Examination of the Relationship between L2 Perception and Production: An Investigation of English /r/-/l/ Perception and Production by Adult Japanese Speakers

Kota Hattori and Paul Iverson

Division of Psychology and Language Sciences, University College London, London, UK
kota.hattori@gmail.com, p.iverson@ucl.ac.uk

Abstract

This study took an individual differences approach to examine the relationship between L2 speech perception and production, with the aim of examining whether they share common underlying representations. All Japanese speakers were assessed in terms of their /r/-/l/ identification, discrimination, best exemplars, and production. The results demonstrated that, although there was a moderate correlation between English /r/-/l/ identification and production, all other perceptual behaviors poorly related to /r/-/l/ production, suggesting that L2 speech perception and production processes and representations may be somewhat autonomous.

Index Terms: speech perception, speech production, second language acquisition

1. Introduction

One of the long-standing issues in second language (L2) acquisition is the relationship between speech perception and production. Some current theoretical frameworks hypothesize that speech perception and production processes are closely related, with common underlying mechanisms. For example, motor theory [1,2,3] states that listeners perceive speech using a phonetic module that represents speech in terms of neuromotor commands to the articulators (i.e., intended articulatory gestures), and that humans produce acoustic signals by using the decoder to generate muscle contractions leading to intended vocal tract shapes. Direct realist theory [4,5] states that listeners perceive speech using a general perceptual system, which directly detects the actual articulatory gestures of the speaker's vocal tract. The theory also states that humans perceive speech as a part of learning to use vocal tracts for communicative purposes.

If speech perception and production processes are closely related, it is possible that L2 learners who are good at perceiving L2 speech sounds are likely to be good at producing the sounds. Some previous studies, indeed, provided corroborative evidence that this is the case although the correlations tend to be only moderate [6,7,8,9]. L2 phonetic training studies demonstrated that such training is effective for enhancing both perception and production abilities [10,11,12,13,14]. However, the amounts of improvement in perception and production due to training are uncorrelated [10]. Therefore, it appears that there is a relationship between L2 perception and production, but the connection is not robust enough to be sure that there are common underlying mechanisms for perception and production.

The present study further examined the relationship between L2 perception and production by assessing individual differences in perception and production of English /r/-/l/ by Japanese speakers. Japanese speakers generally have difficulties in identifying these consonants [15] although they can improve by receiving phonetic training [16,17] or by having extensive exposure to English-speaking environments [18]. Likewise, Japanese speakers generally have difficulties in producing /r/ and /l/ [19,20], and they need decades to overcome these production problems [21]. The aim of present study was to examine whether there are underlying common mechanisms between L2 perception and production, by examining the finer grained correlations between a battery of tests. These tests included recordings of /r/-/l/ words, /r/-/l/ identification, discrimination, and best exemplar search.

2. Method

2.1. Subjects

Forty-seven Japanese speakers were tested in London. Their ages ranged from 18 to 67 years (median = 31 years). They started learning English between 9 and 13 years (median = 13 years), and had received instruction for 5 - 12 years (median = 8 years). All participants were born and raised in Japanese-speaking environments in Japan. They had lived in English-speaking countries between 1 month to 17 years and 2 months (median = 12 months). Additionally, eight British English speakers participated in /r/-/l/ discrimination task. They were raised in the southeast of England. Their ages ranged from 20 to 35 years old (median = 28 years old). None of the participants reported having hearing problems.

2.2. Apparatus

All subjects were tested and recorded using Dell Optiplex GX 260 in a sound-treated room. Recordings were made (16-bit depth; 44,100 samples/sec) with Radio Spares (RS) 249-946 microphone, and Edirol USB Audio Capture UA-25.

2.3. Stimuli

2.3.1. /r/-/l/ identification

The stimuli were initial-position /r/-/l/ minimal-pair words (e.g., *lack* and *rack*) used in previous studies [16,22]. Four British English speakers (2 male and 2 female) recorded a total of 120 minimal-pair words.

2.3.2. /r/-/l/ discrimination

Six pairs of synthetic stimuli were generated based on best exemplars for English /rɑ/ and /lɑ/ from our previous study [22], using a Klatt synthesizer [23]. Half of the pairs varied in F2 frequencies, and the other half varied in F3. In each acoustic dimension, there was one pair within the /rɑ/ category, one pair within the /lɑ/ category, and one pair at the /rɑ/-/lɑ/ boundary. There was a 2-ERB difference for all pairs of stimuli except within-/lɑ/ in the F3 dimension; this pair had a 1-ERB difference to make sure that both stimuli were categorized as /lɑ/.

