

Early Acoustic Evidence for Mandarin /j/: Temporal Dynamics in Native and L2 Speech

Yuyan Niu¹, Albert Lee¹

¹ The Education University of Hong Kong

Listeners often identify speech categories before a syllable is complete, relying on dynamic acoustic information rather than discrete segmental boundaries. A growing body of speech perception research shows that early portions of the signal—especially rapid formant transitions—play a critical role in evidence accumulation and category commitment. However, for Mandarin medial glides such as /j/, it remains unclear how early reliable acoustic evidence becomes available, how stable this evidence is across syllable structures, and how it differs between native speakers and second-language (L2) learners.

Traditionally, Mandarin /j/ has been treated as a segment whose onset and offset can be identified acoustically (Xu, 1997). Yet in complex syllables (e.g., *miao*, *mian*), the second formant (F2)—the primary acoustic correlate of palatality—often shows a continuous transition from the consonant into the vowel, with no clear plateau supporting discrete boundary placement. From a dynamic perspective, /j/ may instead be understood as an early-timed tongue-body gesture whose perceptual contribution depends on when and how strongly it shapes the evolving F2 trajectory.

The present study examines the temporal availability of acoustic evidence for Mandarin /j/ using high-density, time-normalized F2 trajectories, with particular focus on the early portion of the syllable. We address three questions: (1) How early in the syllable does robust acoustic evidence for /j/ emerge in native Mandarin speech? (2) Is this early evidence stable across different syllable structures? (3) Do L2 learners differ not only in the magnitude of /j/-related cues, but in the timing of their availability?

A total of 45 speakers participated in the production experiment, divided into three groups: 15 native Mandarin speakers (L1), 15 intermediate Japanese learners of Mandarin (JP_INT), and 15 beginner Japanese learners (JP_BEG). All participants reported normal hearing and had no history of speech or language disorders. They were recorded in three speech rate conditions (slow, normal, fast) across four syllable structures (CGVX, CGV, CGVN, GVX). Rather than segmenting C–G–V boundaries, we treated the entire target syllable as a continuous unit. F2 trajectories were extracted, time-normalized, and sampled at high temporal resolution. To capture early acoustic evidence relevant for perceptual decision-making, we derived a scalar measure: the early F2 slope over the first 30% of the normalized syllable duration, an interval motivated by evidence-accumulation and gating studies.

Results revealed clear temporal differences across speaker groups. At normal speech rate, L1 speakers exhibited strong early F2 rises across all syllable structures, indicating rapid availability of /j/-related acoustic evidence. Intermediate learners showed smaller, but still positive early slopes, while beginner learners exhibited near-flat or negative slopes, suggesting limited early acoustic evidence for /j/. Crucially, syllable structure did not show a significant main effect on early slope, indicating that early acoustic evidence for /j/ is structurally robust in native speech but systematically attenuated or delayed in L2 production.

These findings suggest that native listeners can accumulate evidence for /j/ rapidly, potentially committing to a /j/-bearing interpretation well before the syllable midpoint. In contrast, L2 learners—especially beginners—may force listeners to rely on later portions of the signal, delaying category stabilization. By linking fine-grained production dynamics to the temporal availability of perceptual cues, this study provides a production-based foundation for understanding time-based speech perception in both native and non-native contexts.

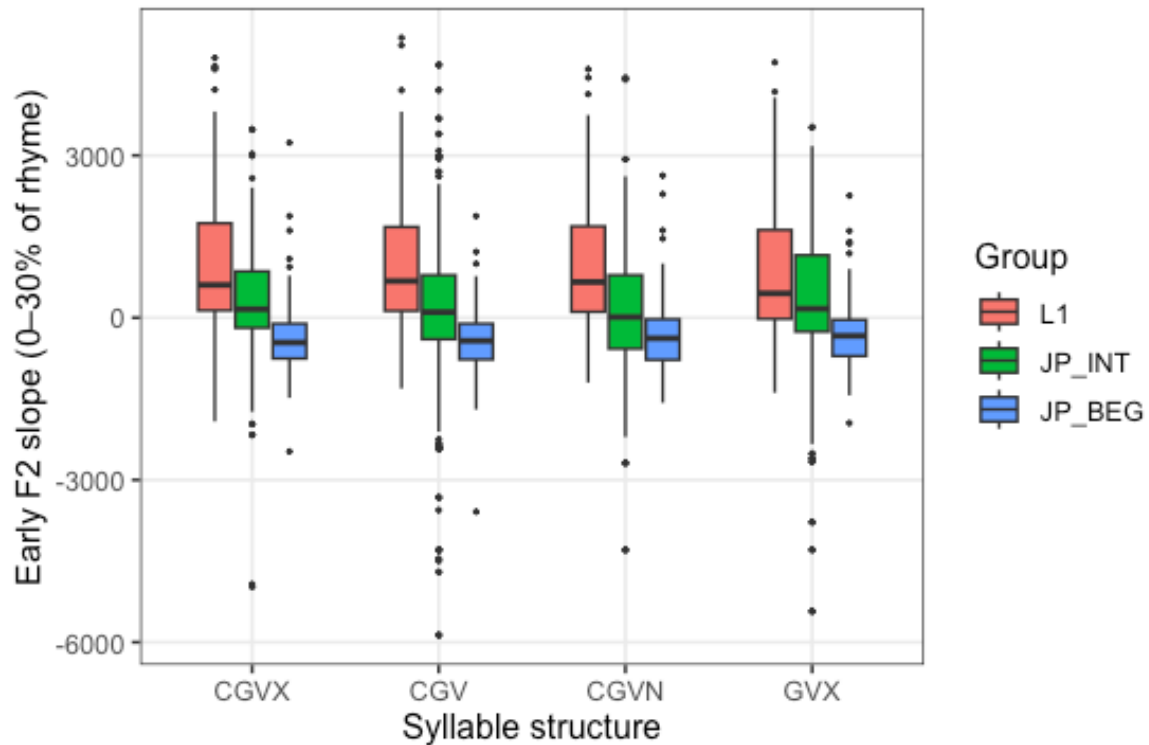


Fig. 1. Early F2 slope (0–30% of the normalized syllable) by speaker group (L1, JP_INT, JP_BEG) and syllable structure (CGVX, CGV, CGVN, GVX) at normal speech rate. Positive slopes indicate a rapid F2 rise, providing strong early acoustic evidence for /j/.

References

- [1] Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins.
- [2] Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155–180.
- [3] Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28, 267–283.
- [4] McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? *Psychological Review*, 118, 219–246.
- [5] Öhman, S. E. G. (1966). Coarticulation in VCV utterances. *Journal of the Acoustical Society of America*, 39, 151–168.
- [6] Xu, Y. (1997). Contextual tonal variation in Mandarin. *Journal of Phonetics*, 25, 61–83.