Speech Timing Adaptation of EFL Learners in Dialogic Reading and Implications for Pronunciation Instruction

In Young Yang[1] & Hyunkee Ahn[2†]

[1]Korea National University of Transportation, [2]Seoul National University

ABSTRACT

This study examined how Korean learners of English adjust their speech rhythm in two settings: dialogue reading with a native speaker and monologue reading. Thirty-five university students read English scripts under both conditions. Speech rhythm was analyzed using the vocalic normalized Pairwise Variability Index (nPVI; Grabe & Low, 2002) and compared based on participants' pronunciation nativelikeness. The results showed that rhythm in the dialogue condition was more stress-timed than in the monologue condition. Repeated measures two-way ANOVA revealed significant effects of reading condition, whereas the effects of pronunciation nativelikeness and its interaction with reading condition were not statistically significant. These findings partially support the role of interactional context in second language (L2) pronunciation and suggest that engaging learners in dialogue practice with more stress-timed speakers may promote rhythmic adaptation. Implications for technology-enhanced pronunciation instruction and the need to examine L2 pronunciation in AI-based interactional contexts are discussed.

Keywords: second language pronunciation, pronunciation in interaction, English rhythm, rhythmic adaptation, phonetic convergence, phonetic alignment, pronunciation instruction

1. Introduction

This study examines Korean learners' production of second language (L2) English rhythm during reading in dialogue compared to reading in monologue. Traditionally, L2 pronunciation research has primarily focused on segmental or suprasegmental features in words, phrases, or sentences produced in decontextualized settings, whether for acquisition or instruction (e.g., Yang, 2014, 2017, among others). While L2 pronunciation has been extensively studied in isolated reading or speaking tasks, only a few studies have explored learners' pronunciation in dialogic or interactional

† Corresponding author: ahnhk@snu.ac.kr

https://doi.org/10.30961/lr.2025.61.3.249

contexts (e.g., Chung, 2013a, b, among others). This research further addresses this gap.

Among various aspects of pronunciation, speech rhythm was selected as the focus of the analysis. This decision was supported by its significance in L2 communication and by the availability of rhythm metrics for quantifying speech timing, developed since around 2000, to differentiate between stress-timed and syllable-timed rhythms (Grabe & Low, 2002; Ramus et al., 1999; White & Mattys, 2007). Language rhythm is closely linked to cognitive processing and the information structure of spoken discourse (Birch & Clifton, 1995; Bock & Mazzella, 1983; Cutler, 1976, 1984; Cutler & Fodor, 1979; Cutler & Foss, 1977), making it an important factor in second language education. Languages have been known to have distinct rhythmic patterns, either stress-timed or syllable-timed (Abercrombie, 1967). English and Korean are known to have distinct rhythmic patterns: English is generally classified as a stress-timed language, while Korean is considered syllable-timed, although this is controversial (Kim et al., 2007).

Although the rhythm class hypothesis has been widely discussed in linguistics, empirical and physiological evidence for isochronous timing is lacking, making objective observation and classification difficult. As a result, some researchers have shifted from viewing isochrony as an objective regularity to describing it as a subjective tendency, while others argue that isochrony is primarily a perceptual phenomenon. (Grabe & Low, 2002). However, advances in the quantification of speech timing (Grabe & Low, 2002; Ramus et al., 1999; White & Mattys, 2007) have made such analysis more feasible and reliable, and diverse research has been conducted based on these metrics so far. The current study builds on these developments to explore the rhythmic features of L2 English speech.

The current study further investigates whether speech rhythm is not just a fixed, learner-specific skill but a dynamic feature shaped by pragmatic and social contexts. Gut (2003) also examined non-native speakers and found that rhythm can vary depending on the type of texts, comparing reading and retelling. This study specifically explores whether the presence or absence of an interlocutor—especially one with a more stress-timed rhythm—may affect how learners produce speech rhythm; reading in a dialogue with a native speaker versus reading alone in a monologue could lead to different stress timing.

A dialogic reading context that offers consistent exposure to the native speaker's more stress-timed speech may help learners imitate or adapt their rhythm. This supports the interactive alignment model proposed by Pickering & Garrod (2004), which states that dialogue encourages linguistic and paralinguistic alignment, including phonological matching. Research also shows that phonetic convergence occurs not only among native speakers but also in L2 environments (e.g., Lewandowski, 2012; Trofimovich & Kennedy, 2014, among others). The current study, therefore, measures Korean learners' rhythm using the normalized Pairwise Variability Index (nPVI) based on vowel duration (Grabe & Low, 2002). The current study further explores how pronunciation proficiency influences L2 English speech rhythm production. Additionally, since pronunciation proficiency does not always match overall language ability in L2 learners (Kennedy & Trofimovich, 2017), participants are rated on their pronunciation nativelikeness and grouped based solely on that, with their rhythmic features examined accordingly.

