Linguistic Research 43(1): 295-328
DOI: 10.17250/khisli.43.1.202603.011

Prosodic fluency measures in Korean learners' Mandarin*

Yan Xiong ** **· Jieun Park · Ok Joo Lee** ***
(Seoul National University)

Xiong, Yan, Jieun Park, and Ok Joo Lee. 2026. Prosodic fluency measures in Korean learners' Mandarin. *Linguistic Research* 43(1): 295-328. Prosodic features of second language (L2) speech, encompassing temporal and intonational dimensions, are closely linked to listeners' perceived fluency. In stark contrast to extensive literature on lexical tone production, the prosodic fluency of Korean learners of Mandarin across different proficiency levels remains underexplored. This study aims to fill this gap by analyzing temporal and intonational correlates of prosodic fluency in Korean learners of Mandarin, using sentence-reading data drawn from a large-scale L2 speech corpus developed for AI training that provides balanced proficiency representation across beginner- and advanced-level learners. Results show 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. These findings not only advance our understanding of the development of L2 Mandarin prosody but also shed light on the fluency features that should be prioritized in the design of automated speaking proficiency assessments for Korean learners. **(Seoul National University)**

Keywords Korean learners, Mandarin, fluency, pronunciation rate, articulation rate, pause, sentential declination

---

* This work was supported by the Hyunsong Educational & Cultural Foundation Academic Research Grant for Humanities at Seoul National University and by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2025S1A5A2A01008329). We thank the reviewers for their constructive comments and suggestions on our manuscript.

** First Author
*** Corresponding Author

1. Introduction

Fluency refers to the production of language in a fluid and smooth manner as speakers produce words and sentences in oral speech. In second language acquisition (L2, hereafter), the term is often used in a broad sense, synonymously with overall language proficiency. In a narrower sense, fluency, along with accuracy and complexity, constitutes three main aspects of proficiency (Baker-Smemoe et al. 2014; Yan et al. 2021, references therein). Further recognizing the multidimensional nature of fluency, Segalowitz (2010) proposes distinctions between cognitive fluency, utterance fluency, and perceived fluency. Cognitive fluency concerns a speaker's capacity to efficiently coordinate the cognitive processes involved in speech production, while utterance fluency denotes the observable features revealing the ease and smoothness of speech. Perceived fluency pertains to listeners' assessments drawn from speakers' utterances (Segalowitz 2010: 48–49).[1]

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 such as speech rate, pausing, hesitation, and repair, as well as intonational and rhythmic features including stress, tone, focus, and sentential intonation, contribute significantly to the perceived non-nativeness of L2 speech (Wennerstrom 1994; Bradlow et al. 1996; Guion et al. 2000; Trofimovich and Baker 2006; Mennen 2007; Kang et al. 2017; Kang 2022). In general, L2 speakers tend to produce utterances that are longer in duration, with more frequent pauses, lower pitch levels,[2] and narrower pitch ranges (Bradlow et al. 1996; Aoyama and Guion 2007; Jilka 2007; Mennen 2007; Tavakoli et al. 2020).

Although extensive research has examined L2 English of speakers with diverse L1 backgrounds, the relative contribution of fluency features to Korean learners' oral proficiency in L2 Mandarin Chinese (Mandarin, hereafter) remains underexplored.

1 While carefully distinguishing among the terms "proficiency", "fluency", and "accuracy", the present study adopts the terminology and assessment framework of Han et al. (2024), from which the assessment criteria and speech data scores used in this study are derived.

2 The terms "pitch level" and "pitch range" have been defined and measured differently across previous studies. Although these discrepancies are not discussed in detail due to space limitations, the present study uses the terms "mean pitch" and "pitch range" with explicit definitions to maintain conceptual and methodological clarity (see Section 3.2). While mean pitch provides an ideal measure of pitch level or overall pitch height, it should be noted that this measure has been implemented differently in the literature.

Investigating the fluency patterns of Korean learners in Mandarin is, however, essential for both theoretical insight and applied purposes. First, it helps expand our understanding of L2 development when salient prosodic differences exist between L1 and L2. As is well known, Mandarin uses pitch to mark lexical tones, whereas in Korean it serves post-lexical intonational functions. Pausing, an important cue for prosodic phrasing, is often facilitated by word-based spacing in the Korean orthography, *Hangul*, but such spacing is absent in the Mandarin writing system, *Hanzi*. Second, this line of inquiry is also crucial for developing speech assessments for Korean learners of Mandarin. Although Korean learners constitute more than half of the overseas takers of the HSK (*Hanyu Shuiping Kaoshi*), the official Mandarin exam for non-native speakers, and represent the largest group of Mandarin learners in China (Kang 2017), research on their speaking fluency has been largely limited to segmental pronunciation and lexical tone accuracy (Han 2017; Zhang, H. 2018; Oh 2021; Fu and Lee 2022; Lee and Park 2025, among others). Thus, understanding the temporal and intonational features of Korean learners' Mandarin and their relationship to perceived fluency is essential for developing valid oral proficiency assessments.

To address this gap, the present study examines acoustic measures associated with speaking fluency in Korean learners of Mandarin and identifies the key predictors of fluency level. More specifically, acoustic analyses are conducted on both temporal and intonational dimensions. The temporal measures include pronunciation rate (including pauses), articulation rate (excluding pauses), pause count, and pause duration, while the intonational measures include mean pitch, pitch range, and characteristics of sentential declination. As a pioneering attempt to investigate fluency development in Korean learners of Mandarin, this study analyzes these measures at beginner and advanced levels using read speech drawn from a large-scale multilingual L2 speech corpus (AI Hub, see Section 3). It also evaluates the relationships among fluency measures and their relative importance for L2 fluency perception, thereby providing important insights for the development of automated speaking assessments for Korean learners of Mandarin. While this study focuses on sentence-reading speech, the findings provide broader insights into fluency assessment in spontaneous L2 Mandarin speech.

2. Background and research aims

2.1 Temporal fluency features

Extensive research has shown that L2 speech is characterized by longer utterance durations, whether measured by pronunciation rate (including pauses) or articulation rate (excluding pauses), as well as by more frequent and longer pauses, silent or filled. This is sometimes accompanied by repetition, false starts, reformulations, and repairs, which affect listeners' perception of L2 fluency. This pattern has been observed in L2 speech produced by speakers from diverse L1 backgrounds (Munro and Derwing 1995a, 1995b; Guion et al. 2000; Trofimovich and Baker 2006; Aoyama and Guion 2007; Mennen 2007; Baker-Smemoe et al. 2014; Tavakoli et al. 2020; Yan et al. 2021; Lee 2024, and references therein).

Empirical evidence from cross-linguistic studies further supports this pattern. For instance, a comparison of adult native speakers of Japanese and English in Aoyama and Guion (2007) found that syllable and utterance durations were longer, while the pitch range in content words tended to be larger for L2 English speakers. Yan et al. (2021) reported that pronunciation rate, articulation rate, and number of silent pauses, as well as mean length of runs, juncture pause rate, and repair success rate, were correlated with the speaking fluency scores of L2 English produced by speakers from 15 different L1 backgrounds. Similarly, Baker-Smemoe et al. (2014) showed that syllables per second, pause frequency and duration, run frequency and duration, hesitations, and false starts were associated with fluency in L2 French, German, Japanese, Arabic, and Russian spoken by L1 English speakers.

The Mandarin speech of Korean learners generally conforms to this pattern, as their speech tends to be slower and contains more frequent and longer pauses (Yoon 2013; Lee 2016; Chen and Zhou 2017; Kong and Kim 2018; Zhang, Q. 2018; Lee, Y. 2019, 2020; Bao 2020; Wang, G. 2020; Cui 2021). Despite this general trend, however, a close look at previous studies reveals inconsistencies. For instance, some studies have observed that articulation rate varies with L2 Mandarin level (Chen and Zhou 2016; Zhang, Q. 2018; Lee, Y. 2019; Cui 2021), whereas others report no significant differences (Kong and Kim 2018). Likewise, mean pause length has been shown to differ by fluency level in some studies (Zhu 2009; Lee, Y. 2019), but has not been found to correlate with the length of study in others (Zhang, Q. 2018; Cui

2021). Another challenge in analyzing pauses lies in the absence of a clear definition of what constitutes a perceptual pause, as well as the lack of distinction between semantically appropriate pauses and unnecessary pauses caused by disfluency. Given that appropriate pause placement, such as at clause boundaries or after groups of words forming a meaningful unit, can enhance the perception of L2 fluency (Thornbury 2005: 7), it is essential to distinguish between different types of pauses.

2.2 Intonational fluency features

L2 speech often diverges from L1 speech in intonational features, although the sources of this difference remain a matter of debate. One line of research suggests that languages differ in pitch level and range. For instance, monolingual English speakers have a higher pitch level and a wider pitch range than monolingual German speakers (Mennen et al. 2012), and this difference has been cited as a factor underlying native German speakers' narrower pitch range in their L2 English speech (Jilka 2007). Pitch differences have also been observed across varieties of the same language, with speakers of Beijing Mandarin exhibiting higher pitch levels and wider pitch ranges than speakers of Taiwan Mandarin and Min Chinese (Chen 2005; Torgerson 2005; Huang and Fon 2011). Several studies involving bilingual speakers have also provided evidence for cross-linguistic differences in pitch. For example, Japanese has a higher pitch level and a wider pitch range than English (Graham 2014). Russian has a higher pitch level than English, which has been reported to show no difference from Cantonese (Altenberg and Ferrand 2006), while Ng et al. (2010, 2012) found English to have a higher pitch level than Cantonese among female speakers.

