Prosodic fluency measures for automated AI-based assessment of L2 Mandarin speech

Fluency generally refers to the continuous and smooth production of language as speakers articulate words and sentences in oral speech. A large body of research has demonstrated a strong relationship between measurable prosodic features constituting utterance fluency and overall L2 proficiency. In particular, temporal fluency features (e.g., speech rate, pausing, hesitation, and repair), as well as intonational and rhythmic features (e.g., stress, tone, focus, and sentential intonation), contribute significantly to the perceived non-nativeness of L2 speech (Wennerstrom 1994; Bradlow et al. 1996; Mennen 2007; Guion et al. 2000; Trofimovich and Baker 2006; Kang et al. 2017; Kang 2022). L2 speakers tend to produce utterances that are longer in duration, with more frequent pauses, lower pitch levels, and narrower pitch ranges (Bradlow et al. 1996; Aoyama 2007; Jilka 2007; Mennen 2007; Tavakoli et al. 2020). Despite extensive research on L2 English across diverse L1 backgrounds, prosodic fluency features across different proficiency levels in L2 Mandarin remain underexplored, in contrast to the substantial literature on lexical tone production in L2 Mandarin speech.

The present study examines acoustic measures associated with speaking fluency in Korean speakers of L2 Mandarin and the relative predictive weights of these fluency features in predicting oral proficiency. Acoustic analyses are conducted along both temporal and intonational dimensions, including pronunciation rate (including pauses), articulation rate (excluding pauses), pause count, and pause duration, as well as mean pitch, pitch range, and characteristics of sentential declination. As a seminal attempt to investigate fluency in Korean speakers' L2 Mandarin development, this study analyzes fluency measures at beginner and advanced proficiency levels using read speech drawn from a large-scale multilingual L2 speech corpus encompassing seven languages – English, Japanese, Mandarin, French, German, Spanish, and Russian – developed to support the automatic assessment of Korean speakers' L2 speech. The L2 Mandarin dataset alone consists of approximately 500 hours of read and spontaneous speech across different proficiency levels and is available through the AI Hub website (https://www.aihub.or.kr), an open platform established and supported by the National Information Society Agency (NIA) of Korea to provide the infrastructure necessary for the development of AI technologies, products, and services (Han et al. 2024; Lee and Park 2025). Results reveal that articulation rate, the number of non-comma pauses (i.e., pauses not corresponding to orthographic commas), and the degree of pitch declination across an utterance are most closely related to perceived fluency, with articulation rate emerging as the strongest predictor of perceived L2 Mandarin fluency. While this study focuses on Korean speakers' L2 Mandarin, the findings not only provide broader implications for understanding L2 Mandarin prosodic development but also for identifying fluency features that should be prioritized in the design of automated speaking proficiency assessments for speakers of diverse L1 backgrounds.

Selected References

Aoyama, K. and S. G. Guion. 2007. Prosody in second language acquisition. In Bohn and Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*, 281-297. Philadelphia: John Benjamins Publishing Company.

Han, S., M. Chung, and S. Kim. 2024. Constructing Korean learners' L2 speech of seven languages for automatic pronunciation assessment. *LREC-COLING 2024*: 3772–3781.

Jilka, M. 2007. Different manifestations and perceptions of foreign accent in intonation. In Trouvain and Ulrike (eds.), *Non-native prosody: Phonetic description and teaching practice*, 77-96. Berlin: De Gruyter Mouton.

Lee, O. J. and J. Park. 2025. A critical evaluation of AI Hub Korean learners' Mandarin speech corpus: structures and applications. *Journal of Chinese Literature* 122: 309-333.

Mennen. I. 2007. Phonological and phonetic influences in non-native intonation. in Trouvain and Ulrike (eds.), *Non-native prosody: Phonetic description and teaching practice*, 53-76. Berlin: De Gruyter Mouton.

Trofimovich, Pavel and Wendy Baker. 2006. Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 23(1): 1-30.