2.3.3. */r/-/l/ best exemplars*

A set of synthetic C-/ɑ/ syllables from our previous study [22] was used. These synthetic stimuli were generated by varying five acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration), using a Klatt synthesizer [23]. The stimuli were embedded in naturally spoken English carrier sentences (i.e., *Say [ ] again*).

2.4. Procedure

2.4.1. */r/-/l/ identification*

Participants saw minimal-pair words on a computer screen and listened to one of the words. They gave their response by clicking on the spelled words on the screen. They could not replay the stimuli nor did they receive feedback. Each participant completed a short practice session and six experimental sessions.

2.4.2. */r/-/l/ discrimination*

All participants heard the pairs of stimuli with a 300ms ISI and judged whether they were the same or different. Half were same pairs, containing two repetitions of the same stimulus. Half were different pairs. Participants underwent a practice block of 24 trials (two same and two different for each pair) and an experimental block of 192 trials (i.e., 6 pairs x 4 orders x 8 repetitions = 192 trials). The results were analyzed using a differencing model of signal detection theory [24] to calculate sensitivity (*d′*) for each stimulus pair. Note that the within-/lɑ/ F3 pair had only a 1 ERB distance between the stimuli, so the d′ sensitivity was doubled in data analysis to make it comparable to the other pairs.

2.4.3. */r/-/l/ best exemplar search*

We adapted the goodness optimization procedure that has been used in previous studies [22,25] to find the best exemplars of English /r/ and /l/. Subjects saw a target consonant on a computer screen, heard a sentence, and rated goodness of the consonant using a continuous 1 (far away) -7 (close) scale.

The computer algorithm adjusted the acoustics of the stimuli on each trial based on each subject's rating. The algorithm adjusted five acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration), and subjects searched for best exemplars in each acoustic dimension over 35 trials.

In order to provide normative data, we used English data from our previous study [22]. Note that we report the results of F2 and F3 dimensions only due to space limits.

2.4.4. */r/-/l/ production assessment*

Japanese speakers made recordings of 19 initial-position /r/-/l/ minimal pairs. Five British English speakers listened to these recordings and identified consonant categories (i.e., /r/, /l/, /w/, /d/).

F2 and F3 measurements for the /r/ and /l/ productions were made using Praat [26]. In order to provide normative data, we also measured the frequencies for /r/ and /l/ by English speakers from a previous study [16]. F3 frequencies were normalized for data analyses.

3. Results

For /r/-/l/ identification, Japanese speakers demonstrated a wide range of identification accuracy, ranging from 43% to 95% (mean = 71.46%). On average, they correctly identified English /r/ on 70.78 % of trials, and /l/ on 72.13 % of trials.

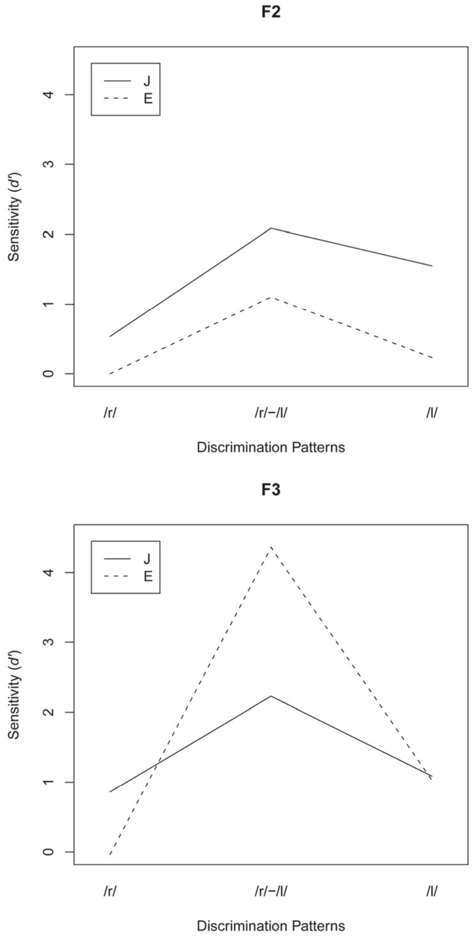

Figure 1: *F2 and F3 discrimination sensitivity within /r/ and /l/ categories and at the English /r/-/l/ boundary for Japanese (J) and English (E) speakers.*