With limited comparative research on Mandarin and Korean, some studies have reported that these languages differ from English in pitch, although the detailed findings are somewhat varied. Monolingual Mandarin speakers have been shown to exhibit a higher pitch level than monolingual English speakers, with no difference in pitch range (Eady 1982). Mandarin has also been found to exhibit both a higher pitch level and a wider pitch range than English in single-word utterances, although only a higher pitch level was observed in prose passages (Keating and Kuo 2012). Several studies on bilingual speakers have further reported that Mandarin and Korean show a higher pitch level and/or pitch range than English during reading tasks (Xue et al. 2002; Lee and Sidtis 2017), while gender- and age-related variation has also been noted (Xue et al. 2002; Cheng 2020). However, other studies have reported that pitch level and pitch range in Mandarin do not differ significantly from those in Korean or English (Lee, O. J. 2019; Zhu et al. 2022). Methodological differences in speaker proficiency levels, dataset characteristics, speech elicitation paradigms, and pitch level measurements may largely account for these mixed results in the literature.

Another line of research on intonational differences between L1 and L2 speech concerns the effect of a speaker's L2 proficiency. That is, L2 speech is often characterized by overall lower pitch levels and/or narrower pitch ranges, along with incorrect placement of pitch prominence (Bradlow et al. 1996; Mennen 2007). These features become more native-like as L2 proficiency advances (Jenner 1976; Backman 1979; Willems 1982; Mennen 2007; Zhang, H. 2018; Kang 2022). The pitch range has also been found to significantly affect intelligibility in L2 English (Bradlow et al. 1996). In fact, the prosody of L2 Mandarin, particularly for lower-level learners of non-tonal languages such as English and Korean, has long been criticized for exhibiting a lower pitch level and narrower pitch range, even though pitch in Mandarin, as a tone language, is expected to fluctuate to a greater extent than in non-tonal languages (Chen 1974; Zhang, H. 2018, and references therein).

However, although tone errors have often been attributed to L2 learners' lower pitch level and range, little attention has been paid to how these factors influence the perception of L2 Mandarin fluency. In addition to pitch level and range, another important intonational characteristic is sentential declination, defined as a gradual downward pitch trend over the course of an utterance, particularly evident in declarative sentences. While it has been attested in many languages (Ladd 1996/2008; Gussenhoven 2004), it plays a crucial role in signaling sentence types and discourse structures in Mandarin. Global declination is particularly important since Mandarin considerably restricts local pitch variations, such as phrase boundary tones, arguably due to the presence of lexical tones (Gårding 1987; Shen 1990; Tao 1996; Lee 2005; Peng et al. 2005; Wang et al. 2013; Yuan and Liberman 2014). Despite its importance, the relationship between declination characteristics and L2 Mandarin fluency remains underexplored.

2.3 Research aims

Although previous studies have advanced our understanding of how L1 background and L2 proficiency level influence L2 fluency, they exhibit several notable limitations in examining the fluency features of Korean learners of Mandarin across different stages of development. First, few studies have employed comprehensive measures encompassing both temporal and intonational features, thereby limiting the ability to evaluate the relative contribution of these measures to Korean learners' Mandarin fluency. While some attention has been devoted to identifying the temporal characteristics of L2 Mandarin, intonational properties beyond those associated with lexical tones have been largely overlooked. Second, learners' speaking proficiency levels have rarely been adequately accounted for. Proficiency has been classified in various ways, such as by HSK scores, length of study, academic year, coursework completed, or evaluations by native and non-native listeners, yet these classifications rarely reflect actual fluency in speech. Third, most studies have employed a variety of measures and tasks (e.g., sentence reading, narration, interviews, picture descriptions), making it difficult to determine whether the conflicting results are attributable to methodological differences. The seemingly contradictory findings may also reflect variation in data size and rater type. Examples include picture description (Kong and Kim 2018; Zhang, Q. 2018; Lee, Y. 2019, 2020), oral interview recordings (Zhu 2009), dialogues with an investigator (Cui 2021), speaking and writing tasks (Wang, G. 2020), and rephrasing and storytelling (Lee 2020). As a result, existing findings do not provide a clear picture of the relationship between Korean learners' prosodic characteristics and the perceived fluency of their L2 speech.

To address these issues, the present study investigates both the temporal and intonational features of Korean learners' Mandarin speech. The data are drawn from an open-source corpus containing over 500 hours of L2 Mandarin produced by Korean learners of differing proficiency levels. The focus is on Mandarin speech at beginner and advanced proficiency levels, elicited through a sentence-reading task. Prosodic fluency scores, a set of fluency assessments made by expert raters on a 1-to-5 scale that achieved high inter-rater agreement (0.73; Han et al. 2024), are employed to evaluate the fluency of each utterance. The specific research questions guiding this study include: (1) What temporal and intonational features are strongly associated with the fluency of Korean learners' Mandarin read speech? (2) Which fluency measures contribute most to the perception of fluency, and which best predict differences across fluency levels?

3. Methodology

3.1 Data

The speech data and metadata analyzed in the present study were drawn from an L2 Mandarin corpus of Korean learners, the final version of which was released in November 2023 on the AI Hub website, 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 (https://www.aihub.or.kr; see Han et al. 2024; Lee and Park 2025 for a detailed discussion of the corpus). This dataset is part of a multilingual L2 speech corpus encompassing seven languages, English, Japanese, Mandarin, French, German, Spanish, and Russian, developed to support the automatic speech assessment of Korean learners' L2 speech.

This study draws on this corpus for three primary reasons. First, to the best of the authors' knowledge, it is the largest L2 Mandarin speech corpus produced by Korean learners of varying proficiency levels, both within and outside Korea. Second, it is not only annotated with both *Pinyin* and Simplified Chinese characters, but also contains pronunciation transcriptions and pronunciation error labels. Third, and most importantly, it provides experts' pronunciation assessment scores, including pronunciation accuracy and prosodic fluency. The corpus therefore enables effective categorization of learners' proficiency levels, allowing for the examination of the impact of L2 development on fluency, while metadata on speakers' gender, age, self-reported Mandarin proficiency, and pronunciation error types offer additional insights into learner-related factors. Given the challenges in defining L2 speaking proficiency, which are not well captured by length of study or HSK and C.TEST (*Shiyong Hanyu Shuiping Rending Kaoshi*, Test of Practical Chinese) scores used in previous studies (Zhu 2009; Yoon 2013; Chen and Zhou 2016; Lee 2016; Kong and Kim 2018; Zhang, Q. 2018; Lee, Y. 2019, 2020; Wang, G. 2020; Cui 2021), the present study addresses this methodological limitation by distinguishing between beginner and advanced levels of

L2 Mandarin based on prosodic fluency scores provided in the corpus metadata.

The L2 Mandarin speech component alone accounts for more than 500 of the over 3,000 hours of multilingual L2 speech in the corpus, categorized into two types: "pronunciation assessment data," which consist of recordings of learners reading a provided script aloud through sentence- and paragraph-reading tasks, and "speaking assessment data," which include recordings from various spontaneous speaking tasks, such as answering questions, picture description, and chart explanation. It was collected from 872 Korean learners, whose proficiency levels were classified as beginner, intermediate, and advanced, labeled "low," "mid," and "high," respectively (1,226 when counting participants in both data types separately). As a pioneering attempt to examine the acoustic measures that best correlate with L2 Mandarin fluency levels, the present study selected data from "pronunciation assessment data", thereby minimizing potential confounding influences of segmental and tonal errors in pronunciation. To acquire balanced speech data from beginner- and advanced-level learners, two subsets containing equal numbers of learners at each level were selected from "pronunciation assessment data" (Set 01 and Set 04). Each subset contains single-sentence readings numbered #1 through #100 in the corpus, with sentence lengths ranging from 5 to 20 syllables (see Appendix for sentence list).[3]

Given that the proficiency levels in the metadata were self-reported, fourteen learners with mean prosodic fluency scores in the 1–2 range (≥1.0, <3.0) were selected as beginners, and another fourteen with scores in the 4–5 range (≥4.0, ≤5.0) were selected as advanced, ensuring adequate representation of the two proficiency levels.[4] As the intermediate-level group exhibited a relatively wide range of prosodic fluency scores, which compromised the validity of learner homogeneity, the present study included only the beginner- and advanced-level groups. In addition, because the corpus contained more female learners than male learners, it was not possible to balance the data for both fluency level and gender. Therefore, this study included only female speech in order to more accurately isolate the effects of fluency. After excluding a few recordings with poor sound quality, a total of 2,497 sentence readings were

3 One sentence (Sentence #58 of Set 01), which contained an English letter, was excluded from the analysis because of potential variability in its pronunciation.

4 The sample standard deviation (SD) across utterances for each speaker in the beginner- and advanced-level groups ranged mostly from 0.36 to 0.72, indicating that within-subject variability was relatively limited.

acquired: 1,246 from beginner-level learners and 1,251 from advanced-level learners. To provide a clear overview of the prosodic fluency scale, the holistic scoring rubrics used in the corpus are presented in Table 1 (Han et al. 2024: 4).

Table 1. Scoring rubrics for prosodic fluency in the AI Hub L2 corpus

score	prosodic fluency
5	Natural stress, rhythm and intonation. The speaking rate is moderate, and the number and duration of pauses are natural. There are few speech mistakes, and the pauses are appropriately used to separate units of speech.
4	Slightly awkward stress, rhythm and intonation. The speaking rate is mostly consistent, with some hesitations and breaks. The pauses are appropriately used to separate units of speech, but their number and duration are slightly awkward.
3	Somewhat awkward stress, rhythm and intonation. The speaking rate is inconsistent and a bit slow, with frequent breaks. The pauses are not appropriately used to separate units of speech.
2	Considerably awkward stress, rhythm and intonation. The speaking rate is slow, with many breaks. The pauses last long and do not appropriately separate units of speech.
1	Terrible stress, rhythm and intonation. The speaking rate is too slow, with too many breaks. The pauses last too long and do not serve to separate units of speech at all.

