EXAMINATION OF SIMILARITY BETWEEN ENGLISH /r/, /l/, AND JAPANESE /ɾ/: AN INVESTIGATION OF BEST EXEMPLARS BY ENGLISH AND JAPANESE SPEAKERS

Kota Hattori and Paul Iverson

Department of Phonetics and Linguistics, University College London, London, UK
k.hattori@ucl.ac.uk, p.iverson@ucl.ac.uk

ABSTRACT

Japanese adults have difficulty learning the English /r/-/l/ contrast, and it has been suggested that this occurs because /r/ and /l/ are similar to the Japanese /ɾ/ category. The present experiment evaluated this similarity by finding best exemplars of these three consonants in a 5-dimensional acoustic space (F1, F2, F3, closure duration, transition duration) for native speakers of Japanese and English. The results demonstrated that /ɾ/ was similar to /l/, but not /r/, for Japanese listeners. However, the /ɾ/ and /l/ best exemplars of Japanese speakers were still significantly different (e.g., /ɾ/ having a shorter closure than /l/), indicating that Japanese speakers maintained separate mental representations for these categories rather than using their L1 /ɾ/ for both consonants.

Keywords: L2 speech perception, best exemplar, category assimilation

1. INTRODUCTION

Adult second-language (L2) learners can encounter great difficulties learning some non-native vowels and consonants. A classic example of hard L2 phoneme learning is the acquisition of the English /r/-/l/ contrast by Japanese speakers; many early learners of English are near chance at categorizing these phonemes and have production difficulties, although they can improve with experience and training [4, 7, 11, 13].

Why is learning the /r/-/l/ contrast so difficult for native speakers of Japanese? Best's [2] Perceptual Assimilation Model (PAM) argues that novice L2 learners assimilate non-native phonemes into existing L1 categories. For example, Japanese listeners are claimed to perceive both English /r/ and /l/ as poor exemplars of their L1 flap category (i.e., /ɾ/), making these phonemes sound the same [3]. Similarly, Flege's [6] Speech Learning Model (SLM) argues that the distance between L1 and L2 phonemes affects L2 phoneme learning. For example, Aoyama et al. [1] have claimed that the Japanese /ɾ/ is closer to the English /l/ than to /r/; learning English /r/ is thus easier than /l/ because there is less interference from the L1 category.

Although the similarity between the Japanese /ɾ/ and English /r/ and /l/ is thus thought to underlie the learning difficulties of Japanese listeners, it is quite difficult to quantify this degree of similarity. One approach has been to compare acoustic measurements of these phonemes. For example, Lotto et al. [12] revealed that the range of F2 and F3 frequencies of /ɾ/ partially overlapped with that of /r/ and /l/. However, such acoustic measurements have limitations because the Japanese /ɾ/ can use additional dimensions (e.g., a burst), and between-talker differences (e.g., vocal tract length and speaking rate) can introduce variability that is hard to control.

The present experiment instead used a perceptual mapping approach [5], finding best exemplars for /r/, /l/, and /ɾ/ in a 5-dimensional multidimensional acoustic space (F1, F2, F3, closure duration, transition duration). Such an approach can be useful because talker differences are eliminated (i.e., all listeners judge goodness based on the same talker's voice), and the best exemplars are constrained to share common acoustic dimensions. Our aims were to evaluate how close /ɾ/ is to /r/ and /l/, and whether Japanese adults assimilate /r/ or /l/ into their L1 /ɾ/ category.

2. METHOD

2.1. Participants

Thirty-six adult native speakers of Japanese were tested in London (median = 25 years old; range: 19-48). The age at which they started learning English ranged between 6 and 13 (median = 13 years). All participants were brought up in monolingual environments in Japan. Their

accuracy for identifying English /r/ and /l/ was 69.6% correct (measured in a pretest). Thirteen British English speakers were also tested to provide normative data (median = 24 years old; range: 19-62). All of the participants reported no hearing problems.