The research questions are as follows.

Research Question 1

How does the Korean learners' production of English rhythm vary between dialogue reading with a native speaker and monologue reading alone?

Research Question 2

How is the English speech rhythm of Korean learners related to their pronunciation nativelikeness, as measured by rhythm metrics?

We expect participants to show higher nPVI (i.e., a more stress-timed rhythm) during dialogue reading than during monologue reading. Among them, those with higher proficiency levels are likely to display more stress-timed speech and adapt more to stress timing, as shown by larger changes in nPVI values from monologue to dialogue than those with lower proficiency levels.

Finally, this research aims to provide pedagogical insights into L2 English pronunciation instruction in the Korean context. If contextual variation significantly affects rhythm production, the findings could lead to more structured teaching strategies, including utilizing Artificial Intelligence (AI) speech resources to improve pronunciation training.

2. Theoretical Backgrounds

2.1. English and Korean rhythm and rhythm metrics

Rhythm in language refers to the phonetic features that create the temporal pattern of speech. It includes elements such as stress, syllable length, and speech rate, which together shape how speakers organize their speech for effective communication and how listeners understand messages efficiently (Arvaniti, 2009; Cutler, 1984). English is generally considered a stress-timed language, where stressed syllables tend to occur at regular intervals, while unstressed syllables are shortened and less prominent. Conversely, Korean tends to produce each syllable with relatively uniform duration, showing minimal variation in syllable length (Kim et al., 2007). This fundamental difference in rhythm structures can make it challenging for Korean learners to acquire the stress-timed rhythm of English, despite its importance in effective communication, especially in English discourse.

Specifically, Ordin & Polyanskaya (2015) measured L2 English stress-timing using rhythm metrics of German speakers whose native language shares the same rhythm typology as English, and French speakers whose native language has a different rhythmic pattern from English. They evaluated the durational variabilities of syllables, vowels, and consonants. The results showed that both groups approached stress-timing as their English proficiency increased, but the French speakers exhibited less stress-timed speech than the German speakers among advanced-level learners of English. The findings suggest that their native language rhythm may limit the ultimate achievement of rhythmic patterns in second language acuiqsition.

Research also indicates that language rhythm exists on a continuum rather than as distinct categories, and efforts have been made to quantify rhythmic variation across languages (Grabe & Low, 2002; White & Mattys, 2007). Several rhythm metrics have been proposed, including %V, ΔV, and ΔC (Ramus et al., 1999), as well as the Pairwise Variability Index (PVI) and its normalized version (nPVI) (Grabe & Low, 2002). The nPVI measures the normalized average difference in duration between adjacent sound units such as vowels, consonants, or syllables. Higher nPVI values reflect greater durational variability, typical of stress-timed languages, while lower values are associated with syllable-timed patterns. These rhythm metrics are recognized and used in many empirical studies, although some point out that they cannot fully capture the nature of rhythm (Arvaniti, 2009; Gibbon, 2003). Nonetheless, rhythm metrics at least quantify aspects of speech rhythm, which has led to their widespread use in second-language speech rhythm research (Choe, 2022; Jang, 2008; Kim, 2016).

2.2. Second language English rhythm

English rhythm plays a crucial role in effective and accurate communication. It closely relates to the information structure, emphasis strategies, and discourse design that take into account listeners' background knowledge and message understanding (Birch & Clifton, 1995; Bock & Mazzella, 1983; Cutler, 1976, 1984; Cutler & Fodor, 1979; Cutler & Foss, 1977). Therefore, it goes beyond just timing patterns (Arvaniti, 2009). Despite its importance in communication, English rhythm is considered notoriously difficult to acquire in English as a Foreign Language (EFL) settings. Jang (2008) quantitatively examined the English rhythm of Korean learners and discovered that some of the rhythm metrics correlated with the learners' overall English proficiency.

English rhythms produced by Korean learners have been consistently studied using rhythm metrics. In a large-scale study by Kim et al. (2007), controlled read speech samples from 111 Korean learners of English were analyzed using various speech rhythm indices. They specifically examined the pairwise variability between stressed and unstressed vowels and compared it to that of native speakers. Additionally, the instructional effect was tested, and the results showed positive improvements. Xue & Chung (2021) examined the English rhythm of Korean and Chinese speakers by measuring vowel durations and calculating nPVI. The 10 Korean participants learning English showed relatively lower nPVI values than native speakers and Chinese learners, with Koreans scoring about 58 in their English speech and 50 in their native Korean speech, while native English speakers averaged about 70.