3.2 Measurement and analysis

The temporal and intonational features were measured for each sentence recording. For temporal features, pronunciation rate (including pauses), articulation rate (excluding pauses), pause count, and pause duration were measured, while for intonational features, mean pitch, pitch range, and sentential pitch declination were analyzed. Pause durations were measured in seconds, and pitch values were measured in Hertz and subsequently normalized to semitones to account for inter-speaker pitch variation. For measurement, audio was automatically aligned with *Pinyin* transcriptions and Simplified Chinese characters using the Montreal Forced Aligner (MFA; McAuliffe et al. 2017), generating “words” and “phones” tiers in a TextGrid. A Praat script added a “silences” tier with comma labels, which was later merged with the “words” tier using the praatIO library in Python. A "pitch tier" labeling the rime portion of each syllable was created, from which pitch values were measured. This sentence-level labeling is illustrated in Figure 1. The statistical analyses were performed in Python (version 3.10.16; Python Software Foundation 2024) and the R environment (version 4.3.3; R Core Team 2024) using the lmerTest (version 3.1.3; Kuznetsova et al. 2017) and stats packages.

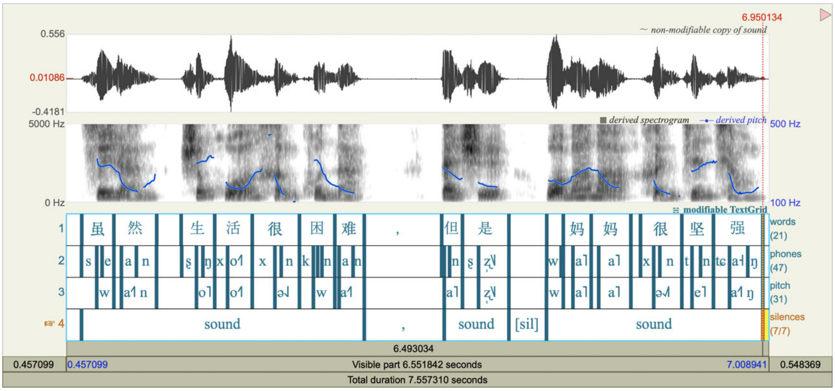


Figure 1. Example of measurement annotation

To quantify temporal features, pronunciation rate was calculated by dividing the number of syllables in a sentence by its total duration in seconds, whereas articulation rate was computed by dividing the number of syllables by the articulation time, excluding all pause intervals. Pauses were analyzed in terms of both count and duration on a per-sentence basis, and each was further examined with respect to two pause types: comma-aligned pauses and non-comma-aligned pauses (hereafter comma pauses and non-comma pauses, respectively). This distinction is particularly relevant given that, unlike Korean, Chinese orthography does not employ spaces between words. As a result, commas serve as the sole explicit visual indicators of intra-sentence pauses available to L2 learners. Chinese orthography distinguishes between two types of commas: the standard comma (*douhao*), which separates clauses or indicates intra-sentential pauses, and the enumeration comma (*dunhao*), which separates list items. Only the standard comma was analyzed for comma pauses, as the enumeration comma appeared only once in the dataset.[5] Pause duration was assessed using two measures: (1) the mean length of comma pauses; and (2) the mean length of non-comma pauses. Pause count was analyzed using three measures: (1) the number of comma pauses; (2) the number of non-comma pauses; and (3) the number of pauses per syllable. The number of pauses per syllable was normalized by dividing the number of non-comma pauses by the total number of syllables in each sentence. To identify pauses, this study adopted a 250-millisecond threshold, consistent with previous research on pause detection (Goldman-Eisler 1968: 12; Campione and Véronis 2002; Heldner and Edlund 2010; De Jong and Bosker 2013; Kahng 2018; Liu et al. 2022; Huang and Qian 2023).

For intonational analysis, fifteen equally spaced pitch values were extracted from each syllable within a sentence and were subsequently converted into semitones using a reference frequency of 100 Hz. To avoid consonantal perturbation, measurements were taken from the rime, defined as the nuclear vowel, optionally preceded by an onglide and followed by an offglide or nasal coda. In instances where pitch-tracking errors (e.g., pitch doubling or halving) occurred, erroneous values were manually corrected through careful inspection of the acoustic signal, corresponding annotation tiers, and the pitch contour. The mean pitch was computed by averaging all extracted pitch values within each sentence, and the pitch range was defined as the difference between the maximum and minimum pitch values. Given that the intonational patterns of questions in Mandarin, shaped by syntactic and contextual factors as well as lexical tones, remain controversial (Shen 1990; Lee 2005; Liu and Xu 2005; Peng et al. 2005; Liu et al. 2006; Lin and Li 2011), the present study examined sentential pitch declination patterns in a total of 1,258 declarative sentences without internal commas to minimize potential confounds due to question intonation and intonational phrase boundaries associated with pauses.[6] Sentential pitch declination was quantified using three methods: (1) an overall slope derived from a linear regression fitted to all pitch values and their corresponding time points; (2) a slope based on the maximum F0 value and its corresponding time point; and (3) a slope based on the minimum F0 value and its corresponding time point. Given that Mandarin exhibits pitch fluctuations caused by lexical tones, which differentially affect pitch range as defined by maximum

5 Only Sentence #2 of Set 01 has an enumeration comma.

6 Of the 1,258 sentences, 634 and 624 were produced by the beginner- and advanced-level groups, respectively. Sentence lengths range from 5 to 20 words and do not differ from those in the full speech dataset. See the Appendix for the 101 declarative sentences used for intonational analysis, as indicated by an asterisk (*).

and minimum F0 values, we used these three measures to capture sentential intonation characteristics more comprehensively. This approach also helps avoid potential misinterpretation of measurement results arising from methodological differences. Table 2 provides a summary of the temporal and intonational measurement procedures, while Figure 2 illustrates an example of declination analysis of the sentence *Ta shi wo gege de haizi* ("He is my brother's kid").

Table 2. Temporal and intonational measurements

Temporal measurements		
speech rate	Pronunciation rate	Number of syllables per second, including pauses
	Articulation rate	Number of syllables per second, excluding pauses
pause count	Number of comma pauses	Number of pauses at orthographic comma positions
	Number of non-comma pauses	Number of pauses not at orthographic comma position
	Number of pauses per syllable	(Number of non-comma pauses) / (number of syllables within sentence)
pause duration	Mean length of comma pauses	(Length of comma pauses) / (number of comma pauses)
	Mean length of non-comma pauses	(Length of non-comma pauses) / (number of non-comma pauses)
Intonational measurements		
pitch	Mean pitch	Mean of fifteen pitch values measured from each syllable
	Pitch range	Pitch range across each sentence (max pitch – min pitch)
sentential declination	Overall slope	Slope of linear regression line fitted to all pitch measures from all syllables within sentence
	Max slope	Slope of linear regression line fitted to all maximum pitch measures from all syllables within sentence
	Min slope	Slope of linear regression line fitted to all minimum pitch measures from syllables within sentence

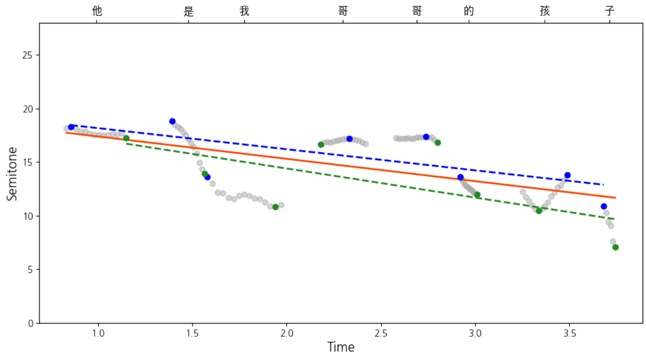


Figure 2. Sentential declination measurements
(red solid line: overall slope, blue dashed line: maximum-pitch slope, green dashed line: minimum-pitch slope)

Temporal and intonational correlates of Korean learners' fluency were examined using Pearson correlation tests and linear mixed-effects models (Section 4.1), while the relative contributions and predictive power of these correlates for perceived fluency were evaluated using linear mixed-effects and Random Forest models (Section 4.2).

4. Results

4.1 Temporal and intonational correlates of L2 Mandarin fluency

To examine whether the temporal and intonational correlates differed between the beginner- and advanced-level groups, a linear mixed-effects model was employed, with proficiency level treated as a fixed effect and learner identity included as a random effect. The results show that Korean learners of Mandarin exhibited distinct temporal patterns across proficiency levels. As presented in Table 3, the beginner-level learners demonstrated significantly slower pronunciation and articulation rates, along with a higher number of pauses per syllable and more non-comma pauses. The mean length of non-comma pauses was also significantly longer in the beginner group. By contrast, neither the number nor the mean length of comma pauses showed significant between-group differences. These results indicate that Korean learners speak significantly faster and use fewer non-comma pauses as their fluency progresses, while both beginner and advanced learners appear to pause in a similar fashion at orthographic comma positions.

