

L1-Dependent Spectro-temporal Dynamics of the English Diphthong [aɪ]: Evidence from Generalized Additive Mixed Models

Yunah Chung¹, Sejin Oh¹

¹ Jeju National University

A diphthong is a vowel that involves a dynamic change in vowel quality within a syllable, such as *bite* [bait] in English. Early acoustic work on English diphthongs suggests that temporal expansion may not shape all aspects of diphthong production uniformly: some components such as onset and F2 transition rate remain stable across different speech rates while offset target positions can be more variable (Gay, 1968). Evidence from Chinese [aɪ] further indicates that longer (slower) tokens may exhibit higher F1 onset and higher F2 offset, and that F1 and F2 transition rates decrease as vowel duration increases, suggesting that duration shapes overall trajectory dynamics (Xiang & Sun, 2025). Building on this literature, the present study examines how L1 background and length of U.S. residency relate to the dynamics of the English diphthong [aɪ] produced by English, Korean, and Chinese speakers, and tests whether vowel duration conditions diphthong dynamics differently across these groups.

Using the Wildcat Corpus, we analyzed productions of the target word *bite* from 12 Korean, 12 Chinese, and 8 English speakers. Overall vowel duration and formant trajectories (F1, F2) were analyzed, with formant values sampled at ten equidistant points across the vowel interval and normalized in time. Vowel duration was modeled using linear mixed-effects regression with random intercepts for speaker. To capture nonlinear trajectory differences, generalized additive mixed models (GAMMs) were fit to F1 and F2 trajectories, including smooth terms over time, language- or group-specific smooths, gender as a parametric covariate, and speaker-level random effects. For non-native speakers, additional GAMMs examined effects of length of residence (LoS). In addition, to investigate how temporal expansion affect [aɪ], additional GAMMs examined effects of duration groups, where each token was classified into short (D1: 100–200 ms), mid (D2: 200–300 ms), and long (D3: >300 ms) duration groups.

Linear mixed-effects modeling revealed no significant effects of L1 or length of residence on the overall duration of the English diphthong [aɪ], with the non-significant 20–30 ms lengthening observed for Chinese and Korean speakers. In contrast, GAMM analyses revealed robust language-specific differences in formant dynamics: Chinese speakers exhibited higher overall F1 and lower overall F2 values than English speakers, while Korean and English speakers did not differ in overall formant values. Crucially, as shown in Fig. 1, English native speakers initiated both F1 and F2 transitions earlier in time, whereas Korean and Chinese speakers maintained a more prolonged [a] target followed by a steeper transition toward [ɪ], a pattern especially pronounced in the F1 trajectory. GAMM analyses further showed that length of residence (LoS) systematically shaped F1 trajectories for both Korean and Chinese speakers. For Chinese speakers, overall F1 remained higher than natives across all LoS groups, but longer residence was associated with progressively more native-like temporal coordination of the diphthong transition. For Korean speakers, short-term residents differed from natives in overall F1, while long-term residents exhibited increasingly nonlinear, native-like temporal patterns.

Across the three language groups, duration grouping (D1–D3) modulated the temporal organization of F1 trajectories in language-specific ways, though no reliable differences across duration groups were observed within any single language. This pattern aligns with Gay (1968), suggesting that diphthongal transition structure can remain stable despite increases in overall duration.