Choe (2019) examined Busan Korean speakers' speech using %V, ΔC, ΔV, VarcoC, VarcoV, and PVIs and compared them to those of native English speakers. Notably, some participants exhibited more stress-timed patterns than native speakers, which was interpreted as the interaction with the participants' tonal dialect. Moreover, rhythm metrics that were not normalized for speech rate showed greater variability among learners than among native speakers, suggesting that normalized metrics are more reliable. Choe (2022) examined the effect of pronunciation instruction on the improvement of learners' rhythm in terms of rhythm metrics. The results showed that learners exhibited greater improvements in consonant-based variabilities than in vowel-based variabilities, indicating that stress-timing regarding vowel reduction is more difficult for learners to acquire.

In second language pronunciation research, key constructs include nativelikeness (or accentedness), intelligibility, and comprehensibility (Derwing & Munro, 2005; Levis, 2005). Studies have shown that these constructs are partially independent of each other. Among these, rhythm—covering timing and nuclear accents—is a common and important factor across all constructs (Kang, 2010; Yang, 2021). Although rhythm metrics have not yet been actively studied in relation to L2 pronunciation constructs, prosodic features such as rhythm and intonation have consistently been shown to affect perceived pronunciation nativelikeness (Kang, 2010; Trofimovich & Isaacs, 2012). In the Korean context, Kim (2021) grouped 52 Korean learners of English into three groups based on their accentedness, such as no accent, weak accent, and strong accent, and compared them in terms of vowel-based variabilities, such as %V, VarcoV, nPVI-V, showing significant differences among the groups.

2.3. Speech contexts and L2 rhythm adaptation

It has been argued that spoken utterances are not just the result of a speaker's internal linguistic knowledge. Instead, they are actively influenced by interactions with the listener, depending on the pragmatic and social context in which the utterance occurs (Pickering & Garrod, 2004). Dialogic and monologic situations affect speech production differently at various levels of linguistic representation (Pickering & Garrod, 2004). In conversations, the presence of a listener often causes speakers to make adaptive changes in their speech. This process is commonly called phonetic entrainment or phonetic convergence, where a speaker's pronunciation becomes more similar to that of the listener during interaction. In second language learning, Lewandowski (2012) studied phonetic convergence between native and nonnative speakers, considering individual differences in phonetic talent, by measuring the amplitude envelope of speech.

Chung (2013b) specifically examined the possibility of phonetic accommodation among speakers. Chung (2013b) studied rhythm produced by Korean speakers of English in London during spontaneous conversations with individuals from different language backgrounds, including native English speakers, Korean speakers, and other non-native speakers. The study aimed to determine whether rhythmic convergence and accommodation occur. It assessed rhythm using a variability index based on syllable durations and speech rate by the number of syllables per second.

The analysis showed no significant evidence of convergence and accommodation related to the interlocutor group.

Kim (2016) compared read speech, which is an excerpt from *The Sun and the Wind* from Aesop's fables, and spontaneous speech, such as self-introductions, in terms of nPVI among Korean, Vietnamese, and Chinese learners of English. She confirmed the transfer of native rhythmic patterns to English speech, and learners with higher proficiency scores on certified English tests showed higher nPVI values, indicating more stress-timed speech. However, the two types of speech did not show significant differences in nPVI values.

Although it did not include phonetic measurements, Trofimovich & Kennedy (2014) confirmed that phonetic alignment occurs at various levels of perception of language use during conversations between bilingual speakers. For individual interlocutors, native speaker evaluators also assessed comprehensibility, accent, fluency, and other factors. Although it was a global rating, the evaluators reported that the speakers' pronunciations sounded more similar as the task progressed. In the study by Trofimovich et al. (2014), it was also observed that second-language speakers aligned their stress patterns with those of other speakers during the interactive practice of multisyllabic English words in classroom tasks.

Following these efforts, the current study extends to compare dialogue reading and monologue reading. This has importance, especially in educational settings, as it contrasts speech timing between traditional monologue reading, long used in pronunciation classes, and controlled, educational turn-taking with native-like speech that has become easier thanks to recent technological advancements.

Recent advances in AI have led to an increase in studies exploring the use of AI interlocutors in language classrooms (e.g., Park & Lee, 2022; Yoo & Ahn, 2024, among others). In response to this emerging technology-enhanced language learning environment, the present study also aims to provide a theoretical foundation by investigating how learners adjust their speech rhythm across dialogue and monologue settings. For example, consistent exposure to a native model's rhythm—especially in interactional dialogue reading or cooperative speech tasks—may promote rhythmic adaptation and lead to the acquisition of more target language-like prosody. From this perspective, pronunciation instruction, including rhythm, should go beyond traditional monologic reading or drills. Instead, it should incorporate interactive practices such as dialogue-based reading, shadowing, and AI-mediated conversations. More importantly, such pedagogical approaches should be grounded in theoretically supported and empirically validated frameworks. In the same vein, the current study seeks to examine how speech rhythm varies depending on the reading context (dialogue versus monologue) and to explore the pedagogical and theoretical implications of these differences.

