l'anglais. Pourtant, on peut montrer qu'ils sont capables de réagir comme s'ils étaient sensibles aux gestes articulatoires différents qui sont nécessaires pour produire /l/ et /r/. Dans une étude, des locuteurs natifs du japonais et des locuteurs natifs de l'anglais devaient classer des stimuli le long d'un continuum /da/-/ga/ lorsque le stimulus était précédé par des occurrences naturelles de /s/ ou /ʃ/, de /al/ ou /ar/. Chaque paire de "prédécesseurs" avait des effets différents sur l'emplacement de la frontière catégorielle entre /da/ et /ga/, et ni la direction ni l'étendue de l'effet ne dépendait de l'expérience linguistique. De manière intéressante, /al/ donnait naissance à plus de percepts de /ga/ que /ar/, à la fois pour les locuteurs japonais et anglais, indépendamment de leur aptitude à identifier /al/ et /ar/ en tant que tels. L'interprétation des résultats repose sur des observations plus anciennes selon lesquelles les effets perceptuellement contrastifs de /al/ vs. /ar/ et de /s/ vs. /ʃ/ ont un correspondant dans la structure acoustique des énoncés naturels de /al-da/, /ar-da/, etc. du fait de la co-articulation du geste qui produit la consonne qui précède et de celui qui produit le /ga/ ou /da/ suivant. Il semblerait que les

locuteurs natifs du japonais soient sensibles aux conséquences acoustiques de la co-articulation de /l/ ou /r/ avec /d/ ou /g/, alors qu'ils sont incapables de catégoriser /l/ et /r/ en tant que phonèmes distincts. Il se pourrait donc qu'il existe, en deçà du niveau de perception propre au langage dans lequel les sons linguistiques sont représentés conformément aux contraintes d'un système phonologique, un niveau universellement partagé où la représentation des sons linguistiques correspond de plus près aux gestes articulatoires qui donnent naissance au signal linguistique.