Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners

John C. L. Ingram and See-Gyoon Park
Department of English, University of Queensland, Brisbane 4072, Australia

(Received 12 December 1995; revised 12 September 1997; accepted 6 October 1997)

Japanese and Korean listeners' identification and discrimination of English /r/ and /l/ were compared using a common set of minimal pair stimuli. The effects of speakers (two native speakers of Australian English), position of the contrast within the word (word initial, initial consonant cluster, and medial positions), and listening task (forced choice identification versus oddball discrimination) were examined, with a view to assessing the relative importance of language-specific and language-independent factors operating at the acoustic–phonetic and phonological levels of signal processing in “foreign sound” speech perception. Both prior phonological learning and the relative acoustic discriminability of the items affected subjects’ performance on the identification test. Where both factors were engaged, phonological learning effects predominated over the effects of acoustic discriminability. The extent to which a speaker encoded critical acoustic cues for the /r–l/ distinction was found to affect /r–l/ identification. Dynamic spectral features known to be relevant for the /r–l/ contrast were effective in predicting (in a linear regression analysis) speaker-dependent differences in identification scores. Although the discrimination test may have been influenced by ceiling effects, the performance profiles on the identification and discrimination tests were quite different, indicating that the identification and discrimination tests imposed quite different task demands upon listeners and that phonological processing of the signal was more engaged by the former task. © 1998 Acoustical Society of America. [S0001-4966(98)05301-6]

PACS numbers: 43.71.Hw, 43.71.Es, 43.71.Bp, 43.66.Lj [WS]

INTRODUCTION

Difficulties that Japanese and others, whose native language has only a single phonemic liquid, encounter with the English /l–r/ contrast are so well known as to have become a linguistic stereotype. Phonetic studies of this phenomenon have raised, but not resolved, a number of fundamental questions for understanding how novel phonemic contrasts are acquired in second language learning; such as whether beyond a certain age of initial contact with English, the /l–r/ contrast can be reliably perceived acquired in second language learning; such as whether beyond a certain age of initial contact with English, the /l–r/ contrast can be reliably perceived (Goto, 1971; Mochizuki, 1981), though it may be consistently realized in production (Sheldon and Strange, 1982); whether phonetic training can facilitate acquisition of the /l–r/ contrast (Logan et al., 1991; Lively et al., 1992, 1993); how language-specific transfer effects influence acquisition of the /l–r/ contrast (Gillette, 1980; Sheldon and Strange, 1982; Henry and Sheldon, 1986), or whether language-independent factors, based upon acoustic phonetic considerations or phonological markedness (Broselow and Finer, 1991; Eckman and Iverson, 1993) play a significant role in acquiring the /l–r/ contrast in different phonological environments in English.

It seems likely that perceptual responses of non-native listeners to foreign speech sounds reflect the simultaneous operation of factors at various levels of signal analysis from the basic acoustic to the linguistic, and that the relevance of prior learning of the first language (L1) may vary depending on parameters of the listening task. It may be that precedence relations operate between different classes of effects in second language speech perception or production. This was proposed by Henry and Sheldon (1986) to account for the relative difficulty that Cantonese listeners experienced with the /l–r/ contrast in syllable codas, an environment that Japanese listeners find relatively easy, presumably because spectral cues for distinguishing /l/ and /r/ syllable finally are distributed over a longer time frame and are hence more discriminable than in other environments. Cantonese (unlike Japanese) has a “clear” (nonvelarized) /l/ in final position and this evidently interferes with the perception of the velarized English /l/, whereas no such interference effects operate for Japanese listeners.

A priori, four major classes of influence may be identified as relevant for the perception of a novel phonemic contrast by second language learners. These are indicated in Table I.

Language specific or language universal influences may express themselves at either the phonetic or the phonological levels of speech processing, and all four combinations of these two factors have found their way into contemporary accounts of interlanguage speech perception. Phonologically oriented theorists, such as Broselow and Finer (1991) who advocate a Parameter Setting model of acquisition, or Markedness theorists, such as Eckman and Iverson (1993), seek to invoke language universal principles of phonological organization to explain patterns of difficulty in mastering L2 phonological contrasts. Phonetically oriented theorists, such as Mann (1986), make an appeal to language universal phonetic constraints, which have their origin in the way articulatory gestures are coded into the acoustic signal of speech.

However, both phonetic and phonological effects on
second language speech perception may be language-specific in their operation. The traditional account of interference effects, expressed in the contrastive analysis hypothesis, makes an appeal to structural differences in the sound patterns of L1 and L2, and therefore, to language-specific phonological properties of the languages concerned. But it is also possible that language-specific phonetic influences may play an important role in interlanguage speech perception. For example, perceptual parsing strategies, that operate at a prephonological level of signal processing to aid in the detection of word or morpheme boundaries, and which vary with the language background of the listener, have been found to operate in both monolingual and bilingual speech perception (Cutler et al., 1986, 1992).

In order to understand how these potentially competing influences affect interlanguage speech perception, it is necessary to have some way of separating factors operating at different levels of analysis, specifically the phonological and the phonetic, and for distinguishing the domain or scope of their operation, whether they are language particular or universal in their application. Only comprehensive cross-language studies comparing speakers of diverse linguistic backgrounds will enable us to distinguish language-specific from universal influences, with the relevant data accumulated over a series of cross-language experiments.

In the present study, phonetic and phonological transfer effects on /l–r/ perception (the first type of effect mentioned in Table I above) were examined by comparing two groups of Japanese and Korean learners of Australian English. Korean and Japanese phonologies differ in ways that are potentially significant for acquisition of the English /l–r/ contrast. These differences enabled the prediction of the pattern of perceptual errors in the two language groups across the three phonological environments where the /l–r/ contrast occurs in Australian English, namely, as single consonants in syllable or word onset position (red–led), as second elements of an onset consonant cluster (pray–play), and in intervocalic position (berry–belly). (Being a nonrhotic dialect, there is no /l–r/ contrast in coda position in Australian English.)

Previous studies have found that /l–r/ perception is more difficult in consonant clusters than other environments (Goto, 1971; Gillette, 1980; Sheldon and Strange, 1982; Lively et al., 1992). This has been attributed to the lower acoustic discriminability of phonemic contrasts within consonant clusters, where segments tend to be shorter, and masking effects due to gestural overlap of adjacent segments are greater. In our experiment, this language-independent effect (type 2 in Table I) was found to operate in an additive manner with transfer effects mentioned above.