3. Methodology

3.1. Recording

Thirty-five Korean female students from a university participated in the recording. Only female participants were recruited to avoid confounding variables related to gender effects. They were undergraduate and graduate students studying various majors, including education, humanities, sciences, social sciences, engineering, and others. More than half of the participants, though not all, had official English test scores such as TOEIC, TOEFL, TEPS, and OPIc, with scores over 100 on IBT TOEFL, over 900 on TOEIC, 600-900 on TEPS, and IH (Intermediate high) on OPIc, indicating their overall English proficiency was intermediate to advanced. The participants had a wide range of experience in English-speaking countries, which was intentional because we needed to ensure a broad range of pronunciation proficiencies. Therefore, the participants had varied lengths of residence in English-speaking countries, ranging from those who had been bilingual from a very early age to those with no experience at all.

Each participant read two types of texts in a soundproof recording room, and their voices were recorded using a SONY PCM-M10 recorder. The recording materials included a short dialogue with four rounds of turn-taking. The content focused on everyday conversational topics. To compare rhythm realization under the two conditions, two additional decontextualized utterances from the same participants were included in the analysis. The utterances analyzed in this study include the following (for the entire dialogue, see Appendix).

Dialogue (Hahn & Dickerson, 1999, p. 11)

Utterance 1: I'm really in a rush.

Utterance 2: Let's see. Maybe you could help me finish these graphs.

Utterance 3: Today I'm especially busy. This morning, I'm finishing a review of an article, and chairing a staff-meeting. Then I need to get ready to present a seminar.

Utterance 4: You'll never believe it. It's "how to manage your time."

Monologue

Utterance 5: This room has a harbor view whereas that room has an obstructed view.

Utterance 6: He was staying close behind me.

In the dialogue reading condition, participants engaged in a role-play activity, taking turns reading lines from a scripted conversation with a male native speaker. If a turn included multiple sentences, such as utterance 3, it was considered a single utterance. That is, they contributed one nPVI. The imbalance in the number of sentences and words between the recording materials in the two conditions remains a limitation of the study. This study primarily investigates rhythm realization using nPVI in the dialogue reading context, with the monologue condition serving as a reference for comparison. Also, utterances of varying lengths were included, and they are discussed in more detail in section 4.3.2 through a readability score to compare speech timing from multiple perspectives, such as the influence of utterance length.

3.2. Rating

As mentioned earlier, L2 pronunciation proficiency differs from other aspects of L2 skills in several ways (Kennedy & Trofimovich, 2017). Therefore, the nativelikeness of the participants' pronunciation was rated based on their production of another English script from the same pronunciation textbook (Hahn & Dickerson, 1999). Each speech sample from the participants was evaluated for pronunciation nativelikeness (i.e., degree of foreign-accentedness) by 11 native speakers of North American English. The listeners heard dialogues read by the male interlocutor and participants, focusing on the speech from female participants. The evaluations used a 7-point Likert scale and included a brief description explaining the definition of foreign-accentedness. All rating sessions were conducted individually in the presence of the researcher to ensure consistency and address any procedural questions from the raters. To assess the inter-rater reliability of the pronunciation ratings, intraclass correlation coefficient analysis (average measures) was used, and the results indicated high agreement (ICC[C, 11]=.959).

Participants were divided into three groups: High, Mid, and Low, based on their rating results. Because the pronunciation score differences among participants were gradual, the levels were assigned solely based on having a balanced number of participants in each group. The group information is as follows.

Table 1. Pronunciation rating results and group formation

Group	N	Mean	S.D.	Min	Max
High	12	5.7	0.7	4.6	6.8
Mid	12	4.1	0.3	3.7	4.6
Low	11	3.1	0.3	2.6	3.6

3.3. Speech rhythm measurement

To compare speech rhythm across two reading conditions, this study used nPVI based on vowel durations, a measure commonly used in previous L1 and L2 rhythm research. Research also reported that rhythm metrics without normalization for speech rate produced inconsistent results (Choe, 2019; Jang, 2008). Therefore, this study used the vocalic nPVI to ensure comparability and clarity in speech rhythm across conditions. The formula for the nPVI used in this study is shown below (Grabe & Low, 2002):

$$\text{nPVI} = 100\left[\sum_{k=1}^{m-1}\left|\frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2}\right| / (m-1)\right]$$

Note. d_k and d_{k+1} are the durations of consecutive units; m is the number of intervals. Higher nPVI values correspond to more stress-timed rhythms.

