

Conflation of L2 vowel phonemes in production, perception, and imitation

Daiki Hashimoto¹, Takuma Sato¹, Hatsumi Ishii¹

¹*Joetsu University of Education*

Some languages have a larger number of vowel phonemes than others; for example, English has a denser vowel system than Japanese (Nishi et al., 2008). Traditionally, American English is analysed as having ten monophthongs, whereas Japanese has only five. This difference in vowel inventory size leads L2 learners to conflate certain vowel contrasts. For instance, Japanese learners of English often have difficulty distinguishing /æ/ and /ʌ/, and tend to realize both vowels as a single category similar to the Japanese vowel /a/ (Hashimoto et al., 2025a).

The relationship between L2 production and perception, as well as that between L2 imitation and perception, remains an open question (Jia et al. 2006; Llompарт & Reinisch 2019). The aim of this study is to examine the relationship between vowel perception and production, with a particular focus on vowel conflation, as well as the relationship between imitation and perception. Specifically, the study asks whether L2 learners who conflate an English vowel contrast in perception also exhibit vowel conflation in their production and imitation.

To address this question, we conducted an experiment consisting of production, perception, and imitation tasks. Thirty-nine Japanese learners of English participated. First, participants read aloud twenty words containing /æ/, twenty words containing /ʌ/, and additional forty English words. The recordings from this session served as baseline production data. Participants then completed a set of perceptual tasks. For the perception tasks, we created acoustically manipulated vowel stimuli whose formant values ranged from *bat* /bæt/ to *but* /bʌt/ in five steps. Participants listened to each stimulus and indicated which word they perceived. Those who consistently identified stimuli as either /æ/ or /ʌ/ were considered phonologically aware of the vowel contrast. Following this task, participants completed an AX discrimination task in which they listened to pairs of stimuli drawn from the manipulated continuum and judged whether the two sounds were the same or different. Sensitivity (d') scores were calculated for each participant. After completing the perceptual tasks, participants were explicitly instructed to imitate model speech stimuli produced by an American English speaker. The words used in the imitation task were identical to those used in the baseline production session. Orthographic information was not provided, as spelling is known to influence phonetic imitation (Hashimoto et al., 2025b). At the end of the experiment, participants listened to the auditory stimuli again and identified the words they heard. Tokens that were not correctly identified were excluded from the statistical analysis. The recordings from the baseline and shadowing sessions were manually annotated using Praat (Boersma & Weenink, 2020), and formant values were extracted using a Praat script.

Linear mixed-effects regression analyses were conducted on baseline and shadowed utterances. The results showed that participants who were aware of the vowel contrast produced /æ/ and /ʌ/ more distinctly than those who were unaware of the contrast (see Fig. 1). In contrast, no significant difference was observed between the two groups with respect to L2 phonetic imitation (see Fig. 2). The statistical models further indicated that sensitivity (d') scores significantly predicted imitation performance but were not associated with baseline production.

In sum, perceptual phonological awareness was closely associated with baseline production. This finding can be interpreted to mean that L2 perception and production are based on a shared phonological representation to some extent. By contrast, sensitivity (d') scores were closely associated with imitation, indicating that L2 imitation is closely linked with phonetic acuteness. This suggests that phonological representations may not be necessarily involved in L2 imitation. L2 imitation may reflect lower-level processing than baseline production.

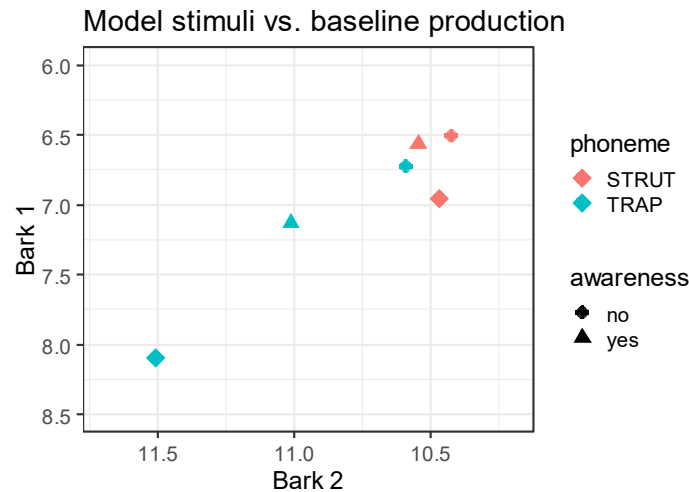


Fig. 1. Baseline production compared with model stimuli. Red indicates the STRUT vowel /ʌ/, and green indicates the TRAP vowel /æ/. Circles represent the mean values for participants who were aware of the vowel contrast, triangles represent the mean values for participants who were unaware of the vowel contrast, and diamonds represent the mean values of the model stimuli.

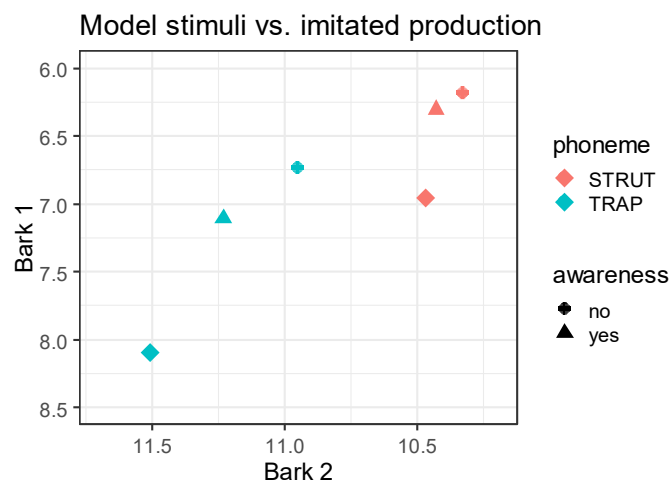


Fig. 2. Imitated production compared with model stimuli. Colours and shapes are identical to those in Fig1.

References

- [1] Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., & Trent-Brown, S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *The journal of the Acoustical Society of America*, 124(1), 576–588. <https://doi.org/10.1121/1.2931949>
- [2] Hashimoto, D., Ida, A., & Jozuka, H. (2025a). Word frequency effects on L2 learners' phonetic imitations. *JASA Express Letters*, 5(5): 055201. <https://doi.org/10.1121/10.0036512>
- [3] Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America*, 119(2), 1118–1130. <https://doi.org/10.1121/1.2151806>
- [4] Llompart, M., & Reinisch, E. (2019). Imitation in a second language relies on phonological categories but does not reflect the productive usage of difficult sound contrasts. *Language and Speech*, 62(3), 594–622. <https://doi.org/10.1177/0023830918803978>
- [5] Hashimoto, D., Tatsuya, K., & Asada, R. (2025b). Effects of orthography presentation and loanword frequency on L2 speech shadowing. *Amper*, 14: 100218. <https://doi.org/10.1016/j.amper.2025.100218>
- [6] Boersma, P. & Weenink, D. (2020). Praat: Doing phonetics by computer (version 6.1.14) <http://www.fon.hum.uva.nl/praat/>