2.2. Stimuli and apparatus

The stimuli were synthetic C-/ɑ/ syllables embedded in naturally spoken English and Japanese carrier sentences (i.e., *Say [ ] again*, and *mata [ ] to itte kudasai*). The speaker was an English/Japanese bilingual who spent her childhood in both the UK and Japan, and was highly fluent in both languages.

The synthetic syllables were modeled based on the natural talker, using a Klatt synthesizer [10]. A 5-dimensional set of stimuli was created by orthogonally varying F1, F2, F3, closure duration, and transition duration (from the consonantal articulation to the following vowel). The values were chosen so that they would span an acoustic space that included /r/, /l/, and /ɾ/. F1 varied from 123-603 Hz. F2 was always at least 1 ERB [14] greater than F1, and less than 2489 Hz. F3 was always at least 1 ERB greater than F2, and less than 3951 Hz. The closure duration varied from 66 to 209 ms, and the transition duration varied from 5 to 209 ms. The frequency values were quantized in 1-ERB steps and the duration values were quantized with a log spacing (5 steps for closure duration, and 12 for transition duration). There were a total of 60,660 stimuli for each language.

2.3. Procedure

We adapted the goodness optimization procedure that has been used previously for vowels [4] to find the best exemplar of each consonant. On each trial, subjects saw a target consonant presented on the screen (i.e., "R", "L", or "Japanese R"), heard a sentence, and rated on a continuous scale whether the consonant that they heard was close or far away from the target consonant.

On each successive trial, the computer algorithm adjusted the acoustics of the stimuli to converge on a good exemplar of that consonant. The algorithm adjusted the stimuli along 7 vectors (i.e., straight-line paths) through the multidimensional space. Subjects first adjusted the stimuli on a vector that co-varied all 5 dimensions, and passed through a location in the space that corresponded to acoustic measurements of that consonant, allowing subjects to converge on a best exemplar quickly. They then searched on vectors that varied individual dimensions (F3, F2, transition duration, closure duration, and F1). Finally, they searched along a vector that varied all 5 dimensions and allowed them to fine-tune the best exemplar that they had found thus far.

Along each of these search vectors, subjects found the best exemplar along that vector in a series of 5 trials. Subjects first heard the first two most extreme stimuli on the search vector (i.e., at the limits of the synthesis set), in a random order. Then the next 3 trials were chosen based on fitting parabolas to the previous goodness judgments (i.e., to predict which stimulus would be the smallest distance away from being a good exemplar).

