Categorial Discrimination of English and Japanese Vowels and Consonants by Native Japanese and English Subjects

J. E. Flege*, S. G. Guion, R. Akahane-Yamada† and J. Downs-Pruitt†

*Department of Rehabilitation Sciences, VH 503, University of Alabama at Birmingham, Birmingham, AL 35294 and †ATR Human Information Processing Labs, 2-2 Hikaridai, Seika-cho Soraku-gun, Kyoto 619-02 Japan

Abstract: We examined the discrimination of four pairs of English vowels, and four pairs of English consonants, by three groups of native Japanese (NJ) subjects. The rate at which the English stimuli were produced had a similar effect for the NJ subjects and NE controls. The relative difficulty of the English phonetic contrasts for the NJ subjects appeared to depend on how the English stimuli were perceptually mapped onto Japanese vowels and consonants. Three years of exposure to English in the US led to a significant increase in A’ scores for the consonant but not the vowel contrasts.

A number of researchers have hypothesized that nonnative subjects’ accuracy in producing second-language (L2) vowels and consonants is constrained by perceptual accuracy. Many previous studies have shown that as adult learners gain experience in English, their production of English vowels and consonants improves measurably. Fewer studies have examined the effect of L2 experience on speech perception. One aim of this cross-sectional study, therefore, was to determine if the discrimination of English vowels and consonants by native Japanese (NJ) adults improves as a function of how much exposure they have received to English as it is spoken by native speakers of English. Nonnative subjects might be expected to have greater difficulty perceiving English vowels and consonants which have been produced at a relatively fast than normal speaking rate. A second aim of the study was to determine if an increase in speaking rate would lower the discrimination scores of the NJ subjects to a greater extent than the scores obtained from the subjects in a native English (NE) control group.

Four groups, each with 5 male and 5 female subjects, were tested. The subjects in NE were English monolinguals (mean age = 25 years). Those in NJ-I were Japanese professionals (mean age = 30 years) who had lived in the US for 3.2 years on average (range: 1.8 to 5.5 years) when they were tested in Birmingham, Alabama. The subjects in group NJ-2 (mean age = 29 years), were matched to the NJ-I subjects for educational level. They used English in their work but had never lived outside of Japan. The subjects in NJ-3 were college students who did not use English professionally and had never lived outside of Japan.

A categorial discrimination test (CDT) was used to assess the subjects’ perception. Four English consonant contrasts (/l/-/l/, /s/-/s/, /b/-/b/, /l/-/w/) were tested in the first session, and four contrasts between English vowels (/i/-/u/, /i/-/l/, /a/-/a/, /e/-/e/) were tested in a second session. The vowel stimuli were spoken in a /b/bo/ context, and the consonants were spoken in an /l/ cl/ context. Five native speakers of American English produced the stimuli at both normal and fast rates. The three edited CV stimuli presented on each trial were always spoken by different talkers. There was an odd item out in the eight “different” trials (e.g., /l/, /s/, /l/) but not in the eight “catch” trials (e.g., three tokens of /l/, or three tokens of /s/) that tested each of the phonetic contrasts. The speaking rate at which the vowel and consonant stimuli were produced (normal vs. fast) was counterbalanced within the two sessions.

The listeners’ task was to choose the odd item out if they thought that one of the three stimuli in a trial was different from the other two. They were to select a fourth response alternative, “none”, if they heard three instances of the same vowel or consonant. The proportion of times that the odd item was correctly chosen in different trials (maximum = 8) and that an “odd” item was incorrectly chosen in catch trials (maximum = 8) was determined. The proportions of hits and false alarms were then used to calculate A’ scores. These scores provided an unbiased estimate of the listeners’ sensitivity to each of the eight phonetic contrasts. The A’ scores for the vowels and consonants were then submitted to separate (4) group by (4) phonetic contrast by (2) stimulus speaking rate ANOVAs.

