EFFECTS OF PERCEPTUAL ASSIMILATION: THE PERCEPTION OF ENGLISH /æ/, /ʌ/, AND /ɑ/ BY JAPANESE SPEAKERS

Yasuaki Shinohara[1], Chao Han[2], Arild Hestvik[2]

Waseda University[1], University of Delaware[2]
y.shinohara@waseda.jp

ABSTRACT

In the present study, Japanese speakers' perception of English vowels /æ/, /ʌ/, and /ɑ/ was examined. According to the Perceptual Assimilation Model [1, 2, 3], discriminating the /ʌ/-/ɑ/ contrast should be more difficult than the /æ/-/ʌ/ contrast, although all three vowels are categorized as a single vowel /a/ in Japanese. In the experiment, an identification test with a resynthesized stimulus continuum of vowels varying in F2 was given to both Japanese and English speakers. Based on the results, three stimuli of English /æ/, /ʌ/, and /ɑ/ for which acoustic distance was controlled were selected and used for a discrimination test. The results demonstrated that while English speakers discriminated both the /æ/-/ʌ/ and /ʌ/-/ɑ/ contrasts equally well, discriminating the /ʌ/-/ɑ/ contrast was more difficult than the /æ/-/ʌ/ contrast for Japanese speakers. That is because English /ʌ/ and /ɑ/ were equally good exemplars of Japanese /a/, compared to the English /æ/-/ʌ/ contrast.

Keywords: perceptual assimilation, non-native vowels, Japanese speakers, discrimination

1. INTRODUCTION

Identifying and discriminating non-native phonemes are often difficult for any language speakers. However, the degree of the difficulty in discriminating non-native phonemes can be predicted. According to the Perceptual Assimilation Model (PAM) [1, 2, 3], when people perceive unfamiliar non-native phonemes, they tend to assimilate them into the most articulatorily similar first-language (L1) phonemes, and therefore the discriminability of non-native phonemes depends on the listeners' L1 phonemic categories. For example, when listeners assimilate two non-native phonemes into two different L1 phonemes, i.e., Two-Category (TC) assimilation type in PAM terms, it will be easy for listeners to discriminate them. On the other hand, when listeners assimilate two non-native phonemes into one L1 phonemic category, discriminating those phonemes will be harder. If the two phonemes are equally good exemplars of a single L1 phoneme, i.e., Single-Category (SC) assimilation type, discrimination will be difficult. However, if the similarity to an L1 phoneme is different between the two non-native phonemes, i.e., Category-Goodness (CG) assimilation type, discriminating those two phonemes will be relatively easy.

For decades since the PAM was established, much research has been conducted to examine the prediction of the discriminability based on the assimilation patterns for consonant [4, 5, 8, 10, 19] and vowel perception [6, 7, 12, 13, 17, 18, 20]. For example, Japanese speakers often judge the most similar Japanese vowel for each of English /æ/, /ʌ/, and /ɑ/ is Japanese /a/ [12, 17], and vowel contrasts which belong to the CG assimilation type are easier to discriminate than vowel contrasts which belong to the SC assimilation type [12]. However, since the stimuli used in Lengeris [12] were natural recordings, the acoustic distance between the vowel stimuli was not controlled for. Therefore, the discriminability may not be attributed only to the listener's phonetic perception, but it may also be affected by the acoustic differences in the stimuli. Although many studies have been conducted using synthetic or resynthesized stimuli [9, 14], none of them have examined Japanese speakers' perception of the English /æ/, /ʌ/, and /ɑ/ vowels for which acoustic distance was controlled.

The present study examined Japanese speakers' discrimination accuracy of English /æ/, /ʌ/, and /ɑ/ vowels for which acoustic cues were manipulated. Using resynthesized stimuli, the perceptual assimilation patterns for both the /æ/-/ʌ/ contrast and the /ʌ/-/ɑ/ contrast were examined with an identification test with goodness rating first, and discrimination accuracy for each vowel contrast was investigated for both English and Japanese speakers. Following the PAM, it was hypothesized that if the /æ/-/ʌ/ contrast belongs to the CG assimilation type and the /ʌ/-/ɑ/ contrast belongs to the SC assimilation type [12], the /ʌ/-/ɑ/ contrast should be more difficult to discriminate than the /æ/-/ʌ/ contrast.