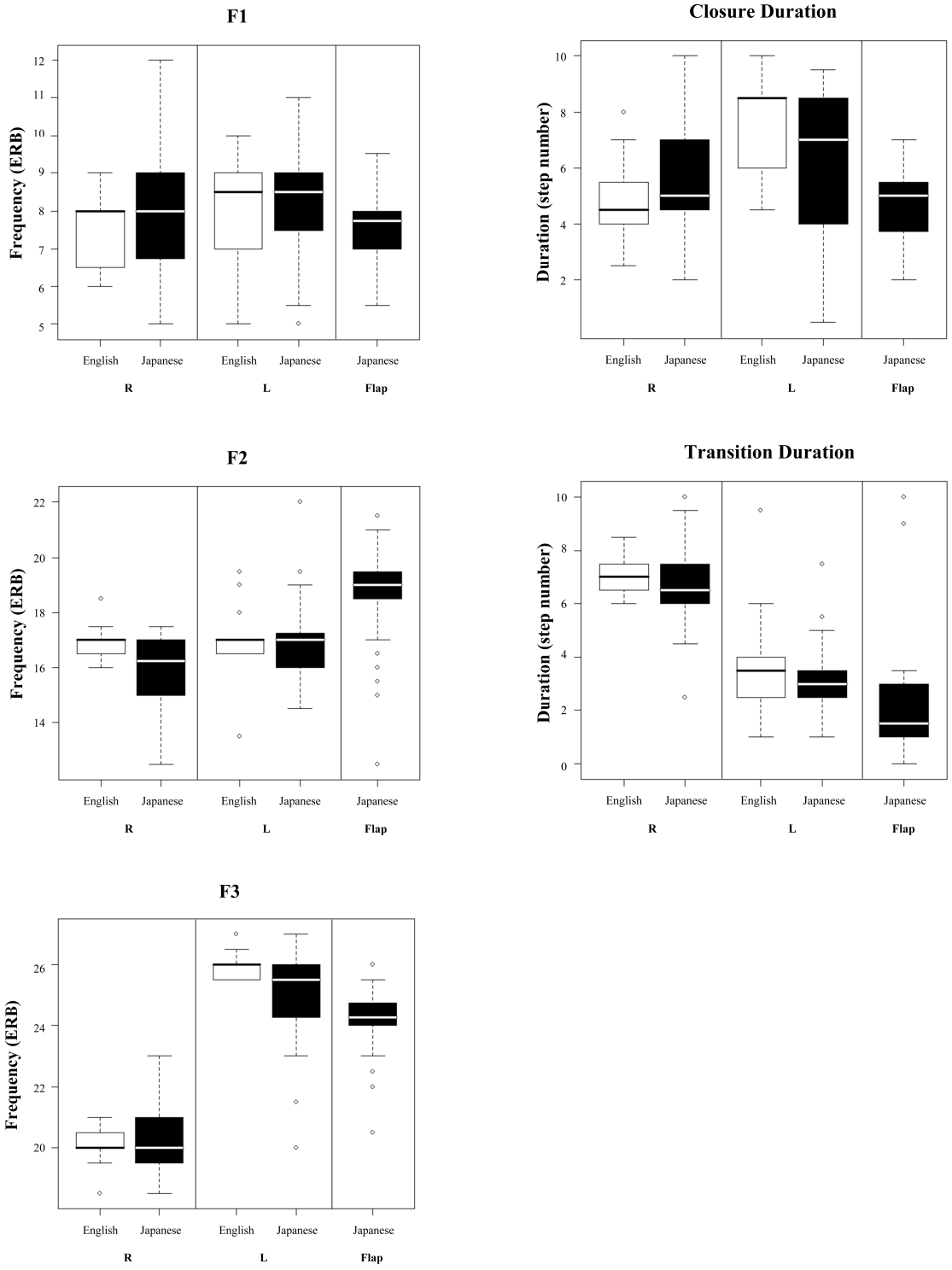
Subjects thus were able to find best exemplars for each consonant after 35 trials. Japanese subjects searched for all three consonants in both carrier sentences (Japanese and English). English speakers searched for /r/ and /l/ only, in both carrier sentences (i.e., they were told just to attend to the target syllable in Japanese, and ignore the fact that they did not understand the entire sentence).

3. RESULTS

Figure 1 displays the acoustic values of the best exemplars of English /r/, /l/, and Japanese /ɾ/ by English and Japanese speakers. Aoyama et al. [1] suggested that the Japanese /ɾ/ was very similar to the English /l/. The present results are somewhat supportive of that claim, in that /ɾ/ and /l/ are broadly similar in F3 and transition duration particularly comparison to /r/. However, paired-samples t-tests revealed that there were significant differences between /ɾ/ and /l/ in F2, $t(35) = -5.181$, $p < 0.01$, F3, $t(35) = 3.128$, $p < 0.01$, closure duration, $t(35) = 3.573$, $p < 0.01$, and transition duration, $t(35) = 3.112$, $p < 0.01$. This indicates that, despite the fact that /ɾ/ and /l/ are broadly similar, the L1 Japanese speakers maintained separate /ɾ/ and /l/ categories in their underlying mental representations. That is, Japanese adults were not simply assimilating /l/ into their L1 /ɾ/ category (i.e., using the same category for both sounds).

The results also demonstrated that the best exemplars of /r/ and /l/ for Japanese speakers were similar to those of native speakers. Separate

Figure 1: Boxplots of the acoustic parameters of best exemplars of /r/, /l/, and /ɾ/ by L1 speakers of English and Japanese



ANOVA analyses analyzed this relationship on each dimension, with language group (English or Japanese) and consonant (/r/ or /l/) as independent factors. On each dimension, there was no significant main effect of language groups or interaction, $p > 0.05$, indicating that Japanese and English speakers did not significantly differ in their underlying /r/ and /l/ categories.