Another consideration which is relevant for the acoustic discriminability of phonemic contrasts in L2, concerns pronunciation variability and the adequacy of tokens provided by different L2 speakers. Recent studies have shown that exposing learners to a range of speaker models during the perceptual training phase enhances retention of the /l–r/ contrast and facilitates generalization to new items by Japanese learners of English (Logan et al., 1991; Lively et al., 1992, 1993). In the present study, two male native speakers of Australian English provided stimulus tokens for the identification and discrimination experiments. Speaker effects were also found to be an important source of variance in the perceptibility of the /l–r/ contrast, and subsequent spectrographic analysis indicated an acoustic basis for the differential discriminability of certain tokens over others, in terms the clarity of encoding of key acoustic properties known to be important for distinguishing /l/ and /r/.

However, language-independent effects may not be confined to factors affecting discriminability and basic auditory processing of speech signals. Considerations of phonological markedness (factor 3, Table I) may constitute a separate class of language-universal constraints on phonological restructur- ing or accommodation to L2 sound patterns. Markedness constraints were originally proposed to remedy perceived inadequacies of contrastive phonological analyses to account for error patterns in second language speech production (Eckman, 1977). In recent years, markedness constraints have been incorporated into the principles and parameters of Universal Grammar (Chomsky, 1986) that supposedly shape the course of first, and possibly also, second language learning. The role of parameter setting in second language learning is controversial, particularly in the phonological domain, where the competing view holds that language-dependent transfer effects exert a much stronger influence (Sato, 1984). Phonological markedness constraints gain expression in cross-language preferences for segment selection and syllable structure. No clear case can be made for the relative markedness of the liquids /l/ or /r/ (Bhat, 1974; Maddieson, 1984). However, it was possible to assign markedness values to items for the range of consonant clusters included in this experiment, based upon proposals by Broselow and Finer (1991) and Eckman and Iverson (1993).

Finally, the perceptual demands of the listening task (factor 4, Table I) exercise a degree of control over listeners’ perceptual responses and the foregrounding or backgrounding of factors mentioned above. Mochizuki (1981) found that Japanese listeners’ ability to discriminate and label minimal pair /l–r/ contrasts presented in triadic comparisons was highly dependent on the echoic memory load induced by the

### Table I. Factors affecting perception of novel phonemic contrasts.

<table>
<thead>
<tr>
<th>Type</th>
<th>Label</th>
<th>Source of effect or influence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Transfer effects</td>
<td>Language specific, phonetic, or phonological learning.</td>
<td>A language-dependent factor.</td>
</tr>
<tr>
<td>3. Phonological markedness or parameter setting</td>
<td>Universal properties of sound systems and the human language acquisition device.</td>
<td>Universal defaults, but language dependent settings.</td>
</tr>
<tr>
<td>4. Task variables</td>
<td>Perceptual or information processing demands of the listening task, e.g., identification versus discrimination.</td>
<td>May modulate the effects of factors 1–3 above.</td>
</tr>
</tbody>
</table>
response task. Werker and colleagues (Werker and Logan, 1985; Werker, 1994) have manipulated interstimulus intervals (ISIs) to control the amount of phonological processing required of listeners to hold items in echoic memory in a discrimination task. Identification tasks, which involve matching the stimulus to internally stored representations, typically impose greater echoic memory loads than discrimination tasks. More generally, an identification task may be expected to engage linguistic processing of the speech signal more fully than a discrimination test. In their classical study of discrimination and identification of synthetic vowel sounds, Stevens et al. (1969) found that Swedish and English subjects exhibited similar performance on vowel discrimination tests, though they yielded somewhat different identification functions, related to phonemic status of the lip rounding.

Lexical familiarity effects have been observed in /l–r/ identification by inexperienced and experienced Japanese learners of English (Yamada and Tohkura, 1992b; Flege et al., 1995), though such effects are likely to be observed in native listeners only under marginal listening conditions or when the stimuli are phonetically ambiguous (Ganong, 1980). It may be hypothesized that lexical familiarity effects will be stronger in an identification task with non-native listeners than in a discrimination task with the same listeners. Such a finding would support the hypothesis (above) that identification tasks engage higher-order linguistic processing of the stimulus to a greater extent than discrimination tasks. Such processing is likely to involve language-specific learning, and consequently, an identification task is also more likely to reflect the influence of the listener’s language background than a discrimination task.

Notwithstanding the above, Mann (1986) found that Japanese listeners were sensitive to the acoustic consequences of coarticulation effects in the discrimination of synthetic /l–r/ tokens, in a manner similar to native English speakers, even though her subjects were unable to consistently identify /r/ and /l/ as phonological targets. This suggests that a language-independent level of phonetic processing at which articulatory constraints govern speech signal processing may be engaged by a basic discrimination task, even when perceiving “foreign sounds” or incompletely acquired phonemic contrasts. Consequently, the impact of variation of task demands on perceptual performance are difficult to predict. Nevertheless, comparisons of performance on identification and discrimination tasks may throw light upon the perceptual mechanisms involved in acquiring a novel phonemic contrast.

To summarize: In the identification and discrimination experiments of the /l–r/ contrast by Japanese and Korean learners of English reported below, we sought to bring together a range of potentially competing effects operating at different levels (the phonetic and the phonological) and in different domains (the language specific and the language universal) to observe their relative importance and interaction. We argue that the results suggest, (1) an important role for language particular phonetic and phonological factors which reflect the operation of prior perceptual learning, (2) no role for language universal phonological markedness or parameter setting, and (3) mixed evidence for the operation of language universal factors operating at a phonetic or prephonological level speech signal processing, depending on the source of phonetic variability and how the term “phonetic level of processing” is construed.

I. JAPANESE AND KOREAN LIQUIDS

Japanese possesses a syllable initial liquid, usually phonemized /l/, which has a range of phonetic variants that have been variously described by linguists as “an alveolar flap,” “a retroflexed flap,” a “combination of an alveolar lateral and a retroflexed flap,” a “lax alveolar stop,” and an “alveolar lateral” (Tsuzuki, 1992). There appears to be a wide range of phonetic variation for the Japanese liquid, from a slightly retroflexed lateral approximant, through to an alveolar tap or flap. This phonetic variation occurs across speakers and does not appear to be strongly conditioned by phonological environment. The Japanese liquid is confined to syllable onset position, word initially, or word internally.