2. METHOD

2.1. Participants

Thirty Japanese speakers were recruited at Waseda University, Tokyo Japan, and 27 American English speakers participated at the University of Delaware in the U.S. The Japanese subjects were native

monolingual speakers of Japanese (15 females, 15 males) aged from 18 to 24 years (M = 20.2, SD = 1.49), and the U.S. subjects were monolingual English speakers (25 females, 2 male) aged from 18 to 22 years (M = 19.6, SD = 1.21). All participants reported no history of speech or hearing impairments; no experience living outside their own countries for more than two months; with both parents also being the same native language speakers as the participants; and speaking only their native language in daily life.

2.2. Stimuli

Using LPC analysis and resynthesis in Praat, a stimulus continuum varying in the F2 frequency was created. Neutral LPC residuals from a female speaker with the duration of 308 ms and 138 ms were filtered with stable formant information, F1 at 979 Hz, F3 at 2886 Hz, and F4 at 4151 Hz. F2 was set at 995 Hz for a front vowel and set at 2698 Hz for a back vowel. F2 frequency was then interpolated between the two vowels into 66 stimuli in Hertz. The fundamental frequency was 220 Hz, and the stimuli with 308 ms were used for the identification test. The stimuli with 138 ms were used for the discrimination test.

2.3. Procedure

2.3.1. Identification with goodness rating

In the first identification task, participants randomly heard 66 stimuli from the stimulus continuum. For American English speakers tested in Delaware, five words (*head, had, hud, hod, hawed*) were displayed on the screen. Participants clicked on the one which included the vowel they thought they heard, then gave a 7-point goodness rating for each token (1 for worst, 7 for best). It should be noted that English speakers in Delaware often do not make a distinction between /ɑ/ and /ɔ/ [11]. For Japanese speakers, three Japanese moraic letters such as 'へ' /he/, 'は' /ha/, and 'ほ' /ho/ were displayed. Japanese speakers then clicked on the letter which included the Japanese vowel they thought they heard, and gave a goodness rating.

Based on the results of the identification test with goodness rating, three stimuli were selected for the subsequent discrimination test. Each of those three stimuli was identified as /æ/, /ʌ/, or /ɑ/ (or /ɔ/) for English speakers, but all of them were identified as Japanese /a/ for Japanese speakers (see the Result section 3.1). In addition, to test the hypothesis based on the PAM (SC vs. CG assimilation type), goodness rating results were taken into account for the stimuli selection. The selected English /ʌ/-/ɑ/ stimuli were equally good exemplars of Japanese /a/ (SC assimilation type), and therefore, should be hard to discriminate. On the other hand, English /æ/ and /ʌ/ were not equally good exemplars of Japanese /a/, thus easier to discriminate (CG assimilation type; see the Results section 3.1). The acoustic distance between /æ/ and /ʌ/ in F2 was set to be the same as the one between /ʌ/ and /ɑ/ in Hertz (but not in Bark scale). Sixty-six stimuli repeated four times each provided participants with 264 trials in this identification task.

2.3.2. Discrimination

Participants who took the identification test also participated in the discrimination test. However, only 21 Japanese speakers out of 30 took the discrimination test.

Both English speakers' and Japanese speakers' discrimination accuracies of the two stimulus pairs (i.e., /æ/ vs. /ʌ/, /ʌ/ vs. /ɑ/) were measured. Numbers 1, 2, and 3 were displayed on the screen, and participants heard three stimuli from either of the two stimulus pairs with a 1.2 sec interstimulus interval (ISI). They were asked to click on the one which sounded different from the other two stimuli. Since the F2 difference in Hertz between two stimuli was the same between the two stimulus pairs, the discriminability should be acoustically the same. However, discriminating the /ʌ/-/ɑ/ contrast was hypothesized to be more difficult than the /æ/-/ʌ/ contrast, due to the perceptual assimilation type difference (i.e., CG vs. SC). Participants had 72 trials for this discrimination task in total (2 stimulus pairs * 3 positions of correct answers * for both vowels * 6 repetitions of each token).