The recorded speech samples were initially annotated using the Montreal Forced Aligner (MFA; McAuliffe et al., 2017), and then each speech sample and annotation in TextGrid files were carefully examined, correcting errors with Praat speech analysis software (Boersma & Weenink, 2025). The resulting boundary markings and errors were manually refined through both auditory inspection and visual analysis of the waveform and spectrogram. Common corrections included adding unrecognized words like “charing” (which are not found in the MFA English dictionary), adjusting boundaries, and inserting a retroflex /r/ in the post-vocalic position. Several participants produced speech errors, such as repeating parts of sounds or words from the scripts, mispronouncing words or sounds, and omitting sounds. All these speech errors were manually annotated and incorporated into the analysis.

A notable speech characteristic of Korean learners of English was the insertion of epenthetic vowels after coda sounds, marked as "epen" in the TextGrid files. First, both speech errors and epenthetic vowels were included in the analysis, but epenthetic vowels unexpectedly resulted in much higher nPVI values because of their very short duration, indicating that more epenthetic vowels lead to higher nPVI (i.e., more stress-timed). This conflicts with our understanding of stress timing, so the epenthetic vowels were excluded when calculating nPVI. Additionally, individual variations in how some words were produced were noted as they appeared. For example, "you'll" was pronounced either as "you'll" or as "you will." These differences caused the number of vowels produced by participants to vary.

Using these manually refined annotations, vowel durations were measured with Praat scripts. These measurements were then used to calculate vocalic nPVI values for each utterance—four utterances per speaker in dialogue and two in monologue conditions, across all 35 speakers (a total of 210 nPVI values).

3.4. Statistical testing

Statistical analysis was performed in *R* (version 4.4.2), beginning with descriptive statistics to summarize the overall distribution of rhythm measures. A repeated measures two-way ANOVA was then conducted to evaluate the main effects of reading context and pronunciation proficiency, as well as their interaction. For the pronunciation proficiency rating, the reliability of the 11 raters was assessed using intraclass correlation coefficients based on average measures.

4. Results and Discussion

4.1. Descriptive statistics

To provide an overview of the dataset and key variables, descriptive statistics were calculated for the entire dataset, including all participants and conditions. Table 2 summarizes the descriptive statistics for the two main variables: 210 nPVI (35 participants * 6 utterances [four utterances in a dialogue and two utterances in a monologue reading setting]) and 35 pronunciation nativelikeness.

Table 2. Descriptive statistics for nPVI and nativelikeness ratings

	N	Mean	Min	Max	25%	50%	75%
nPVI	210	57.7	29.2	91.6	49.5	57.5	65.4
Nativelikeness	35	4.3	2.5	6.8	3.4	4.1	5.5

The average nPVI value across all samples was 57.7, with values ranging from 29.2 to 91.6. This large difference between the minimum and maximum values indicates the speech samples in the study include the full range of speeches, from very syllable-timed speech to very stress-timed speech. For pronunciation nativelikeness, the mean rating on the 7-point Likert scale was 4.3. To allow for a more detailed comparison of rhythmic variation across conditions, descriptive statistics for nPVI were further calculated separately for the two reading contexts—dialogue and monologue. The statistics are summarized in Table 3.

Table 3. Descriptive statistics by reading condition and utterance

Setting	Utterance	N	nPVI (Mean)	S.D.	Min	Max	Vowel count (Mean)
Dialogue	1	35	55.4	12.8	33.1	83.6	7.06
	2	35	65.2	11.9	38.6	90.2	12.5
	3	35	54.2	5.9	40.8	64.8	45.7
	4	35	62.1	8.7	39.0	76.1	13.1
	Total	140	59.2	11.1	33.1	90.2	
Monologue	5	35	50.0	8.7	29.2	63.8	17.8
	6	35	59.4	11.8	44.3	91.6	7.97
	Total	70	54.7	11.3	29.2	91.6	

The dataset contains 140 observations for the dialogue reading condition and 70 observations for the monologue condition. The average nPVI value was 59.2 (*S.D.*=11.1) in the dialogue setting, and 54.7 (*S.D.*=11.3) in the monologue setting. Overall, the descriptive results show that nPVI values were generally higher in the dialogue condition, indicating a more stress-timed rhythmic pattern. Additionally, nPVI values for each utterance are presented, and high variability among the utterances is noted. In dialogue reading, the nPVI ranged from 54.2 to 65.2, while in monologue reading, it ranged from 50.0 to 59.4.

To better understand the relationship between nPVI values and the reading context regarding the participants' pronunciation nativelikeness, the data were visualized using scatter plots. Figure 1 displays a comprehensive scatter plot of dialogue and monologue reading, respectively. Participants are ordered along the x-axis from lowest to highest nativelikeness score. The y-axis shows the corresponding nPVI values, with higher values indicating a more stress-timed rhythm. Since each participant produced four and two utterances per condition, the same number of nPVI values is plotted for each individual. When two or more participants received the same pronunciation nativelikeness score, their data points are aligned vertically.