Table 3. Temporal measurement results (group means and *t*-values)

Temporal measurements		Beginner	Advanced	*t*-value	*p*-value
speech rate (syll/sec)	Pronunciation rate	2.94	3.78	-5.71	$p<0.01$
	Articulation rate	3.06	3.86	-5.75	$p<0.01$
pause count	Number of pauses per syllable	0.02	0.00	2.30	$p=0.03$
	Number of comma pauses	0.17	0.17	0.02	$p=0.98$
	Number of non-comma pauses	0.27	0.07	2.19	$p=0.04$
pause duration (sec)	Mean length of comma pauses	0.46	0.45	0.69	$p=0.50$
	Mean length of non-comma pauses	0.40	0.34	2.34	$p=0.03$

Figure 3 illustrates strong positive correlations between pronunciation and articulation rates and prosodic fluency scores, while strong negative correlations were observed for the number of pauses per syllable and the number of non-comma pauses. Notably, while non-comma pauses were substantially related to perceived fluency, their frequency of occurrence was more closely associated with fluency judgments than increases in pause duration once the 250 ms threshold was exceeded. By contrast, neither the frequency nor the duration of comma pauses exhibits a significant relationship with perceived fluency. Commas, which serve as strong visual cues to pause, may attenuate fluency-related differences, particularly in read speech. Nevertheless, the duration of comma pauses was slightly longer in beginner-level learners, albeit without reaching statistical significance.[7]

7 As one of the reviewers suggested, no significant effects of comma pauses were found, possibly because commas serve as strong visual cues that may neutralize fluency-related differences in the reading paradigm. However, results from the present study, together with further analyses of additional datasets from our ongoing research, indicate that the presence of commas does not fully neutralize fluency-related differences between beginner- and advanced-level learners. Specifically, beginner-level learners exhibited modestly longer pause durations than advanced-level learners.

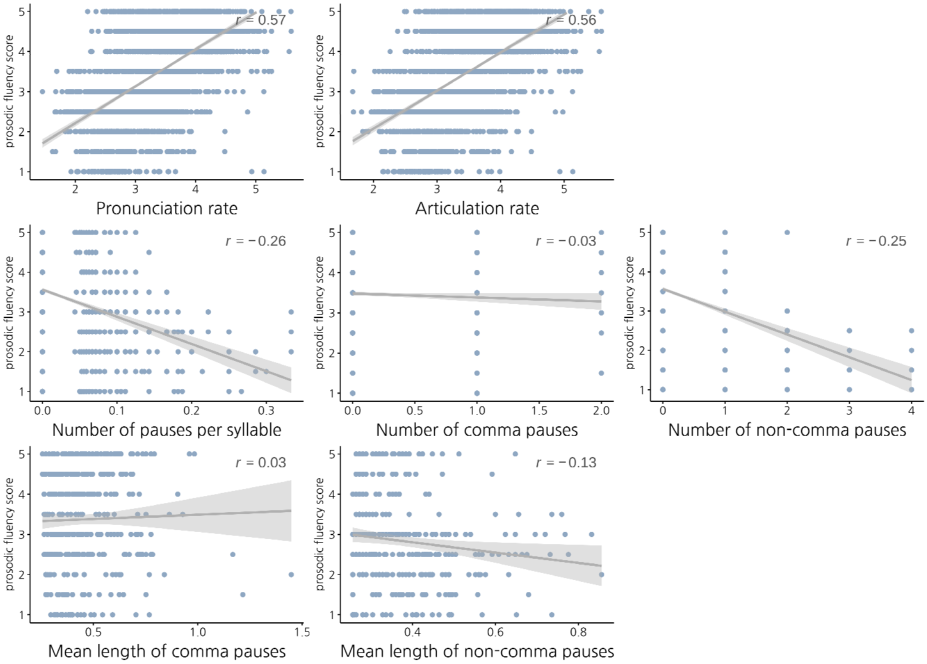


Figure 3. Pearson correlations between temporal measures and prosodic fluency scores

Intonational patterns also differed between beginner- and advanced-level learners, with significant differences observed in sentential pitch declination but not in mean pitch or pitch range. Differences in declination were observed across all three measures: overall slope, maximum pitch slope, and minimum pitch slope. These findings indicate that advanced-level learners tend to produce a significantly steeper overall pitch downtrend in statements, while mean pitch and pitch range show minimal variation with proficiency level. Figure 4, which illustrates the relationships between each intonational measure and perceived fluency, confirms the negative correlations between the degree of overall pitch declination and prosodic fluency scores. No such correlations were observed for mean pitch or pitch range.

Table 4. Intonational measurement results (group means and *t*-values)

Intonational measures		Beginner	Advanced	*t*-value	*p*-value
pitch (semitone)	Mean pitch	13.28	13.32	-0.10	p=0.92
	Pitch range	11.79	13.16	-1.46	p=0.16
sentential declination (semitone/sec)	Overall slope	-0.79	-1.80	3.85	p<0.01
	Max slope	-0.63	-1.62	4.43	p<0.01
	Min slope	-0.99	-2.15	3.65	p<0.01

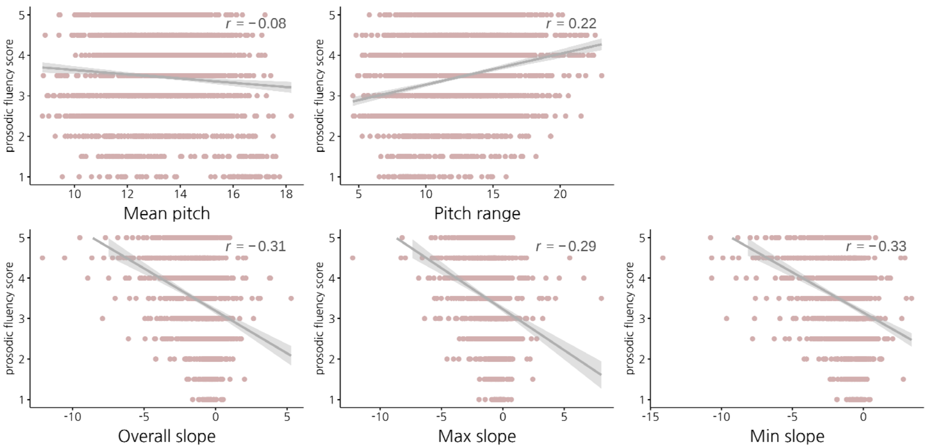


Figure 4. Pearson correlations between pitch measures and prosodic fluency scores

In Mandarin, shorter utterances tend to exhibit steeper declination, even when the effects of initial pitch rise and final lowering are excluded. This pattern suggests that declination is not merely a by-product of tonal concatenation but a global prosodic feature that reflects whole-phrase planning (Shih 2000; Huang et al. 2009; Yuan and Liberman 2014). Therefore, an additional analysis was conducted to examine whether fluency differences correlate with sentential declination slopes as a function of utterance length.

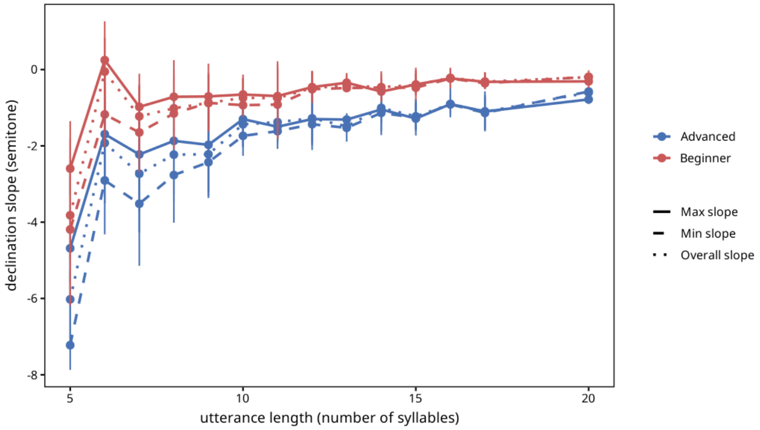
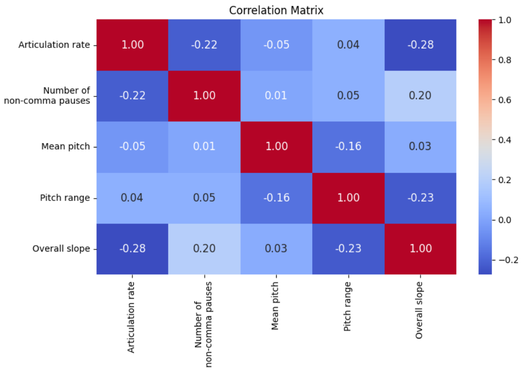


Figure 5. Relationship between declination slope and utterance length

As shown in Figure 5, as sentence length increases, declination generally became less steep in both learner groups. This relationship between declination slope and utterance length was more evident in the advanced group (beginner: r = 0.25; advanced: r = 0.34). This suggests that utterance-level prosodic planning in advanced learners may more closely approximate that of native speakers (Yuan and Liberman 2014). While sentences of five or six syllables appear to deviate from the general declination patterns observed in longer sentences, this likely reflects the stronger influence of individual lexical tones and prosodic structure, particularly in utterance-initial positions, on the overall pitch characteristics of shorter utterances (Wang, A. et al. 2004; Wang, B. et al. 2004; Huang et al. 2009). Although the influence of utterance-initial lexical tones on sentential declination is beyond the scope of this study, the relationship between lexical tones and declination in L2 Mandarin certainly warrants further investigation.

4.2 Predictive contribution of fluency correlates to L2 Mandarin fluency

Building on the analysis of temporal and intonational differences across proficiency levels, the relative contributions of fluency measures were examined using a linear mixed-effects model fitted with statsmodels (Seabold and Perktold 2010) in Python (version 3.10.16). Articulation rate, the number of non-comma pauses, and sentential pitch declination, defined by the overall slope,[8] mean pitch, and pitch range, were treated as fixed effects, while learner identity was included as a random effect. The prosodic score was treated as the dependent variable. To reduce multicollinearity, pronunciation rate, the number of pauses per syllable, the mean length of non-comma pauses, maximum pitch slope, and minimum pitch slope were not included. Low correlations among the predictor variables in the model were confirmed by the variance inflation factor (VIF) values presented in Figure 6. Although some variable pairs exhibited weak correlations (all below 0.3), all predictor variables had VIF values below 5, indicating that multicollinearity was not a concern in the model. The coefficient of determination (r^2) for the model, calculated from the full dataset, was 0.820, and articulation rate, the number of non-comma pauses, and the overall declination slope were identified as significant predictor variables, as summarized in Table 5.