3. RESULTS

3.1. Identification with goodness rating

Figure 1 displays American English speakers' identification of the 66 stimulus continuum. Since 11 out of 27 English-speaking participants reported they do not have a distinction between English /ɑ/ and /ɔ/, the identification rates for those two vowels were combined. Stimuli with higher F2 were identified as front vowels (i.e., /e/ and /æ/), and stimuli with lower F2 were identified as back vowels (i.e., /ɑ/ and /ɔ/). English speakers identified the stimuli with F2 from 1571 Hz to 1807 Hz as the English vowel /ʌ/ more frequently than the other vowel stimuli.

Figure 2 displays the identification of the 66 stimuli from the continuum by Japanese speakers. Stimuli with higher F2 were identified as Japanese /e/, and stimuli with lower F2 were identified as Japanese /o/. Japanese speakers identified the stimuli with F2 from 1415 Hz to 2070 Hz as Japanese /a/, more frequently than /e/ and /o/.

Based on the identification results, we selected three vowel stimuli for the subsequent discrimination

test. Those stimuli were identified as /æ/, /ʌ/, and /ɑ/ (or /ɔ/) by native English speakers, whereas all of them were categorized as Japanese /a/ by Japanese speakers. The F2 frequencies of the three stimuli were 2017 Hz, 1755 Hz, and 1493 Hz, respectively.

Figure 1: Identification of the 66 stimulus continuum of vowels varying in F2 by English speakers. The dashed lines represent the F2 frequencies of the three stimuli selected for the subsequent discrimination test.

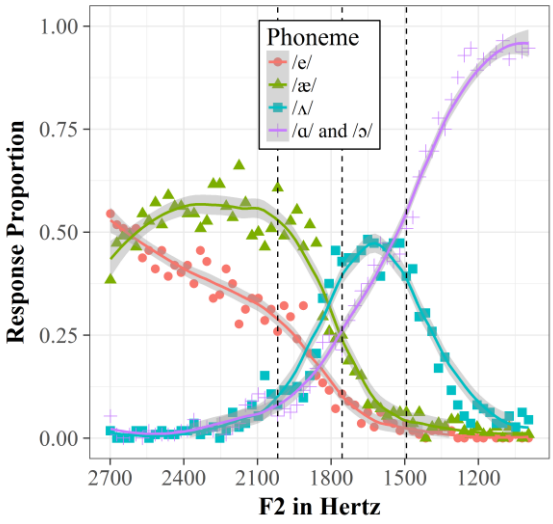


Figure 2: Identification of the 66 stimulus continuum of vowels varying in F2 by Japanese speakers. The dashed lines represent the F2 frequencies of the three stimuli selected for the subsequent discrimination test.