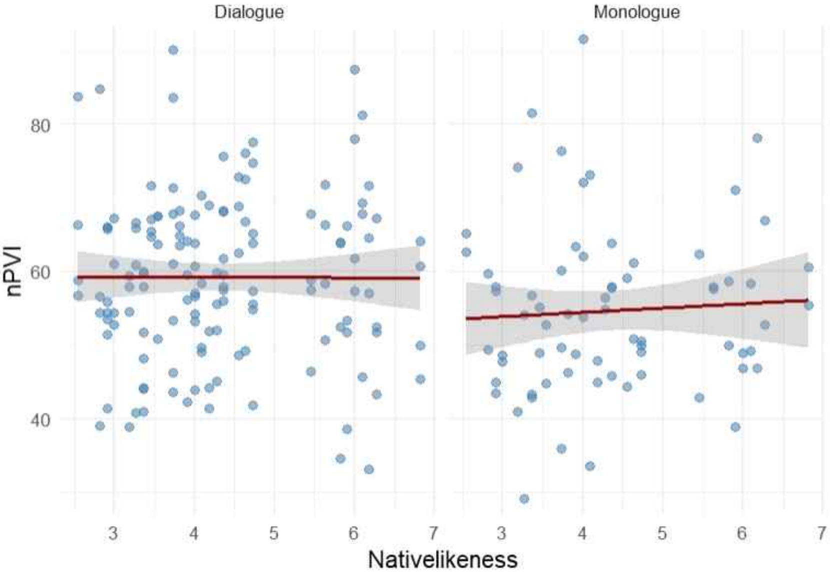
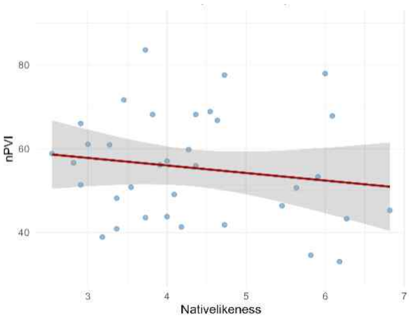
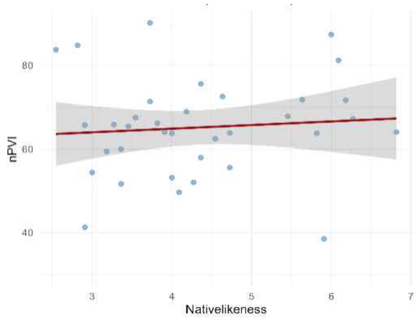


Figure 1. Scatter plot of nPVI values by reading condition and pronunciation nativelikeness.

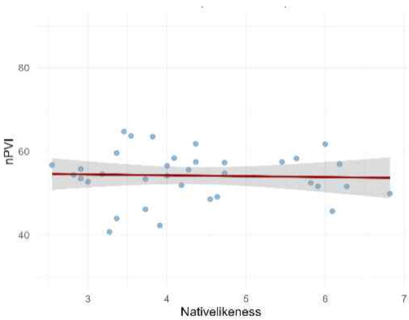
Figure 1 shows that, overall, in a monologue setting, as pronunciation nativelikeness increases, participants tend to produce a more stress-timed speech rhythm, whereas no such tendency is observed in a dialogue setting, as shown by a fairly level regression line. However, the difference in slope is not very large. To further clarify the influence of reading context, Figure 2 separates the scatter plots by utterance.



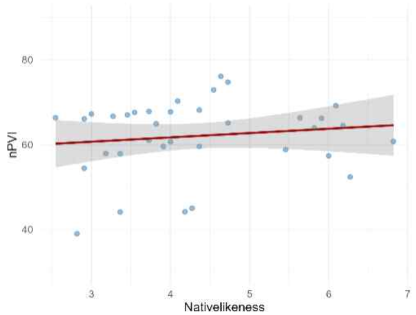
Dialogue: Utterance 1



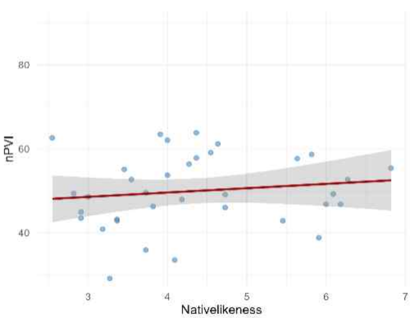
Dialogue: Utterance 2



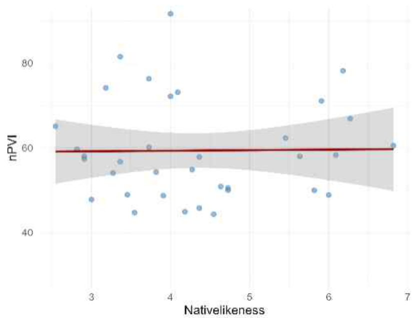
Dialogue: Utterance 3



Dialogue: Utterance 4



Monologue: Utterance 5



Monologue: Utterance 6

Figure 2. Scatter plot of nPVI values by each utterance and pronunciation nativelikeness.