Variable	VIF
Articulation rate	1.120
Number of non-comma pauses	1.084
Mean pitch	1.030
Pitch range	1.097
Overall slope	1.174

Figure 6. Correlation matrix and variance inflation factors (VIF) for predictor variables in the linear mixed-effects model

8 The overall slope, rather than the maximum pitch slope or minimum pitch slope, was included as it better represents sentential declination patterns, given that maximum and minimum F0 values may be influenced by lexical tone types, which could not be controlled due to the nature of the speech datasets used in this study.

Table 5. Linear mixed-effects model results for predictor variables

Fixed effects	Estimate	Std. Error	df	*t*-value	*p*-value
(Intercept)	2.712	0.295	182.5	9.197	< 0.001 ***
articulation rate	0.162	0.037	1247	4.353	< 0.001 ***
number of non-comma pauses	- 0.071	0.032	1230	-2.230	0.026 *
overall slope	- 0.026	0.010	1228	-2.675	0.008 **
mean pitch	0.010	0.013	1243	0.747	0.455
pitch range	0.003	0.007	1241	0.523	0.601

(Signif. levels: '***' < 0.001, '**' < 0.01, '*' < 0.05)

Random effects	Name	Variance	Std. Dev
learner	(Intercept)	0.884	0.940
residual		0.223	0.472

However, when the dataset was split into training and test sets at an 80:20 ratio and the model was fitted and evaluated, the coefficient of determination (r^2) decreased to 0.128, indicating a substantial reduction in explanatory power. This reduction may be attributable to inter-speaker variability, given the relatively small dataset consisting of 28 speakers who each produced 100 utterances.[9] To improve prediction accuracy and account for interactions and nonlinear relationships among predictors, a Random Forest analysis was conducted using scikit-learn (version 1.6.1; Pedregosa et al. 2011) to compute feature importances. The Random Forest model ensemble learning method improves prediction accuracy by reducing variance through the construction of multiple decision trees, each trained on random subsets of the data. To further decorrelate the trees and enhance generalizability, each tree was also trained on a randomly selected subset of predictor variables. The dataset was divided into training

9 This substantial reduction in explanatory power may primarily stem from inter-speaker variability. Although the model has limitations in its use as a fully generalizable automatic scoring tool at this stage, it nevertheless provides meaningful information as a tool for explanatory inference from a phonetic perspective, particularly in identifying the variables that contribute to perceived fluency beyond speaker-specific differences. While within-speaker consistency was ensured through the analysis of a dataset in which 28 speakers each produced 100 utterances, the substantial inter-speaker variability characteristic of L2 Mandarin speech data may limit the model's ability to generalize to new speakers. Future research incorporating a larger number of speakers is expected to yield more robust generalization performance.

and test sets in an 80:20 ratio, and a smaller number of randomly chosen variables were used in each iteration. Figure 7 presents the results of the permutation importance analysis, where greater values indicate more influential predictors. The results show that articulation rate was still the most important predictor, followed by mean pitch, overall slope, and pitch range, none of which showed substantial differences. The number of non-comma pauses was found to be of relatively low importance.

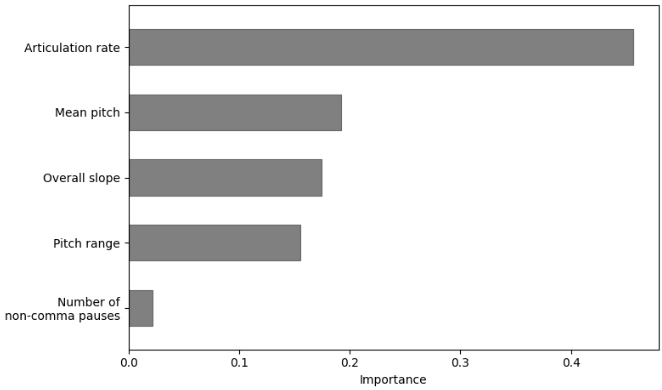


Figure 7. Random Forest model results for predictor variables

Note that while articulation rate was the most critical fluency predictor, it did not show a linear relationship with perceived fluency. As shown by the LOWESS trendline (locally weighted scatterplot smoothing) in Figure 8, the relationship between articulation rate and prosodic fluency score, based on a two-break segmented regression analysis, revealed a clear nonlinear trend: fluency scores increased substantially between approximately 2.72 and 4.15 syllables per second. Perceived fluency showed little change below this range and even slightly declined beyond it. This indicates that excessively slow or fast articulation is associated with lower fluency ratings.

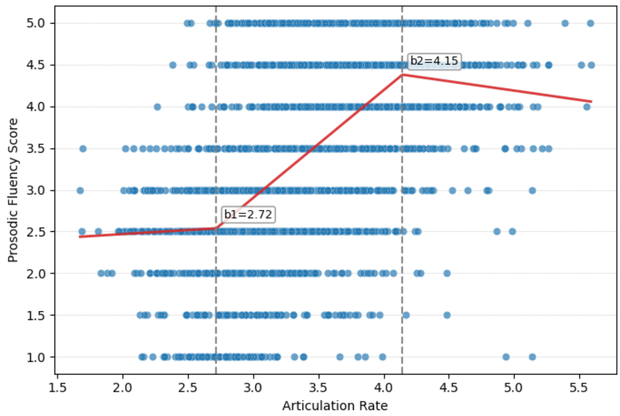


Figure 8. Relationship between articulation rate and prosodic fluency scores

The differences between the linear mixed-effects model results and the Random Forest results require explanation, as the latter indicate that mean pitch and pitch range were relatively important predictors, despite not reaching statistical significance in the former. These differences can be attributed to the distinct modeling approaches employed: mixed-effects models examine the independent linear relationship between each predictor and the dependent variable, whereas Random Forest models estimate variable importance by incorporating complex interactions and nonlinear relationships among variables. Consequently, the sets of variables identified as significant may differ across models. In particular, the finding that mean pitch and pitch range were relatively important only in the Random Forest model suggests that, although these variables may not exert significant linear effects individually, they play an important role in predicting prosody scores when combined with other variables. While the primary goal of the present study is to identify the extent to which specific acoustic features contribute to perceived fluency, future research should provide a more detailed account of the nonlinear relationships suggested by the Random Forest results, which may lead to a better understanding of the complexity of fluency judgment.

5. Discussion and conclusion

The present study examined the temporal and intonational features contributing to Korean learners' Mandarin speaking fluency using sentence-reading speech drawn from a large-scale corpus constructed to support the automatic assessment of Korean learners' L2 speech. The results of the analysis of temporal features showed that pronunciation rate (including pauses) and articulation rate (excluding pauses) were key factors correlated with perceived fluency. The importance of speech rate aligns with its strong negative effect on perceived L2 fluency across languages (Munro and Derwing 1995a, 1995b; Guion et al. 2000; Trofimovich and Baker 2006; Aoyama and Guion 2007; Mennen 2007; Baker-Smemoe et al. 2014; Tavakoli et al. 2020; Yan et al. 2021; Lee 2024). Pauses were also closely correlated with fluency perception: the number of pauses produced without corresponding orthographic commas was particularly important, whereas pauses at orthographic commas were produced in a similar manner across learners of differing proficiency levels. Given that no word-based spacing is present in Chinese orthography, Korean learners, regardless of their fluency level, tend to rely on commas for pausing. Intonational patterns also contributed significantly to fluency perception. It is noteworthy that sentential declination played a significant role in fluency perception: advanced-level learners tend to produce a significantly steeper overall pitch declination in statements, as observed across the overall, maximum-pitch, and minimum-pitch slopes. This finding suggests that not only lexical tones but also global intonational characteristics contribute to L2 Mandarin fluency.

Another novel finding of this study pertains to the relative contributions of temporal and intonational features to the perception of fluency in Korean learners' Mandarin. The results of the predictive modeling revealed that speech rate, operationalized as articulation rate, was the most critical predictor of perceived fluency. At the same time, the relationship between speech rate and perceived fluency was nonlinear: increases in rate exerted a positive influence only within a certain range (approximately 2.72–4.15 syllables per second), suggesting that further changes in rate may not affect fluency perception when speech is markedly slow or fast. Given that native Mandarin speakers produce approximately 3.18 syllables per second in slow read speech and 4.72 syllables per second in normal read speech (Wang, M. 2020), both overly slow and overly fast speech relative to native norms appear to be dispreferred in L2 speech. In addition, the results showed that sentential declination measures were stronger predictors than pause measures, implying that global intonational characteristics play an important role in fluency perception in a language where pitch is used for lexical tone contrasts. The significance of sentential declination effects may reflect fluency differences in speech planning, which are linked to cognitive fluency. However, given that the present findings are based on sentence-reading speech, no conclusive claims can be made regarding the relationship between sentential declination and cognitive fluency in Segalowitz's (2010) framework. Further research using spontaneous speech is therefore needed.

While prosodic correlates of Korean learners' Mandarin were the primary focus of this study, the potential influence of pronunciation accuracy on fluency perception also warrants consideration. Given the significance of the random intercept in the mixed-effects model summarized in Table 5, the random intercepts for each learner were additionally examined. In Figure 9, the y-axis, which represents the estimated random intercept for each learner, indicates the extent to which their predicted fluency scores deviate from the model's overall intercept.[10] Positive values correspond to learners whose predicted scores exceed the model average, whereas negative values indicate scores below the average. Note that the advanced group exhibits positive values, whereas the beginner group displays negative values.

