

Avatar-mediated pitch gestures in L2 intonation perception

Michiko Toyama¹, Tomoko Hori², Mari Akatsuka³, Kaori Sugiura⁴

¹Bunkyo University, ²Juntendo University, ³Nagoya University of Foreign Studies, ⁴Ritsumeikan University

Recent research has increasingly highlighted the role of pitch gestures in supporting the perception of L2 prosodic patterns. Pitch gestures refer to iconic, dynamic hand movements that visually trace the intonation contour of speech, such as upward motions for rising intonation and downward for falling. A growing body of experimental research has shown that observing congruent pitch gestures facilitates the perception of pitch-based prosodic contrasts in L2 learning [1–4]. These effects are consistent with theories of multimodal speech perception, which suggest that gesture and prosody are processed in an integrated manner. Visual representations of pitch movement may thus support sensitivity to fine-grained prosodic detail. However, previous work has relied exclusively on human gesture models, leaving open the question of whether such perceptual benefits depend on the presence of a human demonstrator or whether they can also be achieved through non-human visual agents. Given the increasing use of animated agents in digital language learning environments, it is important to determine whether such visual models can provide comparable perceptual benefits.

This study examined the effects of observing pitch gestures on L2 intonation perception among Japanese undergraduate students. Participants were English majors with an estimated CEFR B1 proficiency level. A pre-test/post-test design assessed their ability to identify English intonation patterns before and after gesture-based training. During training, participants viewed videos demonstrating pitch gestures visually aligned with intonation contours. Two experimental conditions were implemented. In the human gesture condition ($N = 50$), participants viewed videos of a human model using pitch gestures congruent with spoken English stimuli. In the avatar gesture condition ($N = 67$), a separate group viewed an animated avatar performing the same gestures, aligned with identical audio recordings. Thus, auditory input was held constant, with only the visual modality differing. The training targeted three American English intonation patterns: rising, falling, and fall-rise. Participants' performance was assessed using a forced-choice identification task in which they judged whether each utterance conveyed one of the three target patterns. Learning outcomes were analyzed using paired- and independent-samples t -tests as well as ANCOVA controlling for pretest performance.

Results revealed significant improvements in intonation perception accuracy in both conditions. In the human gesture condition, mean accuracy increased from 64.42% to 70.82% ($t(49) = 3.18, p = .0025, d = 0.45$). In the avatar condition, accuracy increased from 68.64% to 76.21% ($t(66) = 6.92, p < .001, d = 0.85$). Between-group comparison of gain scores revealed no significant difference between conditions. An ANCOVA controlling for pretest performance further confirmed comparable posttest outcomes across the human and avatar conditions.

These findings demonstrate that pitch gestures support perceptual learning of English intonation in adult L2 learners, regardless of whether the gestures are produced by a human or an avatar. Both conditions showed significant pre-to-post improvements, indicating that visual representations of pitch movement facilitate sensitivity to intonational contrasts. Importantly, learning gains in the avatar condition were comparable to those in the human condition. This suggests that the effectiveness of pitch gestures in intonation perception does not depend on the presence of a human demonstrator, but on the availability of clear, temporally aligned visual cues mapping onto pitch movement. These results contribute to research on speech across the lifespan by highlighting the continued plasticity of prosodic perception in adulthood and suggest that avatar-mediated pitch gestures offer a scalable, technology-based approach to supporting L2 prosody perception.

References

- [1] Morett, L. M., & Chang, L.-Y. (2015). Emphasizing sound and meaning: Pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, 30(3), 347–353. <https://doi.org/10.1080/23273798.2014.923105>
- [2] Baills, F., Suárez-González, N., González-Fuente, S., & Prieto, P. (2019). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition*, 41(1), 33–58. <https://psycnet.apa.org/doi/10.1017/S0272263118000074>
- [3] Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, 41(1), 5–32. <https://psycnet.apa.org/doi/10.1017/S0272263117000316>
- [4] Hori, T., Akatsuka, M., & Toyama, M. (2025). Effects of observing pitch gestures on the perception of English intonation by Japanese learners of English. *Journal of Second Language Pronunciation* 11(1), 26-45. <https://doi.org/10.1075/jslp.24015.hor>