Korean also has only one phonemic liquid with a similar range of phonetic variation as its Japanese counterpart. However, the Korean liquid may occur in coda as well as onset position, where there are clearly phonologically conditioned allophones. Syllable initially, the Korean liquid is phonetically realized as a flap [ɾ]. There is, however, a strong sanction against liquids appearing in absolute word initial position in Korean. Word initial liquids are found only in loan words, not in items of native vocabulary. In this special environment, there is, like Japanese, a broader range of phonetic variation, from a flap or tap to a lateral approximant. In coda position, the Korean approximant is consistently realized as a lateral approximant.

Intervocically, the singleton liquid is phonetically realized as an alveolar flap. In geminates (where a liquid in coda position is followed by another in onset position in the next syllable), an alveolar lateral is invariably produced. Thus Korean speakers have a phonetic model available to them for English /l–r/ contrasts, in nonword initial position. They can treat English /l/’s as “singleton” liquids, and /r/’s as geminate liquids, and thereby obtain a phonetic approximation ([ɾ][ɾ]) for the English /r–l/ contrast, as is attested by the phonetic structure of English loan words such as:

- *orange* /’oʊrɪn/ (origin)
- *olympic* /’ɒlɪmpɪk/ (Olympic)

However, this model cannot be invoked for word initial position, where, not only is the contrast between the geminate and the singleton liquid unavailable, but any kind of liquid violates canonical native word forms. For Koreans, /l–r/ perception in word initial position would be expected to be doubly difficult, and more so than for Japanese listeners.

With respect to the /l–r/ distinction in onset clusters, Korean phonotactic constraints are practically identical to Japanese. Neither language permits C+liquid clusters and English loan words containing such clusters are typically resyllabified by vowel epenthesis. However, once resyllabification has occurred, Korean speakers will have available to them the phonetic strategy which they employ in word medial position, of treating English /l/ as a singleton and /r/ as a geminate liquid.
TABLE II. Predicted difficulty of /l–r/ perception, based on L1 phonological and phonetic considerations. 1 = least difficult, 2 = more difficult, 3 = most difficult.

<table>
<thead>
<tr>
<th>L1 background</th>
<th>Phonological environment</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial</td>
<td>Cluster</td>
</tr>
<tr>
<td>Korean</td>
<td>CV</td>
<td>CVC</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Japanese</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Table II summarizes cross-language predictions about the relative difficulty of the /l–r/ contrast, based upon phonological differences between Korean and Japanese, in the three environments of occurrence in Australian English.

It was predicted that Koreans would perform somewhat better than the Japanese in intervocalic and cluster environments, but not as well word initially. The predicted patterns of phonological difficulty are neutral with respect to perception and production and should apply equally to each domain. The reduced discriminability of the /l–r/ contrast in the consonant cluster (a language-independent phonetic effect) may be expected as an additional effect upon perception, which would apply equally to listeners of both language backgrounds.

II. IDENTIFICATION EXPERIMENT

A. Stimuli

Two Australian male speakers recorded 50 test words, consisting of (a) 30 /l–r/ words i.e., 15 minimal pairs, five in each of the three phonological environments, and (b) 20 distracter words (see Table III), which were also minimal pair contrasts with the target words. Recordings were made in a sound-attenuated recording booth. The distractor items served to provide a range of perceptual targets, consistent with an identification task, and to promote a linguistic level of stimulus processing. The same 30 target words were used in both the identification and discrimination tests. The 20 distracter words were used only in the identification test. Only common words, likely to be familiar to non-native speakers were used as test items. But it proved difficult to balance word familiarity in the choice of /l–r/ words for minimal pairs (see Sec. III D).

The stimuli were digitized at 20 kHz and 16 bit resolution for construction of identification and discrimination test tapes. The identification test contained 200 items (50 words × 2 speakers × 2 repetitions) in random order, completed in a single session, with a short rest break after every 50 items.

B. Subjects and procedure

Twenty subjects (ten Japanese and ten Koreans, equal numbers of males and females, all students at the University of Queensland) participated in both an identification and a discrimination test. Their age range was 18–37 years. Residency in Australia ranged from 2 to 18 months. Except for one Japanese subject, none had lived in any other English speech community, none had hearing disabilities. Eight of the ten Koreans and four of the Japanese students were undertaking English language courses, to qualify for entry to regular academic programs, thus indicating that the level of English competence may have been higher in the Japanese group.

The identification and discrimination tests were carried out under similar conditions in the same listening environment. Both tests were group-administered in the Language Laboratory of the University of Queensland, using a Tandberg IS9 master station controlling Tandberg 5600 listening stations. The identification test was carried out in two stages. In the initial training session, subjects were familiarized with the test items. The test words were presented on cards and the meaning and correct spelling of each word was reviewed. Subjects were also introduced to the method of filling in the answer sheet, and two practice trials were conducted using distracter items. In the main testing session, subjects were asked to write the word that they heard in the corresponding space on the answer sheet. Responses were scored as correct or incorrect against the target words containing /l/ or /r/. Irrelevant spelling errors ( rare) were ignored and responses to distractor items were not scored. Target words were presented twice on a given trial, with an interstimulus interval of 1 s. There was an intertrial interval of 8 s.

C. Data analysis and predictions

The experimental design called for a four way analysis of variance (ANOVA), where the dependent variable was the percentage of correct identifications for the target sound (/l/ or /r/) across the five words in each phonological environment, and the four treatment variables were: (i) The speakers (JI or TM); (ii) the language background of listeners (Japanese or Korean), (iii) the type of target sound ([l] or [r]); and (iv) the position in the word (initial, medial, or cluster). With repetition of the five /l/ or /r/ target words, the maximum score a subject could obtain was 10/10. This score was converted into a percentage correct for the results reported below.

Language-dependent phonological transfer effects will gain expression in main and interaction effects involving the
Language factor in the ANOVA. Specifically, we looked for (i) a significant main effect of language, in which the Korean group should identify /l/ and /r/ more accurately than the Japanese, and (ii) a two-way interaction of language by position, whereby the Koreans should perform best in medial position, next best in cluster position, and worst initially, and the Japanese, by contrast, should perform equally well in initial and medial position and but worse in cluster position. Effects which operate at the (language independent) phonetic level, i.e., at the level of extraction of phonetic features from the acoustic signal, will be reflected in main and interaction effects involving (i) speakers (JI or TM), (ii) phonetic type ([l] or [r]), and (iii) position (initial, medial, cluster), excluding interactions with language.

**D. Results**

A summary of the ANOVA is shown in Appendix A. The results are complex, with a high proportion of main and interaction effects achieving statistical significance at $p < 0.01$.