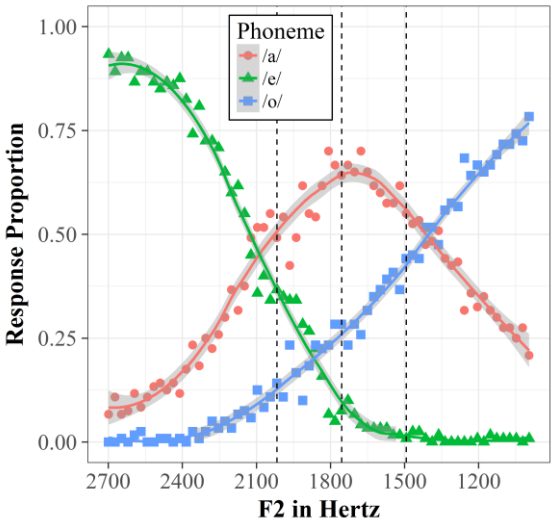
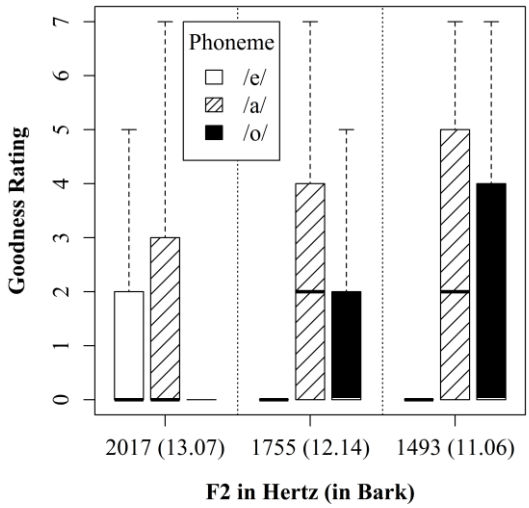


Figure 3 displays the goodness rating of these three stimuli as Japanese vowels judged by Japanese speakers. After Japanese participants identified each of 66 stimuli as either of /e/, /a/, or /o/, they gave a goodness rating (1 for worst, 7 for best) as the selected vowel. A linear mixed effects model was used for the analysis of the goodness rating, and rating 0 (not identified) was added after the experiment for this statistical analysis. The fixed factors were phoneme (identified as /e/, /a/, /o/), stimulus (three levels with F2 at 2017 Hz, 1755 Hz, 1493 Hz), and the interaction of phoneme and stimulus. The random factor was the intercept of Japanese participants. The results demonstrated that there were significant main effects of phoneme, $\chi^2(2) = 153.44$, $p < .001$, and of stimulus, $\chi^2(2) = 8.22$, $p = .016$. The significant interaction of phoneme and stimulus, $\chi^2(4) = 72.29$, $p < .001$, suggests that Japanese participants gave a different goodness rating for each stimulus as a Japanese vowel according to F2.

Figure 3: Goodness rating of the selected three stimuli (F2: 2017 Hz, 1755 Hz, 1493 Hz) as Japanese vowels (/e/, /a/, /o/) given by Japanese speakers.



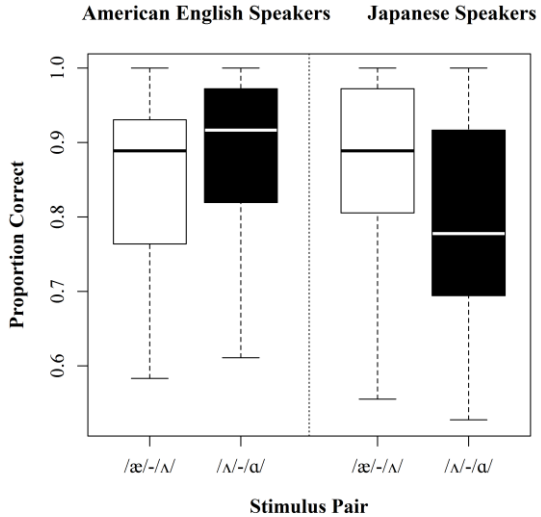
Post-hoc analyses were conducted only for the stimuli identified as Japanese /a/. The results demonstrated that there was a significant difference in the goodness rating as Japanese /a/ between those three stimuli, $\chi^2(2) = 9.59$, $p < .01$. Pairwise comparisons for stimulus (2017 Hz vs. 1755 Hz, 1755 Hz vs. 1493 Hz) showed that there was no significant difference in the goodness rating as Japanese /a/ between stimulus with F2 at 1755 Hz and the one with F2 at 1493 Hz, $\chi^2(1) = 1.18$, $p > .05$. However, there was a significant difference in the goodness rating as Japanese /a/ between the stimulus with F2 at 2017 Hz and the one with F2 at 1755 Hz, $\chi^2(1) = 4.85$, $p = .028$. These results suggest that the stimulus with F2 at 2017 Hz was perceived as the worse exemplar of Japanese /a/ than was the one with 1755 Hz (CG assimilation type), but stimulus with F2 at 1493 Hz was an equally good exemplar of Japanese /a/ as the one with 1755 Hz (SC assimilation type). Although the acoustic distance

between the /æ/ and /ʌ/ stimuli was the same as that between the /ʌ/ and /ɑ/ stimuli in F2 (Hertz), the phonetic perceptual assimilation type was different between those two stimulus pairs.

