

DOES REPEATED EXPOSURE TO SEGMENTAL SOUNDS IMPROVE PERCEPTUAL ABILITY IN NON-NATIVE SPEAKERS?

Hiroki Fujita¹, Ruri Ueda², Ken-ichi Hashimoto²

¹University of Reading, ²Osaka Kyoiku University
h.fujita@pgr.reading.ac.uk

ABSTRACT

Previous studies have shown that native (L1) speakers adapt to regional dialects and foreign accents rapidly. Some studies show that non-native (L2) speakers also adapt to foreign accents, but little is known about how they adapt to L2 segmental sounds difficult to perceive in the development of L2 phonetic acquisition. This study explores whether adaptation (learning effects) occurs as Japanese language learners of English are exposed to the English phonetic contrasts /ɑ-/æ/ and /s/-/θ/ posing perceptual difficulties for them during an identification task. The results showed decreased response times over the course of the experiment only when participants were exposed to the same set of minimal pairs. Accuracy rates did not change, irrespective of whether target items were identical or different. Taken together, the present study observed reduced auditory processing cost, however, this processing advantage resulted from enhanced comprehensibility rather than from the improvement of perceptual ability.

Keywords: Speech perception; L2 perceptual learning; Speech adaptation

1. INTRODUCTION

Successful language comprehension requires perceiving similar phonetic categories correctly. However, precise perception is not always easy even for L1 speakers when speech deviates from native norms like foreign-accented speech and regional accents, which often contain multiple departures from L1 speakers' general accent and are thus often perceived less correctly or as unintelligible [15]. Nonetheless, there is evidence that (short-term) repeated exposure to such unfamiliar accent reduces perceptual difficulty. This learning effect is known as adaptation [7, 11, 19].

Similarly, L2 speakers often have difficulty perceiving phonetic segments precisely when those sounds are not available in auditory representations that have been developed in L2 speakers' first languages [5, 10]. For example, Japanese language learners of English often misperceive the /θ/ sound as the /s/ sound, as a voiceless dental fricative is not

present in Japanese and therefore, /θ/ is often assimilated into /s/ [12]. This L2 fuzzy phonetic encoding property is evidenced by phonolexical representation studies, which show that unlike L1 speakers, L2 speakers show priming effects in a lexical decision task even when primes and targets are different but share a similar phonetic segment (i.e. minimal pair) [e.g. 8].

However, this does not necessarily mean that improving L2 perceptual ability is not possible. Many studies on L2 perceptual training or the effect of L2 language experience on L2 phonetic acquisition report that identification or discrimination performance improves after a certain amount of training or exposure to the phonetic segments [6].

Nevertheless, the contribution of repeated exposure to phonetic segments in L2 is still unclear, as training potentially contains focused effects compared with actual word learning [17]. To examine the effect of mere repeated exposure to segmental sounds in L2 perceptual learning, we explored whether and to what extent Japanese language learners of English adapt to /ɑ-/æ/ and /s/-/θ/ contrasts that they have difficulty perceiving, using the two-alternative forced identification task (Figure 1).

2. METHOD

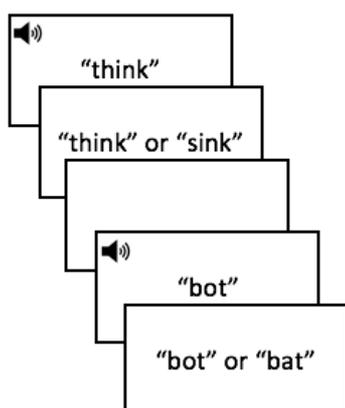
2.1. Participants

In all, 38 Japanese language learners of English (18 males, mean age 19) took part in the experiment for course credits. One participant, whose first language was not Japanese, was removed before data analysis. Each participant filled in a language background questionnaire before the experiment, which showed that the average age when participants started learning English was 10 (range = 3–13, SD = 3). Participants also took an English proficiency test called 'Progress' prior to the experiment; the results indicated that they were low-to-intermediate English language learners. All participants provided informed consent before the experiment and were instructed to be able to withdraw at any time during the experiment.