**1. Language effects**

Language-specific transfer effects arising from phonetic or phonological differences between Korean and Japanese were reflected in the main effect of language background and the interaction of language background and phonological environment (language: position). In both instances, the predicted pattern of responses was obtained. The Korean subjects performed better than the Japanese overall (83% vs 70% correct), and the pattern of performance across the three phonological environments was, except for a minor qualification, in agreement with that predicted in Table II above (see Fig. 1).

The minor qualification was that both groups performed less well than expected in intervocalic position. Post-hoc analysis revealed why. We made the mistake of including the items “lorry” and “lolly,” which had the effect of increasing the apparent difficulty of perceiving the /l–r/ contrast in intervocalic position. Errors occurred on these two items because listeners were uncertain of the locus of the contrast. With these items removed, the expected pattern of means was obtained.

As predicted, Korean speakers had more difficulty with word initial position than with the cluster or medial positions. The Japanese, who had greater difficulties with all three environments, relative to the performance of the Koreans, had the most trouble with /r–l/ perception in the consonant cluster, which is what would be expected on grounds of relative acoustic discriminability of the tokens, in the absence of countervailing L1 transfer effects.

**2. Type effects**

Significant main and interaction effects of segment type, /l/ or /r/ were obtained. The main effect of type ($F[1,198] = 14.86, p<0.0002$) showed a higher percentage of correct responses for /r/ rather than /l/ (81% correct versus 72%). However, all three of the two-way interactions involving segment type were significant as well as two of the three-way interactions.

The language by type interaction ($F[1,198]=8.29, p<0.005$) indicated that while the difference in favor of /r/ over /l/ perception was small (not significant) for the Koreans, and (84% to 82%), it was large for the Japanese (75% to 63%). This may reflect the greater phonetic similarity of the English /r/ rather than /l/ to the Japanese liquid, whereas the availability of two phonologically conditioned perceptual targets to map onto the English forms discouraged the development of a response preference for Korean listeners.

The type by position interaction ($F[2,198]=5.39, p<0.006$) showed that the perception of /l/ was favored in the cluster environment, but that /r/ perception was easier in initial and medial position. This finding agrees with Gillette (1980) and Sheldon and Strange (1982).

The three-way interaction of type by language by position also reached statistical significance ($F[2,198]=4.64, p<0.02$). It indicated that the pattern of difficulty across phonological environments differed for the Korean and Japanese listeners. Two sets of *a posteriori* t tests were conducted to identify the nature of this interaction. The results are summarized in Table IV, which shows the cell means and the comparisons that were tested. For the Koreans, /l/ in initial position scored significantly lower than liquids (ʃl/ or ʃl/) in any of the other environments (with the exception of /l/ in cluster position). Initial /l/ is the only one of the six phonetic type by position combinations which is excluded by Korean phonology, assuming, as argued above, that English liquid+stop clusters are reanalyzed as intervocalic liquids by Korean listeners, following application of vowel epanthesis. For the Japanese, *a posteriori* t tests (Table IV) showed a significant preference for /l/ over /r/ in initial and medial positions, which we noted previously in the two-way language by type interaction, combined with lower scores for both /r/ and /l/ in cluster position.

The other two- and three-way interactions involving segment type and speaker differences are described in Sec. 3 below.
3. Speaker effects

Significant speaker effects were observed. Overall, speaker JI provided clearer tokens than speaker TM (77.4% versus 73.5% correct). However, this main effect is small in relation to the two-way interaction of speaker by segment type, or the three-way interaction (speaker×type of sound×position) shown in Fig. 2.

Comparisons among means showed that speaker JI provided more clearly identifiable /l/ tokens than speaker TM in all three positions. With respect to /r/, the identification scores were comparable for both speakers, except for the cluster environment, where speaker JI’s tokens were clearly less identifiable than the other stimuli. To ascertain whether the speaker effects could be attributed to differences in the acoustic discriminability of the tokens provided by the two speakers, a spectrographic analysis of the test stimuli was undertaken. Observations focused on the extent to which well-known spectral and temporal cues for the /r–l/ contrast were present in the stimuli and correlated with listener identification scores.

4. Spectrographic analysis of speaker differences

Synthesis studies (Yamada and Tohkura, 1992a) usually manipulate three cues for the /l–r/ contrast: (1) the locus of the third formant, (2) the onset frequency of the second formant, and (3) the abruptness of rise of F1 into the following vowel, which is greater in the case of /l/, imposing a clearer acoustic state change between vowel and consonant in the case of the lateral. In the acoustic analysis, we examined time-varying frequency and amplitude characteristics of the formant transitions of /l/ and /r/ in those environments where speaker effects (differences between JI and TM’s productions) had the greatest impact upon listeners’ identification scores.

5. Analysis of /r/ in cluster position

As noted earlier, speaker JI’s /r/ tokens were more accurately perceived in initial and medial positions than those of speaker TM, except for cluster position where TM’s tokens were responded to more accurately. Comparison of spectrograms of JI and TM’s /r/ cluster tokens indicated a
consistent difference in the rate of change (slope) of the third formant trajectory between the tokens of the two speakers. As can be seen from Fig. 3, there was a discernible “elbow” in the third formant trajectory for speaker TM, whereas speaker JI produced a more gradually rising $F_3$ transition to the steady-state target for the following vowel. To quantitatively assess these differences in formant trajectories, particularly the slope of $F_3$, the following measurements were taken.

Starting, midpoint, and end-point measurements were made of the first, second, and third formant frequencies of the /r/ segment. Measurements were made by hand from a spectrogram and power spectrum display (Sensimetrics, SpeechStation™). Starting points for the formant transitions were set so that the first glottal pulse fell within the left skirt of the FFT window. The end point of the formant transition was judged from $F_3$, at the (second) “elbow” of the $F_3$ trajectory into the following vowel. The midpoint of the formant transition was judged, by eye, to fall midway between the start and end points of the formant transitions.

From the formant frequency and duration measurements, the slopes of the $F_1$, $F_2$, and $F_3$ transitions were calculated, from onset to midpoint of the /r/, and from midpoint to end point. The greater of these two slope estimates was defined as the maximum slope of the formant transition for a given formant ($slF_{max}$). These and other formant frequency change measures, taken in the course of calculating the slopes of the formant trajectories, were then correlated with identification scores for JI and TM’s /r/ tokens in the cluster environment. The two variables with the highest correlations with the dependent variable were the maximum slope of the third formant transition ($slF_{3max}$) and the overall duration of the formant transition (i.e., the /r/ segment duration: $rdur$), which were $r = 0.76$ and $r = -0.59$, respectively. These two variables were then used as predictors in a linear regression analysis, with the percentage correct responses to /r/ in cluster position for each of the speakers’ tokens as the dependent variable. In view of the small number of observations (ten tokens), not more than two independent variables were justified. The regression equation yielded a multiple $R$-squared of: $R^2 = 0.64$ ($F[2,7] = 6.34, p < 0.027$).