Figure 1 displays Japanese and English speakers' discrimination sensitivity in the F2 and F3 dimensions. For F2, replicating our previous findings [27], Japanese speakers overall had higher sensitivity compared to English speakers. A 2-way ANOVA revealed that there was a main effect of language group, $F(1,53) = 10.11$, $p < 0.01$. There was a main effect of discrimination pattern, $F(2,106) = 20.34$, $p < 0.0001$, indicating that F2 sensitivity was not uniform across the continuum. There was no significant interaction, $p > 0.05$. For F3, English speakers clearly demonstrated higher sensitivity at the boundary than did Japanese speakers. A 2-way ANOVA revealed that there was no main effect of language group, $p > 0.05$. There was a main effect of discrimination pattern, $F(2,106) = 14.86$, $p < 0.0001$. There was a significant interaction between language group and discrimination pattern, $F(2,106) = 4.90$, $p < 0.01$. Simple effects analyses of the interaction revealed that the effect of language group for discrimination sensitivity at the /r/-/l/ boundary was significant, $t(105) = -2.90$, $p < 0.01$, confirming that English speakers had higher discrimination sensitivity at the boundary.

For best exemplars of English /r/ and /l/, Japanese speakers had mental representations similar to English speakers in the

F2 dimension, but their mental representations in the F3 dimension were inaccurate. For F2, a 2-way ANOVA revealed main effects of language group, $F(1,58) = 4.62$, $p < 0.05$, and consonant, $F(1,58) = 12.26$, $p < 0.01$. There was no significant interaction, $p > 0.05$. For F3, the statistical analysis revealed no significant main effect of language group, $p > 0.05$, but it revealed main effect of consonant, $F(1,58) = 289.18$, $p < 0.01$, and a significant interaction, $F(1,58) = 6.78$, p < 0.05. Simple effects analyses of the interaction confirmed that Japanese speakers were particularly inaccurate for /l/, $t(59) = -2.40$, $p <$ 0.05.

For English /r/-/l/ production (i.e., intelligibility of Japanese speakers' /r/-/l/ productions), English speakers consistently identified /l/ productions of Japanese speakers (mean = 91.18%), with the range of 33.68 to 100%. On the other hand, they poorly identified /r/ productions of Japanese speakers (mean = 73.17 %), with the range of 2.01 to 98.95 %.

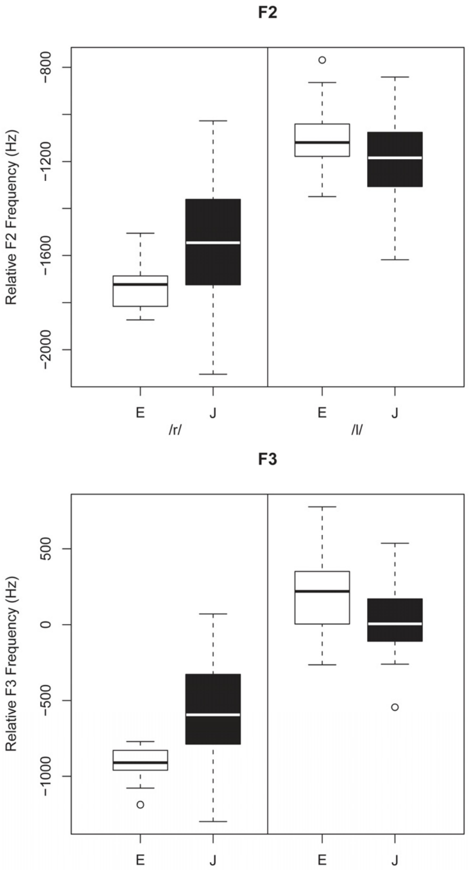

Figure 2: Boxplots of relative *F2 and F3 frequencies of /r/ and /l/ productions by English (E) and Japanese (J) speakers. Boxplots display the medians and quartile ranges of relative frequencies, with outliers marked by circles.*

Figure 2 displays the relative F2 and F3 frequencies of /r/ and /l/ productions by English and Japanese speakers (i.e., normalized to the median F3 for each speaker). Two-way ANOVAs were separately run for each acoustic dimension. The dependent variables were the acoustic dimensions (i.e., F2, F3), language group (Japanese or English) was a between-subjects factor, and consonant (/r/ or /l/) was a within-subjects factor. For F2, there was no main effect of language group, $p >$ 0.05. There was a main effect of consonant, $F(1,60) = 200.08$, $p < 0.01$. There was a significant interaction, $F(1,60) = 18.28$, $p < 0.01$. Simple effects analyses of the interaction revealed that the effect of language group for English /r/ was significant, $t(59) = 4.83$, $p < 0.01$, suggesting that Japanese speakers were inaccurate in producing F2 frequencies. The effect of language group for /l/ was not significant, $p > 0.05$, suggesting that Japanese speakers were similar to English speakers in producing F2 frequencies for /l/. For F3, there was no main effect of language group, $p > 0.05$. There was a main effect of consonant, $F(1,60) = 285.80$, $p < 0.01$. There was a significant interaction, $F(1,60) = 26.75$, $p < 0.01$. Simple effects analyses of the interaction revealed that the effect of language group for /r/ was significant, $t(59) = 5.02$, $p < 0.01$, suggesting that Japanese speakers were inaccurate in producing F3 frequencies. However, the effect of language group for /l/ was not significant, $p > 0.05$, suggesting that Japanese speakers were similar to English speakers in producing F3 frequencies for /l/.