3.2. Discrimination

Figure 4 displays the discrimination accuracy of the two stimulus pairs (i.e., the /æ/-/ʌ/ contrast, the /ʌ/-/ɑ/ contrast) by American English speakers and Japanese speakers. A logistic mixed effects model based on the correct/incorrect binomial responses was used for this analysis. The fixed factors were language group (English speakers, Japanese speakers), stimulus pair (/æ/ vs. /ʌ/, /ʌ/ vs. /ɑ/), and the interaction of language group and stimulus pair. The random factors were the intercepts of subject and stimulus. The logistic mixed effects model demonstrated that there was no significant main effect of language group, $\chi^2(1) = 1.34$, $p > .05$, or stimulus pair, $\chi^2(1) = 0.05$, $p > .05$. However, the interaction between language group and stimulus pair was significant, $\chi^2(1) = 17.69$, $p < .001$, suggesting that the discrimination accuracy was different between the two stimulus pairs for Japanese speakers, compared to English speakers. That is, discriminating the /ʌ/-/ɑ/ contrast was more difficult than the /æ/-/ʌ/ contrast for Japanese speakers, whereas such difference was not observed among English speakers.

Figure 4: Discrimination accuracies of the two English vowel contrasts (/æ/ vs. /ʌ/, /ʌ/ vs. /ɑ/) by American English speakers and Japanese speakers.



4. DISCUSSION

The present study examined English and Japanese speakers' identification and discrimination of vowel stimuli for which only F2 was manipulated. Based on the identification test results, three stimuli which belong to three different English vowel categories, /æ/, /ʌ/, and /ɑ/ but identified as a single Japanese /a/ were selected for a subsequent discrimination test. The discrimination test results demonstrated that English speakers discriminated both the /æ/-/ʌ/ and /ʌ/-/ɑ/ contrasts equally well, whereas Japanese speakers had more difficulty discriminating the /ʌ/-/ɑ/ contrast than the /æ/-/ʌ/ contrast, even though the acoustic distance between two stimuli was the same in both pairs.

The difference in the discrimination accuracy between English and Japanese speakers can be attributed to their L1 vowel categories. On the one hand, English speakers have the three vowels /æ/, /ʌ/, and /ɑ/ as different English phonemes; consequently they should be able to discriminate both the /æ/-/ʌ/ and /ʌ/-/ɑ/ contrasts well. However, for Japanese speakers, all the sounds are categorized as a single Japanese /a/ vowel. As described in Introduction, according to the Perceptual Assimilation Model [1, 2, 3], if two non-native phonemes are equally good exemplars of a single L1 phoneme (i.e., SC assimilation type), its discrimination is predicted to be difficult. However, if the goodness rating as an L1 phoneme is different for the two non-native phonemes (i.e., CG assimilation type), its discrimination should be relatively easy. This was confirmed by the current experiment. Japanese speakers judged the most similar L1 vowel for each of English /æ/, /ʌ/, and /ɑ/ was Japanese /a/. Nevertheless, the goodness rating results showed that English /ʌ/ and /ɑ/ were equally good exemplars of Japanese /a/ for Japanese speakers (SC assimilation type), but English /æ/ and /ʌ/ were significantly different in the goodness rating as exemplars of Japanese /a/ (CG assimilation type). Therefore, the /ʌ/-/ɑ/ contrast was more difficult for Japanese speakers to discriminate than the /æ/-/ʌ/ contrast.