6. Analysis of /l/

As noted previously, in the two-way interaction of type by speaker effects, speaker JI’s /l/ tokens yielded higher identification scores in all three positions than those of TM. Figure 4 shows sample spectrograms of the /l/ tokens by the two speakers. JI’s tokens showed clearer formant structure in the region of $F_3$ and $F_4$ and a more abrupt acoustic state change between the consonant and the following vowel nucleus. (These spectral differences were apparent over various dynamic range settings of the spectrogram display.) We sought to derive quantitative indices of the clarity of the higher formant structure from amplitude measurements based upon the short-term power spectrum, averaged over 3 to 4 pitch periods preceding and following the estimated center point of the /l/ segment. Energy maxima at $F_1$, $F_2$, $F_3$, and $F_4$ were measured in dB and referenced to the major energy peak in the spectrum. Energy minima ($M_1$, $M_2$, $M_3$)
at points midway between formant peaks were also measured. From these spectral maxima and minima, three spectral contrast measures were calculated; one each for \( F_2 \), \( F_3 \), and \( F_4 \), as follows:

\[
\text{contrast}: F_n = \text{amplitude}: F_n - \text{amplitude}: M_{n-1} \quad \text{(in dB)},
\]

where \( 2 \leq n \leq 4 \).

The spectral contrast measure for a given formant was thus the difference in decibels between the formant energy maximum and the trough value measured halfway to the preceding formant peak, thus ensuring that the three contrast measures were independent of each other. Two additional acoustic measures, the duration of the /l/ segment, and the stability of the formant pattern (assessed in terms of the change in \( F_2 \) between the onset and offset of the /l/ segment) were also made. The three formant contrast measures, the energy maxima and minima measurements from which the contrast measures were derived, the /l/ segment duration, and the \( F_2 \) change measure were separately correlated with the identification scores for JI and TM’s /l/ tokens in the three positions.

Correlation coefficients for the formant contrast measures on \( F_2 \), \( F_3 \), and \( F_4 \) were \( r = 0.39 \), \( r = 0.46 \), and \( r = 0.65 \), respectively, supporting the spectrographic observation that the clarity of the higher formant structure contributed to the higher identification scores for JI’s /l/ tokens compared with those of TM. Several linear regression analyses were conducted, taking various combinations of the three formant contrast measures and other acoustic variables mentioned above. These analyses consistently indicated that the only measure which made a statistically significant independent contribution to prediction of the /l/ identification scores was \( cnF_4 \) (spectral contrast of \( F_4 \)).

In summary, the regression analyses of speaker-induced differences in /l/ identification in cluster position and /l/ identification in all three environments found an acoustic basis for the listeners’ performance, in terms of the clarity with which acoustic cues for the /l–r/ contrast were encoded by the two speakers JI and TM. For /l/ clusters, the maximum slope of \( F_3 \) and the duration of the formant transitions jointly accounted for approximately 64% of the variance in the dependent variable (the identification scores). The negative loading of the duration factor is notable. The /l/-sound is typically realized as a flap in Korean and Japanese. Speaker TM’s shorter /l/ segments and more rapid formant transitions, appear to have favored /l/ perception in the cluster environment.

Speaker-induced differences in the accuracy of /l/ identification were associated with the degree of spectral contrast in the region of \( F_3 \) and \( F_4 \), with the latter dominating the regression model. Prediction of speaker differences for /l/ was less successful, possibly because no readily quantifiable means was found to capture the qualitatively observed feature of the “abruptness of change” between the /l/ segment and the following vowel, which, along with the formant contrast feature, distinguished speaker JI’s tokens from those of TM.

7. Phonological markedness and universal grammar

Results presented thus far provide no evidence of language universal markedness constraints or effects of phonological parameter settings affecting listeners’ perceptions. Indeed, the language-dependent transfer effects reported thus far suggest the contrary, that phonological effects on perception reflect language particular exigencies. Phonological parameters can potentially affect second language learners’ responses either through parametric settings required by the second language, or through settings established in first language acquisition. As mentioned earlier, Japanese and Korean are very similar with respect to constraints on initial consonant clusters, which is the domain where parametric settings may express themselves in the data of this experiment. Hence expressions of markedness constraints should be identical for Japanese and Korean listeners and should reflect the markedness ranking of the different English consonant clusters incorporated in the test stimuli.

Two approaches have been taken to assigning markedness values to consonant clusters. The first, and historically prior approach, which has been dubbed the “typological” (Eckman and Iverson, 1993), establishes the relative markedness of a segment or segment cluster on grounds of distribution frequency of occurrence in a given environment, and cross-linguistic implicational relationships. The second approach, which yields similar results, and some would argue, provides an explanatory basis for the distributional findings, depends upon sonority relations within the syllable and has been most extensively discussed by Clements (1990). A “minimal sonority distance parameter” (MSD; Broselow and Finer, 1991) assigns a markedness value to the range of consonant clusters allowed by the phonotactics of a language. Assuming that segments can be rank ordered on a scale of sonority and that onset and coda sequences respect the sonorance hierarchy, MSD provides a means of calculating the markedness value of any given cluster, establishes an implicational hierarchy among clusters, and provides a quantitative characterization of the relative markedness of any two languages with respect to their phonotactic constraints.


Both methods of assigning markedness values to consonant clusters were used to assess the correlation of markedness with rates of identification of the /r–l/ clusters employed in this experiment (see Table V).

Neither the MSD parameter of Broselow and Finer (1991) nor the TM constraint of Eckman and Iverson (1993) accounted for the pattern of identification scores across words or language groups. All four correlations (Spearman’s rank-order coefficient) between the two indices of markedness and the cluster identification scores for the Koreans and the Japanese listeners were computed. The correlation coefficients ranged from 0.03 to 0.28. None approached statistical significance on a one-tail test. Nor can the clear differences in favor of the Korean group be attributed to
parametric settings, because the phonotactic constraints of Japanese and Korean initial clusters are practically identical.