The mean A’ scores shown in Fig. 1(a) for consonants have been averaged over the normal-rate and fast-rate stimuli because the speaking rate factor was non-significant (normal = 0.72, fast = 0.70; p > 0.10) and did not interact significantly with any other factor. The effect of group was significant $[F(3,36) = 28.7, p < 0.01]$. A Tukey’s post-hoc test revealed that the NE subjects had significantly higher A’ scores than NJ-1, NJ-2 and NJ-3, and that NJ-1 had higher scores than NJ-3 (p < 0.01). A group by contrast interaction was obtained $[F(9,108) = 3.9, p < 0.01]$ because the between-group differences varied across the phonetic contrasts: NE > NJ-1, NJ-2 and NJ-3 for /l/-/l/ and /s/-/s/, NE > NJ-2 and NJ-3 for /b/-/l/, and NE > NJ-3 for /l/-/w/ (p < 0.01 by Tukey’s post-hoc tests).

The A’ scores that are shown in Fig. 1(b) for vowels have also been averaged over the two speaking rates. Significantly higher scores were obtained for vowels spoken at a normal than fast rate (0.78 vs. 0.72; p < 0.01), but rate did not interact with any other factor (p > 0.10). The main effect of group was significant (p < 0.01). A
Tukey's post-hoc test revealed that the NE subjects received significantly higher A' scores than the subjects in NJ-1, NJ-2 and NJ-3 (p < 0.01). A group by contrast interaction was obtained \([F(3,36) = 13.6, p < 0.01]\) because the pattern of between-group differences varied over the vowel contrasts: NE > NJ-1, NJ-2 and NJ-3 for /a/-/ʌ/, NE > NJ-2 and NJ-3 for /ɛI/-/ɛt/ (p < 0.01), and no significant between-group differences for /u/-/ʊ/ or /ɪ/-/ɜ/ (p > 0.10). The lack of a simple main effect of group for /u/-/ʊ/ was due to the fairly low scores obtained for the NE subjects; for /ɪ/-/ɜ/, it was due to the fairly high scores obtained for all three NJ groups.

A follow-up experiment was carried out to determine why certain English phonetic contrasts were more difficult for the NJ subjects than others. Japanese monolinguals identified each English vowel and consonant stimulus in terms of a Japanese vowel (or consonant) category, and then rated the stimulus for goodness as an instance of that category. Our preliminary results suggest that NJ subjects' sensitivity to English phonetic contrasts may depend on the extent to which they identify two English vowels (or consonants) as good instances of two different Japanese vowel (or consonant) categories. For example, English /ɪ/-/ɜ/ and /b/-/v/ were readily discriminated by most of the NJ subjects because the two members of both pairs were usually identified as good instances of two different Japanese vowel (or consonant) categories. The English [i] vs. [r] stimuli were usually heard as instances of Japanese /i/ and /ɛ/, respectively. The [b] and [v] stimuli were usually heard as instances of Japanese /b/ and /v/, respectively.

On the other hand, English phonetic contrasts were discriminated poorly when both members of a pair were heard as instances of a single Japanese category, or else as poor instances of multiple Japanese categories. For example, the English [ɪ] and [r] stimuli were heard as poor instances of the Japanese /r/ (a flap) or as /w/. The English [u] tokens were heard as poor instances of Japanese /u/, /u/, /e/ and /ɛt/; and the English [u] tokens were heard as poor instances of Japanese /e/, /a/, /u/, /u/ and /a/.

![Graph](image.png)

**FIGURE 1.** The mean A’ scores obtained in a categorial discrimination test from adult native speakers of English (●), Japanese professionals who had lived in the US for 3 years (■), Japanese professionals who had never lived abroad (○), and Japanese college students (□).

In conclusion, the rate at which the English stimuli were produced had a similar effect for the NJ and NE subjects. The discriminability of English vowels and consonants for the NJ subjects seemed to depend on how well the stimuli matched sounds found in the Japanese phonetic inventory. The NJ subjects who had the most previous exposure to native-produced English, NJ-1, discriminated English consonants more accurately than did the least experienced subjects, NJ-3. Groups NJ-1 and NJ-3 did not differ significantly for vowels, however. These findings suggest that the perceived relation between Japanese and English consonants may change to a greater extent as the result of unaided L2 experience than does the perceived relation between Japanese and English vowels.

**ACKNOWLEDGMENTS**

This study was supported by grant DC00257 from the National Institute for Deafness and Other Communicative Disorders. The authors thank Thorsten Piske for his comments on an earlier draft.