2.2. Materials

The two-alternative forced identification task contained 64 minimally-paired words consisting of either /s-/θ/ (e.g. “sing” vs. “thing”) or /æ-/ɑ/ (e.g. “black” vs. “block”) contrasts. In addition to these target items, 32 minimally-paired words, which are known not to cause perceptual difficulty for native Japanese speakers (e.g. /i-/e/) were created as distractors and distributed in each block equally. These distractors were used not only to avoid participants noticing the purposes of the experiment but also to calculate adjusted response times (see 2.4. Data analysis). Both target and distractor items were matched for length, frequency, and lexical decision speed within each condition according to the English Lexicon Project [2]. All materials were recorded by a male native speaker of American English in a quiet room.

Figure 1: Sequence of two trial examples in the two-alternative forced identification task.



2.3. Procedure

All words were displayed on a PC monitor and responses were collected by the IBEX Farm web-based experimental presentation platform, developed and managed by Alex Drummond (www.spellout.net/ibexfarm/docs). Although the IBEX Farm platform is often used when participant recruitment and data collection are conducted over the web, by distributing the experiment link to as many people as possible to gain a large sample size, the experiment reported in the present study was conducted in the laboratory-based fashion. At the beginning of the experiment, participants received oral and written onscreen instructions on the procedure in Japanese, the participants' first language. To familiarise them with the procedure, several practice items were provided before the

experiment began. During the two-alternative forced identification task, participants listened to a word followed by two visual probe words, and judged which of the visually presented words corresponded to the aurally presented word. They were instructed to press the '1' key for the first option and the '2' key for the second option. The experiment took approximately 15-20 minutes.

2.4. Data analysis

Collected data were analysed in R [16] with the lme4 package [4] by fitting linear mixed effect models [1] to response times and generalised linear mixed effect models [13] to accuracy rates. Response times longer than 10000ms and greater than three standard deviations from each participant's mean response times within each condition were removed based on visual inspections. These procedures resulted in an exclusion of approximately 2% of the data. After this removal procedure, response times were adjusted as in some previous studies [7, 19] by subtracting each participant's mean response times of target words from those of distractor words within each block to remove the adaptation effect to the task itself. Each model consisted of fixed effects of sound (vowel/consonant) and block (Blocks 1-4). These fixed effects were deviation-coded for comparisons between Block 1 and Block 4 before being included in the model. For comparisons of Blocks 1-3, Helmert contrast coding was adopted for block to compare between Block 1 and Block 2, and between Block 3 and the mean of Block 1 and Block 2. All models were initially constructed with maximal random effects structures [3]. If the model failed to converge, correlation parameters were removed, and if it still did not converge, a random effect accounting for the least variance was removed one by one until it successfully converged. A fixed effect was considered significant if *p* values, estimated from the Satterthwaite approximation implemented by the lmerTest package [14] for linear mixed effect models and the Laplace Approximation for generalised linear mixed effect models, were at or below .05.

Figure 2: Adjusted response times by sound and block. Error bars indicate standard errors.

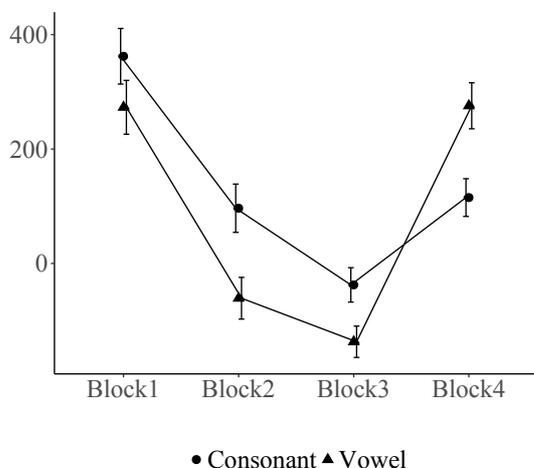
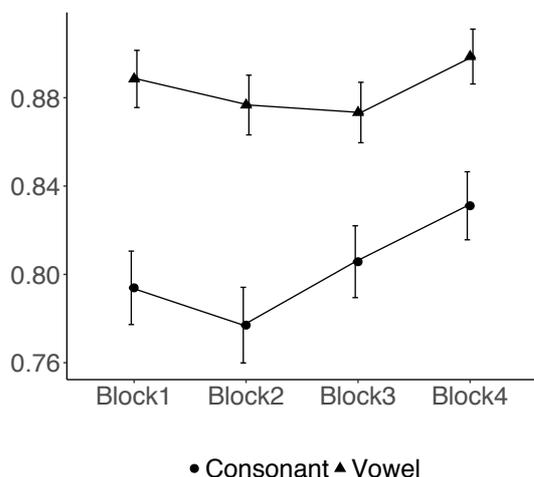