E. Summary of identification experiment results

Results of the identification experiment suggested a dominance hierarchy among the factors contributing to the perceptibility of the /r–l/ contrast by Japanese and Korean learners of English. For the Korean listeners, the possession of a native-language model for the /r–l/ contrast in intervocalic position appeared to be the dominant factor in determining the pattern of identification scores across the three phonetic environments. The contrast was best perceived in medial position. In cluster position, the /r–l/ contrast was marginally better perceived than in initial position, in spite of the reduced acoustic discriminability of /l/ and /r/ in consonant clusters. Enhanced performance in cluster position appears to have been supported by a compensatory phonological strategy of (re)syllabification through epenthesis (CCV→CvCV, where v=epenthetic vowel), in order to access native language templates for the /r–l/ contrast. It seems appropriate to refer to this enhancement of /r–l/ perception in cluster position by Korean listeners as a phonological transfer effect, because any enhancement of the contrast occurs as a result of a linguistic transformation of the input generated by the listener, presumably to render the input signal more compatible with L1 phonotactics, thereby providing a stimulus which is interpretable as an L1 phonological target. Vowel epenthesis may also be observed in Japanese learners’ productions of these consonant clusters, but no processing advantage is obtained, because Japanese lacks an L1 model for the /r–l/ contrast in all environments.

The pattern of performance for Japanese listeners, for whom no L1 model of the L2 contrast was available, was consistent with considerations of acoustic discriminability, with performance best in initial position and worst in cluster position.

The extent to which speakers provided clear exemplars for the /r–l/ contrast in various environments also affected identification rates. Speaker differences interacted with the stimulus type and the environment in which the target sound occurred, but they did not interact with the language factor. Acoustic analysis of the stimulus tokens showed that identification rates were higher for those stimulus tokens which better exemplified spectrographic cues known to be effective for /l–r/ discrimination.

Phonological markedness was found not to be predictive of perceptual error rates among consonant+liquid clusters. Universal markedness constraints apparently had little bearing on the results of this experiment and their relevance for speech perception in second language learning is therefore questionable. On the other hand, language particular factors, arising from a comparative analysis of the relevant phonological structures in Korean and Japanese, were crucial for predicting the pattern of perceptual errors.

III. DISCRIMINATION EXPERIMENT

It was expected that the discrimination task would engage specific linguistic processing of the stimuli less than the identification task reported above. To assess the impact of task variables upon precedence relations and interactions among the factors investigated in the identification experiment, the same group of subjects participated in a discrimination test, using the identical target stimuli. The discrimination experiment was conducted two weeks after the identification task.

A. Procedure

An oddball discrimination test was used, in which, on a given trial, one member of a sequentially presented triad of stimuli differed from the other two, and the listener’s task was to identify the odd one out. The advantage of this procedure over the more commonly used ABX or Same–Different discrimination procedures is that it is relatively economical in terms of the number of trials required to map the set of discrimination judgements. Its main disadvantage is that it imposes a greater sensory storage load upon listener’s and consequently may be more prone to influences of phonological encoding. Hence the oddball discrimination task may yield conservative estimates of task differences between identification and discrimination on /l–r/ performance.

On a given trial, listeners were presented with one of six possible triads made up of minimal pair /r–l/ triplets (LLR, LRL, RLL, RRL, RLR, LRR) and asked to indicate which word was different. Repetitions of the same token were used for the two stimuli which were “the same” within a given triplet. There were 360 items in the stimulus set, which included one replication (5 pairs×3 positions×2 speakers ×6 triplets×2 repetitions). As 360 items were too many for a single experimental session, the subjects were divided into two groups, each of whom responded to half of the trials in the total stimulus set. The interstimulus interval between items in a stimulus triad was approximately 0.5 s and the intertrial interval between triads was 3 s, with a short rest period after each block of 60 trials.

Test conditions were the same as in the identification experiment. Subjects were group tested in the language laboratory listening booths. As in the identification test, there was a pretest in which subjects were familiarized with the test words and then with the response task, using triplets constructed from selected distracter items employed in the identification task.
B. Data analysis

Initial analysis of the data revealed no significant differences between the two groups, who each heard half of the test stimuli, so in further analysis both groups were combined. A four-way analysis of variance was undertaken, with the discrimination scores as the dependent variable, and the same treatment variables as in the identification experiment, except that, in this case, there was no factor for stimulus type (/l/ or /r/), but there was an additional one for the carrier pattern of the discrimination contrast within the triplet (LLR, LRL, LRR, RLL, RLR, RRL).

C. Results

A four-way analysis of variance was conducted. All four main effects were significant at or beyond the $p<0.001$ level (see Fig. 5), but no significant interactions were obtained. The highly significant main effects of language background ($F[1,648], p<0.0001$) and speaker ($F[1,648], p<0.0001$) were in the same direction as for the identification experiment. The Korean listeners performed /l–r/ discrimination better than the Japanese (Mean = 86.7% correct versus 70.2%, respectively). Speaker JI’s tokens were more discriminable than those of TM (82.4% vs 74.5%).

A posteriori comparisons among discrimination score means for initial, cluster, and medial positions indicated that medial position was significantly different from the other two ($t$ test medial versus initial position: $t=-3.03$, df = 478, $p=0.002$, two tailed; $t$ test medial versus cluster position: $t=-2.32$, df = 478, $p=0.021$; $t$ test initial versus cluster position: $t=-0.61$, df = 478, $p=0.54$). This was similar to the pattern of main effects for the identification test. However, the discrimination test did not show any significant interaction effects between the factors of position, language, and speaker.

There was also a strong main effect of the triplet carrier pattern used to elicit discrimination judgments ($F[5,648], p<0.0002$). As Mochizuki (1981) found, the LRL and RLR triplets resulted in lower discrimination scores ($t$ test LRL + RLR versus four other patterns: $t=-4.05$, df = 718, $p<0.001$, two tailed; $t$ test LRL versus RLR: $t=-0.71$, df = 238, $p=0.48$; $t$ test LLL versus RRR: $t=-0.97$, df = 238, $p=0.33$). The LRL and RLR are the two patterns where the “same” stimuli are not contiguous and listeners must hold all three stimuli in sensory storage in order to make a discrimination judgment.