4. DISCUSSION

Despite the fact that Japanese /ɾ/ is broadly similar to English /l/, the results of this experiment indicate that Japanese listeners are able to maintain separate mental representations for these phonemes. That is, they do not simply assimilate English /l/ into their L1 /ɾ/ category. This was surprising because perceived phonetic similarity has been thought to underlie the English /r/-/l/ learning difficulty of Japanese speakers. For example, Best and Strange [3] suggested that Japanese speakers assimilate both /r/ and /l/ to /ɾ/. Moreover, Aoyama [1] claimed that the similarity between /ɾ/ and /l/ made it difficult to learn the /l/ category. Our results instead suggest that Japanese adults, even those that are relatively poor at English /r/-/l/ identification, are able to learn the /l/ category and perceive it as being distinct from /ɾ/.

It was particularly surprising that L1 Japanese and English speakers had similar underlying representations for /r/ and /l/, given that the Japanese listeners were, on average, 69.6% correct when categorizing natural English /r/ and /l/ minimal pairs. Previous work on vowels [8] had suggested that the accuracy of best exemplar locations in an L2 is highly predictive of recognition accuracy, but that more peripheral changes in auditory processing (e.g., cochlear implant simulations) do not change best exemplar locations [9]. This best exemplar task may not be strongly affected by such peripheral distortions because the demands of the task are low (e.g., subjects can repeatedly hear the stimuli and there is no talker variability). The present results thus suggest that Japanese listeners may actually have a good representation of the cues of English /r/ and /l/, but their poor identification performance might be driven by lower-level factors (e.g., the ability to focus attention on the right acoustic cues in naturally variable speech).

5. REFERENCES

[1] Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., Yamada, T. 2004. Perceived phonetic distance and L2 learning: The case of Japanese /r/ and English /l/ and /r/. *J. Phonetics*. 32. 233-250.

[2] Best, C. T. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (ed), *Speech perception and language experience: Issues in cross-language research*. Baltimore: York Press, 171-204.

[3] Best, C. T., Strange, W. 1992. Effetcs of phonological and phonetic factors on cross-language perception of approximants. *J. Phonetics*. 20, 305-330.

[4] Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., Tohkura, Y. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics*. 61, 977-985.

[5] Evans, B. G., Iverson, P., 2004. Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *J. Acoust. Soc. Am*. 115, 352-361.

[6] Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In: Strange, W. (ed), *Speech perception and language experience: Issues in cross-language research*. Baltimore: York Press, 233-277.

[7] Goto. H. 1971. Auditory perception by normal Japanese adults of the sounds “L” and “R”. *Neuropsychologia*. 9, 317-323.

[8] Iverson, P., Evans, B. G. submitted. Learning English vowels with different first-language vowel systems I: Individual differences in the perception of formant targets, formant movement, and duration.

[9] Iverson, P., Smith, C., A., Evans, B. G. 2006. Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement and duration. *J. Acoust. Soc. Am*. 120, 3998-4006.

[10] Klatt, D. H., Klatt, L. C. 1990. Analysis, synthesis, abd perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am*. 87, 820-857.

[11] Logan, J., S., Lively, S. E., Pisoni, D. B. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *J. Acoust. Soc. Am*. 89, 874-886.

[12] Lotto, A. J., Sato, M., Diehl, R. L. 2004. Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/. In Slifka., J., Manuel, S., Matthies, M. (eds), *From Sounds to Sense: 50 + Years of Discoveries in Speech Communication*. Cambriadge, MA: MIT Research Laboratory in Electronics, C-181-C-186.

[13] Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., Fujimura, O. 1975. An effect of language experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception and Psychophisics*. 18, 331-340.

[14] Moore, B., C., J., Glasburg, B. R., and Baer, T. 1997. A model for the prediction of thresholds, loudness, and partial loudness. *J. Audio. Eng. Soc*. 45, 224-240.