The perception and production data clearly demonstrated substantial individual differences. In order to examine whether there are common underlying representations for perception and production, we calculated average discrimination sensitivity (i.e., sensitivity which is averaged across sensitivities within /rɑ/ and /lɑ/ categories and at the /rɑ/-/lɑ/ boundary), and peak discrimination sensitivity (i.e., sensitivity which the averaged sensitivity of /rɑ/ and /lɑ/ categories is subtracted from sensitivity at the /rɑ/-/lɑ/ boundary) in the F2 and F3 dimensions. We also calculated the accuracies of best exemplars and productions in each acoustic dimension, combining /r/ and /l/ using a Euclidean metric. Multiple correlational analyses were conducted using Bonferroni correction. The critical p-value was set to 0.01. The analyses were divided into seven sets. The first family of test involved /r/-/l/ identification and production. Four families of tests involved /r/-/l/ discrimination (i.e., F2 and F3 average and peak sensitivities) and production. The other two families of tests involved /r/-/l/ best exemplars (i.e., F2 and F3 representation accuracies) and production.

For /r/-/l/ identification, there was a significant correlation with /r/-/l/ production intelligibility, $r = 0.56$, $p < 0.001$. However, there were no significant correlations with F2 production accuracy, r = −0.25, and F3 production accuracy, r = −0.24, $p > 0.01$. These results suggest that L2 perception and production processes may be only partially interrelated with each other.

For F2 average discrimination sensitivity, there were no significant correlations with /r/-/l/ production intelligibility, r = 0.09, and F2 production accuracy, r = 0.03, $p > 0.01$. Likewise, F2 peak discrimination sensitivity was not correlated with production intelligibility, r = −0.18, and F2 production accuracy, r = −0.02, $p > 0.01$. For F3 average discrimination sensitivity, there were no significant correlations with /r/-/l/ production intelligibility, r = −0.14, and F3 production accuracy, r = 0.13. Likewise, F3 peak discrimination sensitivity was not correlated with the production intelligibility, r = 0.21, and F3 production accuracy, r = −0.07, $p > 0.01$. These results suggest that sensitivities to acoustic cues have little to do with L2 production; L2 speech perception and productions processes may employ different representations.

For F2 best exemplar accuracy, there were no significant correlations with /r/-/l/ production intelligibility, r = −0.21, and F2 production accuracy, r = 0.32, $p > 0.01$. Likewise, F3 best exemplar accuracy was not related to the production intelligibility, r = −0.10, and F3 production accuracy, r = −0.23, $p > 0.01$. These results indicate that L2 perceptual representations are poorly related to L2 production or vice versa; L2 speakers may employ independent representations for L2 speech perception and production processes.

4. Discussion

The aim of this study was to investigate whether L2 speech perception and production processes are related, sharing common underlying representations. Replicating previous studies [6,7,8,9], there was a relationship between L2 perception (i.e., how accurately Japanese speakers identified English /r/ and /l/) and production (i.e., how accurately English speakers identified /r/ and /l/ productions of Japanese speakers). Given such a correlation, it seems reasonable to assume that other aspects of perceptual knowledge may be related to L2 production. However, this was the only moderate correlation we found in this study despite the fact that we employed various perceptual and production measurements. Sensitivities to important acoustic cues (i.e., F2 and F3 frequencies) have been considered to be potential factors related to /r/-/l/ identification problems, but they were not related to L2 production abilities. Likewise, accuracies of cognitive representations (i.e., best exemplars) in the F2 and F3 dimensions are factors that are related to identification, but they were not related to L2 production abilities either. That is, perceptual knowledge is not necessarily related to L2 production abilities.