D. Comparison of identification and discrimination tasks

While a similar pattern of main effects was found in the discrimination experiment to those observed in the identification test, the lack of interaction effects suggested that the two tasks may have invoked different processing strategies.
on the part of listeners. Phonological transfer effects (language-specific learning effects) were diminished in the discrimination experiment. The main effect of language background was preserved, but no interaction between listeners’ language and the position (initial, cluster, medial) of the /l–r/ contrast was found. Similarly, although a strong main effect of speaker was observed, which was in the same direction as the identification experiment, there were no significant interaction effects between the Speaker factor and listeners’ language background, or the position of the /l–r/ contrast. These results were consistent with the hypothesis that underlying perceptual mechanisms differed in the identification and discrimination tasks. It was apparent from the identification test results that many of the significant higher-order interactions may have been caused by response preferences towards /r/, which varied with the language background of the listener and the position of the target sound. However, such effects were not apparent in the case of the discrimination data.

If, as hypothesised earlier, identification invokes greater linguistic processing of the signal than does discrimination, then word familiarity effects should operate more strongly in the identification than the discrimination test. An attempt had been made to balance word familiarity effects in the initial selection of /l/ and /r/ minimal pairs, by choosing only items which would be familiar to second language learners. However, in view of the /l/ preference biases observed in the identification task, a comparison of the relative subjective familiarity (Flege et al., 1995) of the /l/ and /r/ members of the 15 minimal pairs was warranted, and the impact of any inequalities of lexical familiarity between /l/ and /r/ items on identification and discrimination performance should be assessed.

On the other hand, if identification and discrimination involve similar perceptual processes, then the relative order of difficulty of /l–r/ contrasts across the 15 minimal pairs should be similar across the two tasks. To compare response profiles across the minimal pairs, both the identification and the discrimination data were rescored, so that separate /l/ and /r/ scores were obtained for each of 15 minimal pairs.

1. Scoring the dependent and independent variables

In the case of the discrimination test, a preference towards /l/, or a lexical familiarity effect, could only be expressed as a bias towards or away from the oddball member of a triad. Hence separate /l/ and /r/ scores were obtained by taking account of whether the oddball in a given triad was an /l/ or an /r/ word.

To assess the relative subjective familiarity of the /l/ and /r/ words that comprised the minimal pairs, the authors and two experienced language teachers independently judged which member of each pair would be likely to be more familiar to Japanese and Korean learners of English. There was a high degree of consensus among the four judges. In all but three of the minimal pairs, the /l/ member was rated as more familiar than the /r/ member; the exceptions being: low>row, play>pray, climb>crime. Hence any r bias or lexical familiarity effects would be highly correlated and probably indistinguishable in the present experiment.

FIG. 6. Interaction plots: Identification and discrimination data. Comparison between the interaction of item type (/l/ or /r/) with minimal pair for identification and discrimination responses. The graphs illustrate different response profiles for the identification and discrimination tests and the presence of an interaction effect for the former, but not the latter.

2. Statistical analysis

Parallel two-way analyses of variance of the identification and discrimination tests were conducted, with the 15 minimal pairs and the item type (/l/ or /r/) as experimental factors. Figure 6, which represents an interaction plot of the two experimental factors, shows the profile of /l/ and /r/ scores across the minimal pairs for the identification and discrimination tests.

A significant main effect of minimal pair was found for both identification and discrimination test (F[14,570] = 1.99, p = 0.016; F[14,570] = 2.88, p = 0.003, respectively), but the scoring profiles of /l/ and /r/ over the 15 minimal pairs were quite different for the two tests (see Fig. 6). The identification test yielded a significant main effect for item type, (F[1,570] = 9.17, p = 0.002) revealing an r bias for most of the minimal pair contrasts. No such effect was found in the case of the discrimination test (F[1,570] = 0.72, n.s.). The identification test also yielded a significant minimal pair by item type interaction, (F[14,570] = 1.77, p = 0.039) but no significant interaction effect was found for the discrimination test (F[14,570] = 0.12, n.s.).

The minimal pair lorry–lolly, which was the most difficult on the identification test was the second easiest pair on the discrimination test. The minimal pair red–led showed a strong r bias on the identification test, but a slight preference
in the opposite direction on the discrimination test. These differences in performance profile across the minimal pairs indicate that the identification and discrimination tests were tapping different perceptual processes or making different task demands of listeners.

The effects of lexical familiarity were also assessed. However, this factor proved to be statistically indistinguishable in its effects from that of item type (L1 or L2), which indicated an r bias on the identification test, but no significant effect on the discrimination task. Thus while it was not possible to separate the effects of r bias and lexical familiarity in this experiment, the fact that either or both were operative in the identification task, but not in the discrimination task, supports the hypothesized differential sensitivity of the two tasks to the influence of language-dependent processing strategies and linguistic factors in general.

**IV. GENERAL DISCUSSION AND CONCLUSIONS**

Recent models of cross-language speech sound perception (Flege, 1995; Best, 1995; Kuhl, 1995) stress the importance of naive-language phonemic categorization of non-native ("foreign") sounds in explaining patterns of perceptual errors or the likelihood or ease of accommodation of the perceptual system to such sounds. Although there is disagreement over the phonetic basis for the categorization of non-native sounds, (e.g., whether it be gestural [Best] or acoustic [Kuhl]), all share the assumption that, at least in the initial stages of L2 exposure, patterns of perceptual assimilation to L1 phonemic categories strongly determine learners’ responses to novel sounds. Best (1995) has elaborated different classes of L1 assimilation for the purpose of predicting pairwise discrimination performance, depending on how sounds in the L2 pair are categorized with respect to L1 targets. Kuhl has postulated "perceptual magnet effects," whereby phonetic tokens that fall close to a phonemic prototype are less sensitively discriminated than sounds which fall at a distance from such prototypes. Implicit in all three models is the mediating influence of phonological experience and the language-specific nature of the effects on second language speech sound perception.