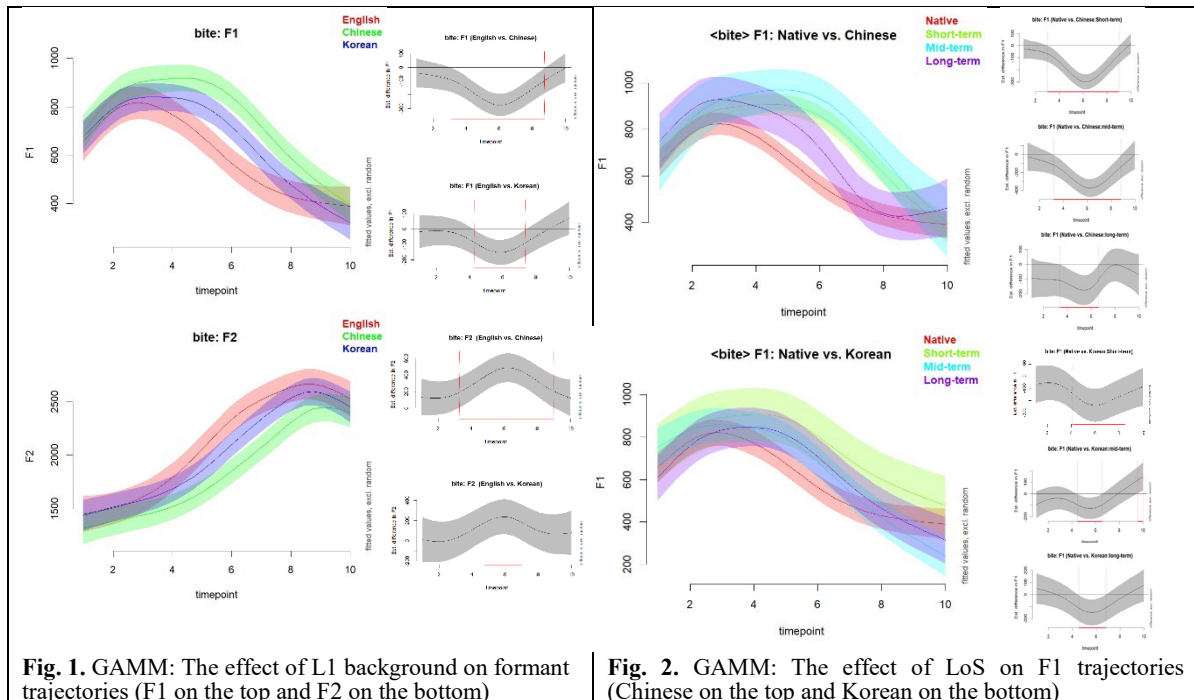


Fig. 1. GAMM: The effect of L1 background on formant trajectories (F1 on the top and F2 on the bottom)

Fig. 2. GAMM: The effect of LoS on F1 trajectories (Chinese on the top and Korean on the bottom)

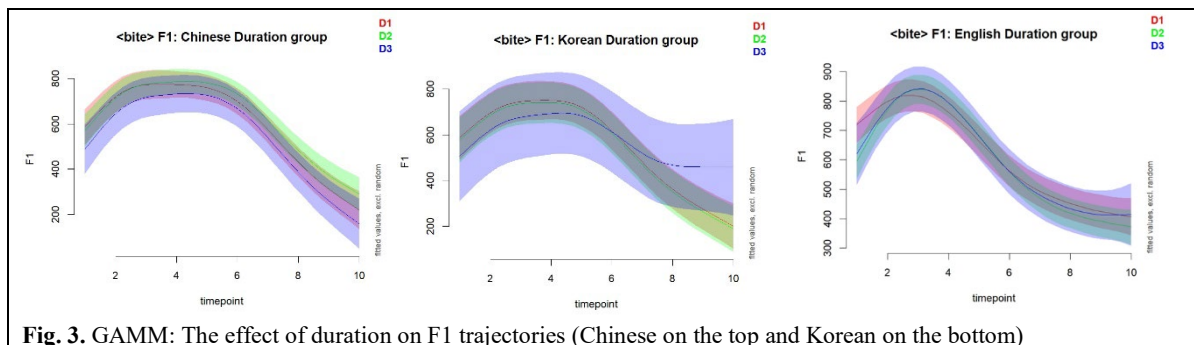


Fig. 3. GAMM: The effect of duration on F1 trajectories (Chinese on the top and Korean on the bottom)

References

- Gay, T. (1968). Effect of speaking rate on diphthong formant movements. *The Journal of the Acoustical Society of America*, 44(6), 1570-1573.
- Tasko, S. M., & Greilick, K. (2010). Acoustic and articulatory features of diphthong production: A speech clarity study. *Journal of Speech, Language, and Hearing Research*, 53(1), 84-99.
- Tjaden, K., & Weismer, G. (1998). Speaking-rate-induced variability in F2 trajectories. *Journal of Speech, Language, and Hearing Research*, 41, 976-989.
- Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, 53(4), 510-540. <https://doi.org/10.1177/0023830910372495>
- Weismer, G., & Berry, J. J. (2003). Effects of speaking rate on second formant trajectories of selected vocalic nuclei. *The Journal of the Acoustical Society of America*, 113, 3362-3378.
- Xiang, Y., & Sun, J. (2025, August). Modeling Formant Dynamics in Mandarin/ai: Effects of Speech Style and Speech Rate. In *Interspeech 2025* (pp. Paper-ID). International Speech Communication Association.