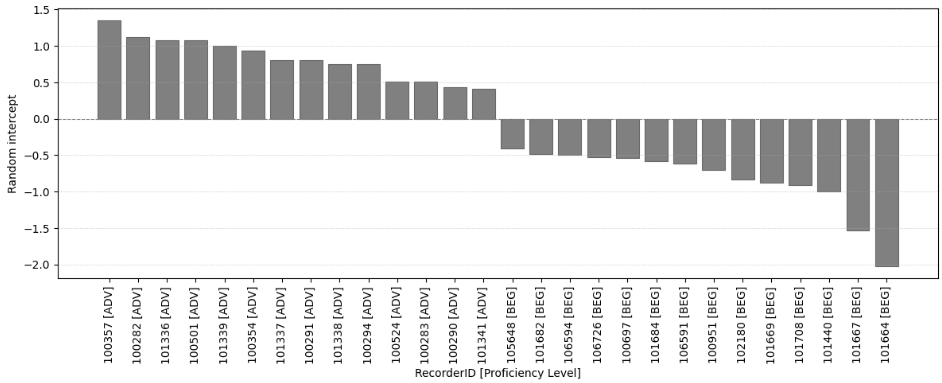


Figure 9. Random intercept values across learners

10 The labels for the learners on the x-axis correspond to the recorder IDs used in the AI Hub L2 Mandarin corpus.

The predictability of the Random Forest model also increased when pronunciation accuracy was incorporated: in Figure 10, this is illustrated by the red dashed line, which represents the ideal prediction line where predicted values would exactly match the actual scores. Note that the plot on the right, which represents the total proficiency score combining prosodic fluency and pronunciation accuracy, demonstrates a tighter clustering of predicted and actual values (r^2 = 0.853), compared with the plot on the left, which represents the fluency score (r^2 = 0.451).

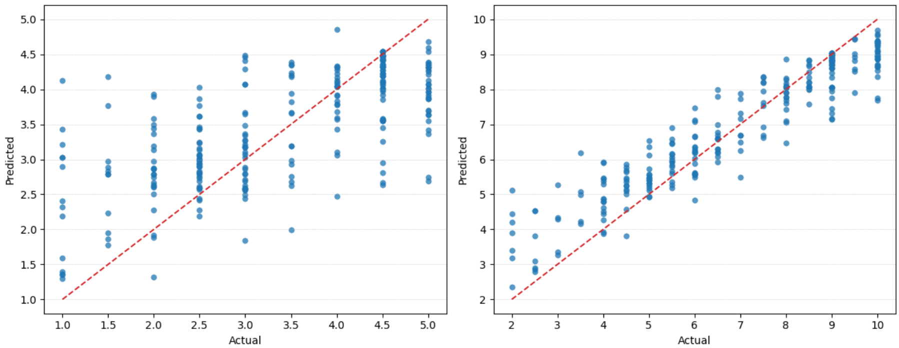


Figure 10. Observed vs. predicted scores from Random Forest model
(left: prosodic fluency scores, right: total proficiency scores)

In summary, the findings of this study identify the key temporal and intonational features associated with perceived fluency in Korean learners' Mandarin and evaluate their relative contributions. Given the paucity of research on the speaking fluency of Korean learners of Mandarin, especially in stark contrast to the extensive literature on lexical tone production, the present study not only advances our understanding of L2 development in contexts where substantial prosodic differences exist between the L1 and L2, but also sheds light on the fluency features that should be prioritized in the design of automated speaking proficiency assessments. Although this study pertains to read speech, the findings provide insights into fluency measures in spontaneous L2 Mandarin speech, an area that warrants further research. In addition, this study found an interdependence between prosodic fluency and pronunciation accuracy, consistent with patterns reported for English learners' Mandarin and Korean learners' English (Yang et al. 2021; Chung 2023; Kim 2024); however, the relationship between fluency and accuracy measures in Korean learners' Mandarin remains an important area for future investigation.

References

Altenberg, Evelyn P. and Carole T. Ferrand. 2006. Fundamental frequency in monolingual English, bilingual English/Russian, and bilingual English/Cantonese young adult women. *Journal of Voice* 20(1): 89-96.

Aoyama, Katsura and Susan G. Guion. 2007. Prosody in second language acquisition. In Ocke-Schwen Bohn and Murry J. Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*, 281-297. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Backman, Nancy Ellen. 1979. Intonation errors in second language pronunciation of eight Spanish speaking adults learning English. *Interlanguage Studies Bulletin* 4(2): 239-266.

Baker-Smemoe, Wendy, Dan P. Dewey, Jennifer Bown, and Rob A. Martinsen. 2014. Does measuring L2 utterance fluency equal measuring overall L2 proficiency? Evidence from five languages. *Foreign Language Annals* 47(4): 707-728.

Bao, Jie. 2020. A multi-dimensional study on the oral fluency of Korean Chinese learners. *Modern Chinese* (7): 112-118.

Bradlow, Ann R., Gina M. Torretta, and David B. Pisoni. 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20(3-4): 255-272.

Campione, Estelle and Jean Véronis. 2002. A large-scale multilingual study of silent pause duration. In Bernard Bel and Isabelle Marlien (eds.), *Proceedings of Speech Prosody 2002*, 199-202. Aix-en-Provence, France: Laboratoire Parole et Langage.

Chen, Gwang-tsai. 1974. Pitch range of English and Chinese speakers. *Journal of Chinese Linguistics* 2(2): 159-171.

Chen, Mo and Qing Zhou. 2016. On the development of Chinese reading fluency of Korean native speakers. *TCSOL Studies* (2): 1-9.

Chen, Mo and Qing Zhou. 2017. Effects of L1 and prosodic hierarchy on prosodic boundary production by advanced CSL learners. *Chinese Teaching in the World* 31(2): 270-288.

Chen, Sheng H. 2005. The effects of tones on speaking fundamental frequency and intensity ranges in Mandarin and Min dialects. *The Journal of the Acoustical Society of America* 117(5): 3225-3230.

Cheng, Andrew. 2020. Cross-linguistic F0 differences in bilingual speakers of English and Korean. *The Journal of the Acoustical Society of America* 147(2): EL67-EL73.

Chung, Hyunsong. 2023. Comparing the effects of letter-based and syllable-based speaking rates on the pronunciation assessment of Korean speakers' English. *Phonetics and Speech Sciences*

15(4): 1-10.

Cui, Ting. 2021. *A case study on the development of speaking fluency and speech self-monitoring in Korean learners of Chinese*. Master's Thesis. Nanjing Normal University.

De Jong, Nivja H. and Hans Rutger Bosker. 2013. Choosing a threshold for silent pauses to measure second language fluency. In Robert Eklund (ed.), *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech (DiSS)*, 17-20. Stockholm: KTH Royal Institute of Technology.

Eady, Stephen J. 1982. Differences in the F0 patterns of speech: Tone language versus stress language. *Language and Speech* 25(1): 29-42.

Fu, Yuping and Yong-cheol Lee. 2022. The production of tone 3 by advanced Korean learners of Mandarin. *Linguistic Research* 39(1): 213-233.

Gårding, Eva. 1987. Speech act and tonal pattern in Standard Chinese: Constancy and variation. *Phonetica* 44(1): 13-29.

Goldman-Eisler, Frieda. 1968. *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.

Graham, Calbert. 2014. Fundamental frequency range in Japanese and English: The case of simultaneous bilinguals. *Phonetica* 71(4): 271-295.

Guion, Susan G., James E. Flege, Serena H. Liu, and Grace H. Yeni-Komshian. 2000. Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics* 21(2): 205-228.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Han, Seoyoung. 2017. A case study of the pronunciation of the Chinese velar fricative by a Korean advanced learner of Chinese. *Korea Journal of Chinese Linguistics* (71): 51-81.

Han, Seunghee, Minhwa Chung, and Sunhee Kim. 2024. Constructing Korean learners' L2 speech of seven languages for automatic pronunciation assessment. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue (eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 3772-3781. Torino: ELRA and ICCL.

Heldner, Mattias and Jens Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics* 38(4): 555-568.

Huang, Changbin and Yulian Qian. 2023. Oral Fluency of L2: Concept, indicator and model. *Chinese Language Learning* (4): 82-92.

Huang, Xianjun, Lu Gao, Yufang Yang, and Shinan Lü. 2009. Experimental study on declination in Chinese intonation. *Acta Acustica* 34(2): 158-166.