Figure 3: Accuracy rates by sound and block. Error bars indicate standard errors.



3. RESULTS AND DISCUSSION

Mean accuracy rates to distractor items were 99% (range = 91–100, $SD = 0.12$), which indicated that participants paid attention to the experiment.

For Blocks 1–3, although there was some weak tendency towards longer response times for consonant than for vowel contrasts ($t = 1.63$, $SE = 69$, $p = .11$), this effect is difficult to treat, as length and familiarity are not controlled for across vowels and consonants. We also observed a main effect of block, as response times decreased over the course of the experiment (Block 1 vs. Block 2: $t = 2.29$, $SE = 45$, $p < .05$; Block 3 vs. mean of Blocks 1 and 2: $t = 3.72$, $SE = 96$, $p < .001$), showing learning effects. Regarding accuracy rates, the model yielded a main effect of sound due to lower accuracy rates for consonant than for vowel contrasts ($z = 3.10$, $SE = 0.36$, $p < .01$). However, there was no effect of block nor interaction between them.

For comparisons of Block 1 and Block 4, there was a main effect of sound in accuracy rates due to higher error rates for consonant than for vowel contrasts ($z = 2.26$, $SE = 0.04$, $p < .05$). However, no learning effects were observed in both response times and accuracy rates unlike Blocks 1–3.

As described, decreased response times with no interaction with sound across Blocks 1–3 suggest that learning effects occurred as participants engaged in the two-alternative forced identification task, irrespective of vowels and consonants. This finding is compatible with some previous L1 studies showing rapid adaptation to a foreign accent [7, 19]. However, contrary to [19], accuracy rates did not improve across the blocks. As mentioned, several phonological representation studies show that unlike L1 speakers, L2 speakers may have fuzzy representations of similar L2 phonetic segments [8]. The finding that L2 participants' accuracy rates remained similar across Blocks 1–3 is compatible with these studies under the assumption that unchanged accuracy indicates that the two phonetic segments were not activated separately by exposure. Importantly [8], which used the ABX task, reports that L2 speakers at least have an ability to notice that two similar phonetic segments are phonologically different even in the early stage of L2 acquisition but cannot encode them in a native-like manner. Given the above, it is conceivable that decreased response times observed across Blocks 1–3 result from either adaptation to the talker's accent, lingering sound patterns, or both, which increased the comprehensibility of the differences between the two phonetic segments, rather than from the improvement of perceptual ability to identify them correctly. This, in other words may indicate that implicit learning hardly occurs in L2 phonetic acquisition considering the implicit nature of adaptation. It is, however, possible that there was some effect on accuracy of phonetic perception, given that there was some tendency towards more correct responses in the latter blocks when target items were consonants, but it was not detectable, potentially due to insufficient input. Relatively low error rates across the blocks, which reduces the power to detect effects [9], may contribute to this undetectability. These factors could account for differences between null effects observed in the present study and the positive effects of training exposure on identification performance in previous studies [e.g. 6]. Another important finding is that the processing advantage observed between Blocks 1–3 disappeared in Block 4, where different sets of minimal pairs were presented. One account of this may be that such processing advantage by adaptation or lingering sound patterns appears only within learnt repeated words, as L2 speakers are potentially less