The major goal of this study was to observe, within a single experiment, the interaction of potentially competing factors, operating at different levels of perceptual processing (the phonetic and the phonological), over different domains (the language specific and the language universal) in the perception of a novel phonemic contrast. The phonetic level of processing was reflected in factors affecting the acoustic discriminability of the tokens, namely, in terms of the present experiment: (a) the position of the target contrast in the word, and (b) the clarity with which the speaker encoded the target contrast in particular tokens across the three contexts. The phonological level of processing was reflected in L1 learning effects, namely: (a) the extent to which Korean provides an L1 model for the foreign /r–l/ contrast in medial position, whereas Japanese does not, and (b) differences in the extent to which the two listening tasks, phoneme identification and oddball triad discrimination, required specific linguistic processing of the test stimuli. Both prior phonological learning and the relative acoustic discriminability of the items affected subjects’ performance on the identification test. Where both factors were engaged, phonological learning effects predominated over the effects of acoustic discriminability. This was evident from the differential impact of positional effects on /r–l/ identification for the Japanese and Korean listeners. Whereas the Japanese listeners’ poorer identification of /r–l/ in cluster position was attributable solely to the lower acoustic discriminability of /r–l/ in this environment (supporting findings of previous studies), the Korean listeners’ responses were predominantly influenced by prior L1 phonological learning. Specifically, their performance on the /r–l/ contrast in cluster position was enhanced by a phonological strategy (vowel ephthnesis), which enabled them to deploy the Korean contrast between flaps and geminate liquids in medial position and to aid perception of the English /r–l/ contrast in initial clusters, such as play—pray. Because this L1-driven enhancement of perceptual discrimination depends upon a misyllabification of the input (from the perspective of L2), it seems appropriate to view it as a phonological transfer effect, generated by learned phonotactic expectancies for L1 rather than by any inherent phonetic property of the stimulus.

The relative acoustic discriminability of the stimulus items also impacted upon identification scores in terms of the clarity with which a given speakers’ tokens encoded critical acoustic features for /r–l/ discrimination. Identification rates for tokens from the two speakers (JI and TM) were found to vary depending upon the type of target (/l/ or /r/) and its position in the carrier word. Speaker JI’s /l/ tokens were more identifiable in all three positions than those of TM. Identification scores for /l/ tokens were found to be quite strongly associated with a measure of the spectral contrastiveness of higher formants F3 and F4. Analogously, speaker TM’s /r/ tokens in cluster position were more accurately perceived than those of JI. Spectral analysis, combined with statistical regression showed that the steepness of F3 in combination with the segment duration were good predictors of /r/ identification in cluster position.

The pattern of main effects for scores on the discrimination test were similar to those found on the identification test. The Koreans performed better than the Japanese. Speaker JI’s tokens were better discriminated than those of TM. Items in medial position were better discriminated than those in initial and cluster positions. However, there was an absence of factorial interactions of the kind observed in the identification test. Rescoring and comparison of performance profiles across the set of minimal pairs indicated that quite different factors were at work in the identification and discrimination tests. The minimal pair scoring profiles for /r/ and /l/ in the identification and discrimination tests were essentially uncorrelated. A lexical familiarity effect, which was statistically inseparable from a response bias in favor of /r/ items, was found to be significant for identification test performance, but no similar effect was found for performance on the oddball discrimination task. These findings clearly indicate that the identification and discrimination differed in terms of task demands on listeners. The lack of higher-order language, position, and speaker interactions and the absence of a lexical familiarity effect on the discrimination task sup-
ported the hypothesis that a phonological level of signal processing was less engaged by the oddball discrimination task.

However, unlike the identification task, where performance on particular items was found to be linked with the presence of acoustic features known to distinguish /r/ and /l/, no comparable analysis of the acoustic features differentiating hard from easy to discriminate items was undertaken. Except for those triads which imposed some short-term memory load (patterns LRL and RLR), performance on the discrimination task was usually close to the ceiling. This is a topic for further investigation. The main significance of the present findings lies in the predominant role of prior phonological learning in the perception of a non-native phonemic contrast, which emerged from comparison of Japanese and Korean identifications of English /l/ and /r/, in the face of speaker- and context-dependent phonetic variation. Perception was enhanced in an environment of acoustic disadvantage by the application of an L1 phonological processing strategy that yielded inappropriate syllabification of the input from the perspective of L2, but which enabled listeners to encode the foreign contrast more effectively. A similar interaction between L1 phonological processing strategies and phonetic properties of speech stimuli is reported for the case of foreign vowel perception by Japanese and Korean learners of English in Ingram and Park (1997).

APPENDIX A

| TABLE AI. Identification of /l/ and /r/ analysis of variance: repeated measures design. a |
|---------------------------------|--------|---------------|---------|------------------|------------------|
|                                | df     | Sum of square | Mean square | F value           | Pr(F)            |
| Language (Jap., Kor.)          | 1      | 102.3773      | 102.3773    | 3.665755         | 0.071579         |
| Residuals                      | 18     | 502.7047      | 27.9280     |                   |                  |
| Error: Within                  | df     | Sum of square | Mean square | F value           | Pr(F)            |
| Main effects:                  |        |               |             |                   |                  |
| Speaker (JL, TM)              | 1      | 10.3128       | 10.3127     | 3.48532           | 0.06339394       |
| Type (/r/, /l/)                | 1      | 43.9898       | 43.9894     | 14.86691          | 0.0001559        |
| Position (Ini., Clu., Med.)    | 2      | 7.6443        | 3.82215     | 1.29174           | 0.2779982        |
| Two-way interactions:         |        |               |             |                   |                  |
| Speaker:type                   | 1      | 47.4815       | 47.4815     | 16.04696          | 0.0000874        |
| Speaker:position               | 2      | 6.0618        | 3.04045     | 1.02767           | 0.3597280        |
| Type:position                  | 2      | 31.9359       | 15.9675     | 5.39657           | 0.0052233        |
| Speaker:language               | 1      | 1.1690        | 1.1690      | 0.39505           | 0.5303648        |
| Type:language                  | 1      | 24.5440       | 24.5440     | 8.29495           | 0.0044132        |
| Position:language              | 2      | 24.9234       | 12.4617     | 4.21140           | 0.0161721        |
| Three-way interactions:        |        |               |             |                   |                  |
| Speaker:type:position          | 2      | 40.7568       | 20.37839    | 6.88713           | 0.0012835        |
| Speaker:type:language          | 1      | 2.1565        | 2.15651     | 0.72882           | 0.3942978        |
| Speaker:position:language      | 2      | 1.4693        | 0.73464     | 0.24828           | 0.7803847        |
| Type:position:language         | 2      | 27.4693       | 13.7346     | 4.64179           | 0.0107134        |
| Four-way interaction:          |        |               |             |                   |                  |
| Speaker:type:position:lang     | 2      | 14.0818       | 7.04089     | 2.37955           | 0.0952351        |
| Residuals                      | 198    | 585.8641      | 2.95891     |                   |                  |

a Responses to the items “lorry” and “lolly” were excluded from the scoring and the ANOVA. See text for explanation.

APPENDIX B. SPECTROGRAMS FOR MINIMAL PAIR TOKEN CONTRASTS FOR ITEMS IN FIGS. 3 AND 4

![Spectrograms](image-url)

FIG. B1.