Huang, Yi-Hsuan and Janice Fon. 2011. Investigating the effect of Min on dialectal variations of Mandarin tonal realization. In Wai-Sum Lee and Eric Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, 918-921. Hong Kong: City

University of Hong Kong.
Jenner, Bryan. 1976. Interlanguage and foreign accent. *Interlanguage Studies Bulletin* 1(2/3): 166-195.
Jilka, Matthias. 2007. Different manifestations and perceptions of foreign accent in intonation. In Jürgen Trouvain and Ulrike Gut (eds.), *Non-native prosody: Phonetic description and teaching practice*, 77-96. Berlin: De Gruyter Mouton.
Kahng, Jimin. 2018. The effect of pause location on perceived fluency. *Applied Psycholinguistics* 39(3): 569-591.
Kang, Hyeon-Seok. 2017. Is English being challenged by Mandarin in South Korea? *English Today* 33(4): 40-46.
Kang, Seok-Han. 2022. *The fluency acquisition and evaluation of English and Korean*. Seoul: Sin-A Press.
Kang, Seok-Han, Hyunkee Ahn, Eunsil Hong, Byeonggon Min, Su-Jin Cho, Sungjun Lee, and Hyun-Jeong Park. 2017. A study of Korean production and fluency as a foreign language. *Bilingual Research* (67): 1-29.
Keating, Patricia and Grace Kuo. 2012. Comparison of speaking fundamental frequency in English and Mandarin. *The Journal of the Acoustical Society of America* 132(2): 1050-1060.
Kim, Mikyoung. 2024. *A basic study for evaluating and providing feedback on the Korean fluency of non-native speakers*. Master's Thesis. Seoul National University.
Kong, Seok-lim and Hyun-cheol Kim. 2018. Research of Korean Chinese learners' speaking fluency. *The Journal of Chinese Linguistics* (77): 333-358.
Kuznetsova, Alexandra, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13): 1-26.
Ladd, D. Robert. 1996/2008. *Intonational phonology*. Cambridge: Cambridge University Press.
Lee, Binna and Diana Van Lancker Sidtis. 2017. The bilingual voice: Vocal characteristics when speaking two languages across speech tasks. *Speech, Language and Hearing* 20(3): 174-185.
Lee, Haneul. 2024. *An acoustic study of voice quality variations in Korean-English bilinguals*. Master's Thesis. Seoul National University.
Lee, Ok Joo. 2005. *The question prosody of Beijing Mandarin*. Phd Dissertation. The Ohio State University.
Lee, Ok Joo. 2019. Global pitch characteristics of tonal and non-tonal language speakers: Evidence from Mandarin-Korean bilinguals. *Journal of Chinese Literature* 98: 267-300.
Lee, Ok Joo and Jieun Park. 2025. A critical evaluation of AI Hub Korean learners' Mandarin speech corpus: structures and applications. *Journal of Chinese Literature* 122: 309-333.
Lee, Yunjung. 2016. Quantitative analysis of oral fluency. *Korea Journal of Chinese Language* (62): 193-216.
Lee, Yunjung. 2019. Comparative study of Chinese and Korean oral fluency. *The Journal of Chinese Linguistics* (81): 403-423.

Lee, Yunjung. 2020. A comparative study on speaking fluency by Chinese speaking task types. *The Journal of Chinese Linguistics* (87): 371-393.

Lin, Maocan and Zhiqiang Li. 2011. Focus and boundary in Chinese intonation. In Wai-Sum Lee and Eric Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, 1246-1249. Hong Kong: City University of Hong Kong.

Liu, Fang and Yi Xu. 2005. Parallel encoding of focus and interrogative meaning. *Phonetica* 62(2-4): 70-87.

Liu, Fang, Dinoj Surendran, and Yi Xu. 2006. Classification of statement and question intonations in Mandarin. In Rüdiger Hoffmann and Hansjörg Mixdorff (eds.), *Proceedings of Speech Prosody 2006*, 5-25. Dresden: TUDpress.

Liu, Shimeng, Yoshitaka Nakajima, Lihan Chen, Sophia Arndt, Maki Kakizoe, Mark A. Elliott, and Gerard B. Remijn. 2022. How pause duration influences impressions of English speech: Comparison between native and non-native speakers. *Frontiers in Psychology* 13: 778018.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. 2017. Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. In Francisco Lacerda (ed.), *Proceedings of Interspeech 2017*, 498-502. Stockholm: ISCA.

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

Mennen, Ineke, Felix Schaeffler, and Gerard Docherty. 2012. Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America* 131(3): 2249-2260.

Munro, Murray J. and Tracey M. Derwing. 1995a. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38(3): 289-306.

Munro, Murray J. and Tracey M. Derwing. 1995b. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45(1): 73-97.

Ng, Manwa L., Gigi Hsueh, and Cheung-Shing Sam Leung. 2010. Voice pitch characteristics of Cantonese and English produced by Cantonese-English bilingual children. *International Journal of Speech-Language Pathology* 12(3): 230-236.

Ng, Manwa L., Yang Chen, and Ellen Y. K. Chan. 2012. Differences in vocal characteristics between Cantonese and English produced by proficient Cantonese-English bilingual speakers – a long-term average spectral analysis. *Journal of Voice* 26(4): e171-e176.

Oh, Hanna. 2021. A study on the limitations of Chinese tone by analyzing phonetic errors. *Culture and Convergence* 43(2): 437-452.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12: 2825-2830.

Peng, Shu-hui, Marjorie K. M. Chan, Chiu-yu Tseng, Tsan Huang, Ok Joo Lee, and Mary Beckman. 2005. Towards a Pan-Mandarin system for prosodic transcription. In Sun-An Jun (ed.), *Prosodic typology I: The phonology of intonation and phrasing*, 230-270. Oxford: Oxford University Press.

Python Software Foundation. 2024. Python: A dynamic, open source programming language (Version 3.10.16). URL https://www.python.org/.

R Core Team. 2024. R: A language and environment for statistical computing (Version 4.3.3). R Foundation for Statistical Computing. URL https://www.R-project.org/.

Seabold, Skipper and Josef Perktold. 2010. Statsmodels: Econometric and statistical modeling with Python. In Stéfan van der Walt, and Jarrod Millman (eds.), *Proceedings of the 9th Python in Science Conference*: 92-96. Austin, Texas: SciPy.

Segalowitz, Norman. 2010. *Cognitive bases of second language fluency*. New York, NY: Routledge.

Shen, Xiao-Nan. 1990. *The prosody of Mandarin*. Berkeley: The University of California Press.

Shih, Chihlin. 2000. A declination of Mandarin Chinese. In Antonis Botinis (ed.), *Intonation: Analysis, modelling, and technology*, 243-268. Dordrecht; Boston: Kluwer Academic Publishers.

Tao, Hongyin. 1996. *Units in Mandarin conversation: Prosody, discourse, and grammar*. Philadelphia, PA: John Benjamins Publishing Company.

Tavakoli, Parvaneh, Fumiyo Nakatsuhara, and Ann–Marie Hunter. 2020. Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal* 104(1): 169-191.

Thornbury, Scott. 2005. *Beyond the sentence*. Oxford: Macmillan Education.

Torgerson, Richard Christen. 2005. *A comparison of Beijing and Taiwan Mandarin tone register: An acoustic analysis of three native speech styles*. Master's Thesis. Brigham Young University.

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.

Wang, Anhong, Ming Chen, and Shinan Lü. 2004. The study of declination in speech database in standard Chinese. *Acta Acustica* 29(4): 353-358.

Wang, Bei, Yufang Yang, and Shinan Lü. 2004. Acoustic analysis on prosodic hierarchical boundaries of Chinese. *Acta Acustica* 29(1): 29-36.

Wang, Guanyang. 2020. *Empirical study on impact of speaking and writing order on oral fluency of initial Chinese learners in South Korea*. Master's Thesis. Shandong Normal University.

Wang, Maolin. 2020. The acoustic effect of speaking rate, focus, and prosodic position on syllables in Chinese. *Journal of Chinese Linguistics* 48(1): 174-205.

Wang, Maolin, Guangling Zi, Wei Xiong, and Maocan Lin. 2013. Pitch declination of Chinese spontaneous speech. *Journal of Chinese Information Processing* 27(6): 128-133.

Wennerstrom, Ann. 1994. Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics* 15(4): 399-420.

Willems, Nico J. 1982. *English intonation from a Dutch point of view*. Dordrecht: Foris

Publications.

Xue, Steve An, Fran Hagstrom, and Jian Ping. 2002. Speaking fundamental frequency characteristics of young and elderly bilingual Chinese-English speakers: A functional system approach. *Asia Pacific Journal of Speech, Language and Hearing* 7(1): 55-62.

Yan, Xun, Ha Ram Kim, and Ji Young Kim. 2021. Dimensionality of speech fluency: Examining the relationships among complexity, accuracy, and fluency (CAF) features of speaking performances on the Aptis test. *Language Testing* 38(4): 485-510.

Yang, Chunsheng, Jing Chu, Si Chen, and Yi Xu. 2021. Effects of segments, intonation and rhythm on the perception of L2 accentedness and comprehensibility. In Yang Chunsheng (ed.), *The acquisition of Chinese as a second language pronunciation: Segments and prosody*, 233-255. Singapore: Springer Singapore.

Yoon, Kiduck. 2013. A rhythmic metrics analysis on the foreign accentedness of Korean speakers' learning Mandarin Chinese. *The Journal of Chinese Linguistics* (45): 491-509.

Yuan, Jiahong and Mark Liberman. 2014. F0 declination in English and Mandarin broadcast news speech. *Speech Communication* 65: 67-74.

Zhang, Hang. 2018. *Second language acquisition of Mandarin Chinese tones: Beyond first-language transfer*. Leiden; Boston: Brill Rodopi.

Zhang, Qiuju. 2018. *A study on the complexity, accuracy and fluency of South Korean students' oral Chinese: Inje university in South Korea*. Master's Thesis. Nanjing University.

Zhu, Shi, Sibie Chong, Yang Chen, Tianqi Wang, and Manwa Lawrence Ng. 2022. Effect of language on voice quality: An acoustic study of bilingual speakers of Mandarin Chinese and English. *Folia Phoniatrica et Logopaedica* 74(6): 421-430.

Zhu, Shifang. 2009. *A study of the dynamic oral text of Korean learners of Chinese*. Master's thesis. Beijing Language and Culture University.

Appendix

Sentences from Korean learners' Mandarin dataset (An asterisk '*' indicates sentences analyzed for intonational measures.)