Given these results, it is fair to assume that L2 perception and production processes are not closely related to each other, with independent processes employing different underlying representations. One way to further examine such a view is to conduct L2 production training without any listening tasks. If such training leads to improvement in perception and production, as perceptual training promotes better L2 perception and production accuracies, it is reasonable to assume that L2 perception and productions processes may share common underlying representations. However, if L2 production training leads to improvement in L2 production only, it is reasonable to conclude that L2 speech perception and production are independent processes employing different underlying mechanisms to process L2 speech.

5. References

[1] Liberman, A. M., Cooper, F. S., Shanksweiler, D. P., and Studdert-Kennedy, M. (1967). "Perception of speech code," *Psychological Review,* **74**, 431-461.

[2] Liberman, A. M., and Mattingly, I. G. (1985). "The motor theory of speech perception revised," Cognition, **21**, 1-36.

[3] Liberman, A. M., and Mattingly, I. G. (1989). "A specialization of speech perception," *Science.* **243**, 489-494.

[4] Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research,* edited by W. Strange (York, Baltimore), pp.171-204.

[5] Fowler, C. A. (1986). "An event approach to the study of speech perception from a direct realist perspective," *J. Phonetics*, **14**, 3-28.

[6] Flege, J. E. (1993). "Production and perception of a novel second-language phonetic contrast," *J. Acoust. Soc. Am.* **93**, 1589-1608.

[7] Flege, J. E., and Schmidt, A. M. (1995). "Native speakers of Spanish show rate-dependent processing of English stop consonants," *Phonetica*, **52**, 90-111.

[8] Schmidt, A. M., and Flege, J. E. (1995). "Effects of speaking rate changes on native and non-native production," *Phonetica*, **52**, 41-54.

[9] Flege, J. E., Bohn, O-S., and Jang, S. (1997). "The production and perception of English vowels by native speakers of German, Korean, Mandarin, and Spanish," *J. Phonetics*. **25**, 437-470.

[10] Bradlow, A. R., Pisoni, D. B., Yamada, R. A., and Tohkura, Y. (1997). "Training Japanese listeners to identify English /r/ and /l/ Ⅳ : Some effects of perceptual learning on speech production," *J. Acoust. Soc. Am.* **101**, 2299-2310.

[11] Iverson, P., Pinet, M., & Evans, B. G. (submitted). "Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels," *Applied Psycholinguistics*.

[12] Lengeris, A. (2009). *Individual differences in second-language vowel learning*. Unpublished PhD thesis, Department of Speech, Hearing, and Phonetic Sciences, Division of Psychology and Language Sciences, University College London.

[13] Rochet, B. L. (1995). "Perception and production of second-language speech sounds by adults," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research,* edited by W. Strange (York, Baltimore), pp. 379-410.

[14] Wang, Y., Jongman, A., and Sereno, J. A. (2003). "Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training," *J. Acoust. Soc. Am.* **113**, 1033-1043.

[15] Goto, H., (1971). "Auditory perception by normal Japanese adults of the sounds "L" and "R"," *Neuropsychologia*. **9**, 317-323.

[16] Iverson, P., Hazan, V., and Bannister, K. (2005). "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults," *J. Acoust. Soc. Am.* **118**, 3267-3278.

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

[18] Yamada, R. A. (1995). "Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research,* edited by W. Strange (York, Baltimore), pp. 305-320.

[19] 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 J. Slifka, S. Manuel, and M. Matthies (Eds.), *From Sounds to Sense: 50 + Years of Discoveries in Speech Communication* (pp. C-181-C186). Cambridge, MA: MIT Research Laboratory in Electronics.

[20] Masaki, S., Akahane-Yamada, R., Tiede, M. K., Shimada, Y., and Fujimoto, I. (1996). "An MRI-based analysis of the English /r/ and /l/ articulations," *The Ninth International Conference on Spoken Language Processing*, Pittsburgh, USA, 1581-1584.

[21] Flege, J. E., Takagi, N., and Mann, V. (1995). "Japanese adults can learn to produce English /r/ and /l/ accurately," *Language and Speech*, **38**, 25-55.

[22] Hattori, K., and Iverson, P. (2009). "English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy," *J. Acoust. Soc. Am.* **125**, 469-479.

[23] Klatt, D. H., & Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.*, **87**, 820–857.

[24] Macmillan, N. A., and Creelman, C. D. (1991). *Detection theory: a user's guide.* New York: Cambridge University Press.

[25] Evans, B. G., and 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.

[26] Boersma, P., and Weenik, D. (2010). "Praat: Doing phonetic by computer," retrieved from http:www.praat.org (Last viewed July, 2010).

[27] Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition, **87**, B47-B57.