The comparison shows similar patterns among the utterances, with a leveled or slightly rising pattern based on pronunciation nativelikeness, except for utterance 1. In utterances 2, 4, 5, and 6, higher pronunciation nativelikeness generally correlates with slightly higher nPVI values, indicating a shift toward a stress-timed rhythm, but the change is not dramatic. However, utterances 1 and 3, especially utterance 1, “I’m really in a rush,” display the opposite pattern, prompting further discussion.

Table 4 below presents the descriptive statistics for the three pronunciation proficiency groups under the two reading conditions.

Table 4. Descriptive statistics by reading condition and group

Setting	Group	Nativelikeness (Mean)	N	nPVI (Mean)	S.D.	Min	Max
Dialogue	High	5.7	12	59.7	12.2	33.1	87.4
	Mid	4.1	12	59.7	10.6	41.4	90.2
	Low	3.1	11	58.2	10.5	39.0	84.8
Monologue	High	5.7	12	54.6	9.1	38.8	78.2
	Mid	4.1	12	56.4	13.1	33.6	91.6
	Low	3.1	11	52.8	11.6	29.2	81.5

nPVI values for the three groups under the dialogue reading condition show very similar mean nPVI values across the groups, while those under the monologue reading condition generally show lower mean values, the mid-level group showing the highest (i.e., most stress-timed). Figure 3 shows boxplots comparing the three groups across the two conditions.

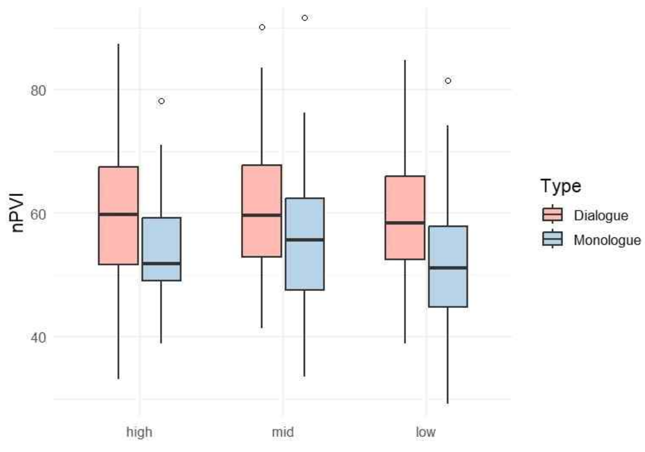


Figure 3. Box plot of nPVIs by reading condition and group.

The box plot clearly shows that participants had different timing patterns in dialogue and monologue settings. The median for the low proficiency group is slightly lower than for the other two groups, but they show more comparable timing to higher-level groups in the dialogue setting than in the monologue setting. The mid-level group has the highest median in the monologue setting, and the high proficiency group exhibits the largest median difference between the two conditions.

4.2. Statistical results

To assess reading context and pronunciation proficiency, a repeated measures two-way ANOVA was conducted with nPVI as the dependent variable, reading context as a within-subject factor, and group as a between-subjects factor. Figure 4 shows the interaction plot for the two main effects.