Set 01

1	同学之间应该互相帮助。*
2	我家有四口人，爸爸、妈妈、姐姐和我。
3	你打算什么时候结婚？
4	虽然生活很困难，但是我妈妈很坚强。
5	最喜欢紫色的人，精神可能不太成熟。
6	小时候女孩子一般比较喜欢爸爸。*
7	她和她丈夫没什么共同话题。*
8	大学一毕业他们就结婚了。*
9	我朋友英语说得很好，我非常羡慕。
10	我们三年前见过一次面，从此再也没联系了。
11	他们结婚十年了，热情也慢慢消失了。
12	你跟你孩子长得太像了。*
13	你心里不高兴，也没有说出来， 我怎么知道你不高兴了呢？
14	孩子没考上好大学，父母很失望。
15	有的夫妻互相没感情了，但是也不离婚。
16	那个人对待客人跟对待自己家人一样。*
17	你跟他确定关系了吗？
18	我奶奶快九十岁了，但脑子一点儿也不糊涂。
19	他们吵了一晚上，吵得邻居都睡不着觉。
20	父母一直夸孩子也不太好。*
21	很多人都会做西红柿炒鸡蛋。*
22	煮猪肉的时候放点儿啤酒，特别香。
23	今天我们去吃自助餐，你想吃什么就吃什么。
24	你喜欢吃什么口味的菜？
25	中国南方人比较喜欢吃甜的。*
26	臭豆腐虽然闻起来臭，但吃起来很香。
27	中国有八个地方的菜很有名，比如广东菜。
28	炖菜最大的优点就是很好做。*
29	饺子吃起来很好吃，但是做起来特别麻烦。
30	我想吃中国菜。*
31	中国人喝茶特别有讲究， 不同茶叶用的杯子都不一样。
32	中秋节吃月饼是中国的传统。*
33	中国人一般不会把主人准备的菜全吃光。*
34	肯德基为了进军中国市场，改了自己的菜单。
35	少放点辣椒，我吃不了太辣的。
36	服装市场上有各种各样的衣服，让人看花了眼。
37	旗袍真是中国婚礼上最漂亮的衣服。*
38	这种衣服早就不流行了。*
39	我想把头发拉直。*
40	你想留刘海吗？
41	那个人穿衣服真的很随便。*

Set 04

1	你知道小王吗？
2	他是我哥哥的孩子。*
3	那个女的是他的妻子。*
4	大家都认识他吗？
5	请你自我介绍。*
6	请问，您贵姓？
7	明天是我弟弟的生日。*
8	这是我们的教室。*
9	那是我姐姐的词典。*
10	我的妹妹很可爱。*
11	我们的教室不小。*
12	这是我爱人。*
13	他们都是留学生。*
14	大家好，认识你们很高兴。
15	这是他的数学老师。*
16	你有兄弟姐妹吗？
17	你家有几口人？
18	他比我大三岁。*
19	昨天他们吵架了。*
20	你跟他一起去吗？
21	你喜欢吃什么？
22	我们去尝尝家常菜！*
23	你觉得哪个更好吃？
24	今天我请你吃饭吧！*
25	你要可乐还是雪碧？
26	这杯奶茶很好喝！*
27	我想吃肉包子和油条。*
28	我们买点儿水果吧。*
29	你要葡萄还是西瓜？
30	这个苹果甜吗？
31	这个菜是煮的还是炒的？
32	我们找个地方吃点东西吧。*
33	这道菜是谁做的？
34	你们先吃饭再说吧。*
35	他长得真帅啊！*
36	我买了一件羽绒服。*
37	这双鞋是多大的？
38	这是最近流行的款式。*
39	你的头发真好看，在哪儿烫的？
40	我可以试试这件衣服吗？
41	学做指甲一般要学多久?

42	你觉得这个牛仔裤怎么样？	42	怎么保养皮肤呢?
43	哪个颜色的眼影适合我？	43	你的头发怎么这么长？
44	这件裙子太小了，有没有更大的？	44	你染过头发吗？
45	你头发颜色很好看，我也想染发。	45	你喜欢玩儿游戏吗？
46	欣赏艺术需要有艺术的眼睛和耳朵。*	46	他是著名的演员。*
47	听说那部电影很好看，有时间一起去看吧。	47	你看过这部电视剧吗？
48	这部文学作品的特色在于它具有鲜明的艺术性。*	48	他从小就对电子游戏有兴趣。*
49	那部电影是根据同名小说改编的。*	49	你喜欢这种音乐节目吗？
50	昨天你有听周杰伦的新歌吗？	50	我没去过美术馆。*
51	不愧是演员，她表演得真好。	51	我特别喜欢看网络小说。*
52	我平时喜欢玩手机游戏。*	52	很多年轻人想成为明星。*
53	那个明星说话很有礼貌。*	53	他唱歌唱得很厉害。*
54	每个明星都会很注重自己的形象。*	54	你会弹奏这首歌吗？
55	韩国综艺节目真的特别好看。*	55	他感冒了，还发烧呢。
56	随着医学技术的发展， 人类征服癌症的日子越来越近了。	56	他看来身体不好。*
57	一般来说，流行性感冒的传染力是很强的。	57	我们要注意个人卫生。*
58	病人的骨折可以通过X射线片子显示。	58	你身体恢复得怎么样？
59	你怎么还在咳嗽，要不我陪你去看医生吧。	59	最近你身体好吗？
60	你打完第三针新冠疫苗了吗？	60	结账时，可直接扫二维码。
61	我们的科学技术还赶不上发达国家。*	61	科学技术的发展给人类带来了许多变化。*
62	科技的发展给人类带来了不少方便。*	62	智能手机无法替代数码相机。*
63	电脑的普及成功实现了办公的现代化。*	63	我手机没电了。*
64	我们采用新技术之后，产品的成本下降了不少。	64	人工智能是否会取代人类？
65	运用先进技术摆脱了落后的状态。*	65	你听过虚拟货币吗？
66	近几年，亚洲的经济发展相当稳定。	66	我办了一张信用卡。*
67	移动支付技术已经普及到了农村地区。*	67	这附近有取款机吗？
68	有了互联网之后，感觉人们之间的关系更疏远了。	68	中国和美国是世界经济强国。*
69	韩国电商平台之间的竞争越来越激烈。*	69	贸易是重要的经济指标之一。*
70	全球经济日益走向全球化。*	70	今天下午有足球比赛，我们一起去看吧。
71	充足的睡眠有助于健康。*	71	为了自己的健康，他每天锻炼身体。
72	如果想保持健康的话，你还是少吃点盐和糖吧。	72	他会打篮球。*
73	长时间打游戏，不仅注意力会不集中， 记忆力也会下降。	73	我比去年胖了五公斤。*
74	每天散散步，还可以减肥。	74	晒太阳对身体有好处。*
75	笑一笑十年少，笑可以让人变年轻。	75	这只猫真可爱。*
76	全世界的重要问题之一是气候变化。*	76	雨下得愈演愈烈，我们快回家吧。
77	因为环境污染问题，动物的生存空间越来越小。	77	外面正在刮大风，今天别出去。
78	你有养过宠物吗？	78	今天天气很晴朗。*
79	所有的宠物都有权受到妥善的对待。*	79	我们要保护生态环境。*
80	气象台预报过，明天会一直下大雨。	80	动物是人类的好朋友。*
81	狗狗看到主人就会摇自己的尾巴。*	81	养宠物不是一件简单的事情。*
82	长颈鹿的脖子是最长的。*	82	一年有四个季节。*
83	熊猫最喜欢吃竹笋。*	83	最近雾霾特别严重。*
84	小猫经常看窗外的小鸟。*	84	春天到了，桃花开了。
85	鲸鱼的祖先是来自陆地的哺乳动物，不是鱼类。	85	我明天七点要去机场接个人。*
86	自行车是最常见的交通工具之一。*	86	寒假去了香港，在香港吃了很多点心。
87	五环之内堵车很严重。*	87	这个季节去哪儿旅游才好呢？
88	要不要跟我去西部旅行？	88	我打算坐地铁去学校。*
89	他们不是坐飞机来的，是坐邮轮来的。	89	他对这个宾馆很满意。*
90	上海人多车多，所以交通很拥挤。	90	飞机快起飞了，她怎么没来？

91	晚上太堵车了，我错过了飞机。	91	平时我骑自行车上学。*
92	他以学生的身份成功购买了半价车票。*	92	这个宾馆虽然不大，但是很安静。
93	坐出租车去的话，大概三十块左右。	93	请问，地铁站离这儿远吗？
94	请出示您的护照。*	94	十一国庆节是中国的重要节日。*
95	请问，公交车站怎么走？	95	这是开往上海的高铁。*
96	火车上不能把头伸出窗外面。*	96	你去过颐和园吗？
97	你要坐高铁还是飞机？	97	我从来没爬过山。*
98	我比较喜欢跟团旅游。*	98	你怎么去北京的？
99	通过旅行社订酒店更优惠。*	99	我要买新的行李箱。*
100	在哪里取行李？	100	这景点的门票太贵。*

Yan Xiong
PhD Candidate
Department of Korean Language and Literature
Seoul National University
1 Gwanak-ro, Gwanak-gu,
Seoul, 08826, Korea
E-mail: xiongyan312@snu.ac.kr

Jieun Park
M.A. Student
Department of Chinese Language and Literature
Seoul National University
1 Gwanak-ro, Gwanak-gu,
Seoul, 08826, Korea
E-mail: jieun22@snu.ac.kr

Ok Joo Lee
Associate Professor
Department of Chinese Language and Literature, Learning Sciences Research Institute
Seoul National University
1 Gwanak-ro, Gwanak-gu,
Seoul, 08826, Korea
E-mail: ojlee@snu.ac.kr

Received: 2025. 11. 25.
Revised: 2025. 12. 31.
Accepted: 2026. 01. 22.