As the PAM posits, if the phonetic-perceptual difference from the L1 phoneme influences the perception of non-native phoneme contrasts, the discrimination accuracy should be different based on the three stimuli selected. In order to thoroughly test the effects of perceptual assimilation, we should expand the range of stimuli with more acoustic parameters, as well as testing the relevant theoretical frameworks such as the Natural Referent Vowel (NRV) framework [15, 16, 20].

5. ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number 16K16884 and Waseda University Grant for Special Research Projects No. 2016K-202, 2017K-221, and 2018K-159. We really thank our participants, assistants, and colleagues for their supports.

6. REFERENCES

[1] Best, C. T. 1994a. Learning to perceive the sound pattern of English. In: Rovee-Collie, C., Lipsitt, L. P. (eds.), *Advances in infancy research* 8. Hillsdale, NJ: Ablex Publishers, 217–304.

[2] Best, C. T. 1994b. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In: Goodman, J. C., Nusbaum, H. C. (eds.), *The development of speech perception: The transition from speech sound to spoken words*. Cambridge, MA: MIT Press, 167–224.

[3] Best, C. T. 1995. A direct realist perspective on cross-language speech perception. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, 171–204.

[4] Best, C. T., McRoberts, G. W., Goodell, E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *J. Acoust. Soc. Am.* 109, 775–794.

[5] Best, C. T., Strange, W. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20, 305–330.

[6] Escudero, P., Williams, D. 2011. Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners. *J. Acoust. Soc. Am. 129*, EL1-EL7.

[7] Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P., Brattico, E. 2014. Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: a behavioral and ERP study. *Frontiers in Human Neuroscience* 8, 1–14.

[8] Guion, S. G., Flege, J. E., Akahane-Yamada, R., Pruitt, J. C. 2000. An investigation of current models of second language speech perception: the case of Japanese adults' perception of English consonants. *J. Acoust. Soc. Am.* 107, 2711–2724.

[9] Ikeda, K., Hayashi, A., Hashimoto, S., Otomo, K., Kanno, A. 2002. Asymmetrical mismatch negativity in humans as determined by phonetic but not physical difference. *Neuroscience Letters* 321, 133–136.

[10] Komaki, R., Akahane-Yamada, R., Choi, Y. 1999. Effects of native language on the perception of American English /r/ and /l/: Cross-language comparison between Korean and Japanese. *Proc. 14th ICPhS*, 1429–1432.

[11] Labov, W., Ash, S., Boberg, C. 2006. *The Atlas of North American English*. Berlin: Mouton de Gruyter.

[12] Lengeris, A. 2009. Perceptual assimilation and L2 learning: Evidence from the perception of Southern British english vowels by native speakers of Greek and Japanese. *Phonetica* 66, 169–187.

[13] Mokari, P. G., Werner, S. 2017. Perceptual assimilation predicts acquisition of foreign language sounds: The case of Azerbaijani learners' production and perception of Standard Southern British English vowels. *Lingua* 185, 81–95.

[14] Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Alku, P., Ilmoniemi R. J., Luuk, A, Allik, J., Sinkkonen, J., Alho, K. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434.

[15] Polka, L., Bohn, O. S. 2003. Asymmetries in vowel perception. *Speech Communication* 41, 221–231.

[16] Polka, L., Bohn, O. S. 2011. Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics* 39, 467–478.

[17] Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., Jenkins, J. J. 1998. Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics* 26, 311–344.

[18] Sun, L., Heuven, V. J. van. 2007. Percetual assimilation of English vowels by Chinese listeners: Can native-language interference be predicted? *Linguistics in the Netherlands 2007*, 150–161.

[19] Takagi, N. 1993. *Perception of American English /r/ and /l/ by adult Japanese learners of English: A unified view*. (Doctoral thesis, University of California, Irvine, United States of America). Retrieved from http://search.proquest.com/dissertations/docview/304042958/fulltextPDF/14286C787BA778B2C85/1?accountid=14511

[20] Tyler, M. D., Best, C. T., Faber, A., Levitt, A. G. 2014. Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71, 4–21.