sensitive to abstract information than L1 speakers. For example, [18] reports that while participants showed priming effects, a phenomenon underlining adaptation, in their first language during the immediate repetition task, irrespective of whether the repeated word was spoken by the same speaker or not, they did not do so in their L2 when the repeated word was spoken in a different voice. As [18] suggests, this finding may indicate L2 speaker's reduced ability to generalise across two instances of spoken L1 words differing in perceptual properties. It is also possible that another set of minimal pairs increased processing costs, which counteracted the processing advantage, given that paying attention to new information can be cognitively costly, especially for L2 speakers. Alternatively, as discussed in the analysis on accuracy, reduced power, due to residuals of the statistical model not being normalised completely in this case, may have affected the result, although we tried to gain normalisation by removing response times that were incredibly long or too deviant from each participant's mean response times within each condition. Given that there is some numerical tendency towards decreased response times in the consonant condition, increasing statistical power may find learning effects across two different sets.

4. CONCLUSION

The present study examines whether Japanese language learners of English show learning effects by repeated exposure to phonetic contrasts that they have difficulty perceiving correctly during a two-alternative forced identification task. We observed learning effects, manifesting as significantly reduced response times over the course of the task. However, these reduced response times seemed to result from adaptation to the talker's accent or/and lingering sound patterns rather than from enhanced intelligibility of the two similar phonetic segments. Further, the observed processing advantage is restricted to the same set of minimal pairs. Overall, L2 speakers benefit from repeated exposure to similar phonetic segments, but it may not lead to global improvement in phonetic learning.

5. REFERENCES

- [1] Baayen, R. H., Davidson, D. J., Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *JML*, 59(4), 390–412.
- [2] Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B.,...Treiman, R. 2007. The English Lexicon Project. *BRM*, 39(3), 445–459.
- [3] Barr, D., Levy, R., Scheepers, C., Tily, H. 2013. Random-effects structure for confirmatory hypothesis testing: keep it maximal. *JML*, 68(3), 255–278.
- [4] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.*, 67(1), 1–48.
- [5] Best, C. T. 1995. A direct realist view of cross-language speech perception. In: W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.
- [6] Bradlow, A., Pisoni, D., Akahane-Yamada, R., 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(4), 2299–2310.
- [7] Clarke, C. M., Garrett, M. F. 2004. Rapid adaptation to foreign-accented English. *J. Acoust. Soc. Am.*, 116(6), 3647–3658.
- [8] Darcy, I., Daidone, D., Kojima, C. 2013. Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, 8(3), 372–420.
- [9] Dixon, P. 2008. Models of accuracy in repeated-measures designs. *JML*, 59(4), 447–456
- [10] Flege, J. E. 1992. The intelligibility of English vowels spoken by British and Dutch talkers. In: R. D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (pp. 157–232). Amsterdam: John Benjamins.
- [11] Gass, S., Varonis, E. 1984. The effect of familiarity on the comprehensibility of nonnative speech. *LL*, 34(1), 65–89.
- [12] 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.
- [13] Jaeger, T. F. 2008. Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *JML*, 59(4), 434–446.
- [14] Kuznetsova, A., Brockhoff, P., Christensen, R. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw.*, 82(13), 1–26. <https://doi.org/10.1111/j.1467-1770.1984.tb00996.x>
- [15] Munro, M. J., Derwing, T. M. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *LL*, 45(1), 73–97.
- [16] R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- [17] Shinohara, Y., Iverson, P. 2018. High variability identification and discrimination training for Japanese speakers learning English /r/-/l/. *JP*, 66, 242–251.
- [18] Trofimovich, P. 2005. Spoken-word processing in native and second languages: An investigation of auditory word priming. *AP*, 26(4), 479–504.
- [19] Xiea, X., Weatherholtzb, K., Baintonc, L., Rowed, E., Burchill, Z., Liu, L., Jaeger, T. F. 2018 Rapid adaptation to foreign-accented speech and its transfer to an unfamiliar talker. *J. Acoust. Soc. Am.*, 143, 2013–2031.