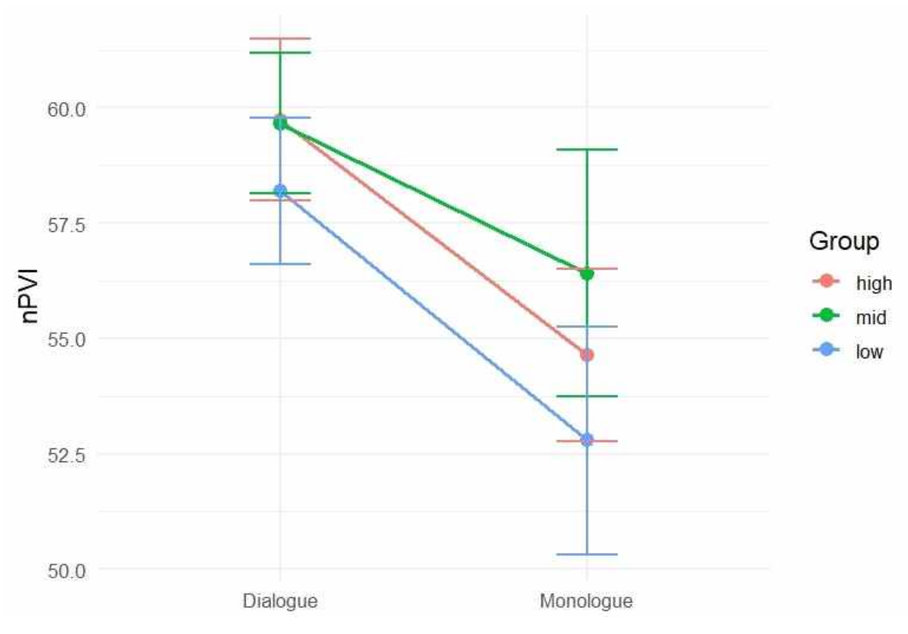


Figure 4. Interaction between reading condition and group.

The three groups showed non-parallel differences between dialogue and monologue reading conditions. The statistical analysis indicated a significant main effect of setting (reading condition) on nPVI values ($F(1, 32)=8.61$, $p<.001$). The main effect of group and the interaction effect of setting and group were not significant.

Table 5. Repeated measures two-way ANOVA on nPVI

Effect	*F*	*Df* (Num, Den)	*p*
Setting	8.61^{**}	(1, 32)	<.001
Group	1.11	(2, 32)	.343
Setting*Group	0.19	(2, 32)	.828

4.3. Discussion

4.3.1. Pronunciation in the different reading contexts and pedagogical implications

This study investigated whether Korean learners' production of English rhythm varies depending on the reading context—dialogue versus monologue. The results revealed that the mean nPVI value was higher during dialogue than monologue, indicating that participants produced more stress-timed rhythm when reading the dialogue with a native speaker. Additionally, in the dialogue context, the mean nPVI values remained consistent across the three groups with different pronunciation proficiency levels. This suggests that practicing dialogue with a speaker who produces more stress-timed speech may encourage learners, particularly those with lower pronunciation proficiency, to produce speech that is more English-like and rhythmically consistent. These differences were statistically significant. The findings imply that the interactional context can help regulate or stabilize rhythm variability among learners, thereby minimizing the influence of individual pronunciation differences and facilitating the acquisition of target-language rhythm.

Overall, the first research question about the effect of reading condition was supported, confirming that speech rhythm interacts dynamically with pragmatic and social context. This pattern suggests that learners with lower pronunciation nativelikeness show more significant shifts toward stress-timed rhythm in the dialogue condition. In other words, engaging with a more advanced learner may help less proficient learners better stabilize and align their speech rhythm, potentially compensating for their lower proficiency.

Based on the findings of the study, the following pedagogical implications can be discussed regarding (1) second language pronunciation acquisition and (2) instructional design.

First, speech production circumstances may affect the production of rhythm, ultimately affecting the acquisition of language rhythm. It is taken for granted that different tasks differently affect the acquisition of linguistic features in various ways, but this is rarely demonstrated in terms of pronunciation, especially the prosodic aspects of L2. As van Lier (1996) conceptualized learning as a socially situated process, he emphasized that social interaction plays a crucial role in language acquisition. In line with this view, pronunciation instruction should engage learners in practicing pronunciation through interaction rather than in isolation. The present study, in part, supports the idea that L2 learners actively attend to and internalize specific linguistic features during meaningful interaction, rather than passively absorbing what is explicitly taught, as also proposed by van Lier (1996), in that they revealed different patterns of a pronunciation feature while reading with another person. Therefore, engaging learners in any form of interaction or dialogue reading can change their language output. This, in turn, may influence second language acquisition processes. While second language research has often focused on other language components such as grammar and vocabulary, pronunciation has received little attention from this interactional perspective (Kim & Michel, 2024). However, the present study goes beyond the observational limits of previous methods.

Second, the finding that the dialogue condition improves the consistency of rhythm production suggests that, in language classrooms, dialogue reading or conversational tasks—whether conducted with a human interlocutor or through AI-based interlocutors—may positively contribute to the acquisition of natural rhythm. Trofimovich et al. (2014) emphasize that linguistic alignment among learners can occur through conversational activities in the second language classroom. They advise teachers that, through meaning-focused activities, imitation and repetition take place, leading to natural pronunciation learning.

For Korean learners to acquire English rhythm, instructional contexts that support rhythm alignment should be prioritized, even in limited conversational settings such as those used in this study. Additionally, using controlled dialogue reading tasks, as shown in this study, can provide further benefits for pronunciation development. Because such tasks reduce variability in grammar and vocabulary, they enable learners to focus more directly on pronunciation aspects such as rhythm, thereby enhancing their potential for improvement.

4.3.2. nPVI and L2 speech characteristics

The difference in nPVI between the two conditions did not vary among the three groups with different levels of pronunciation proficiency. Participants produced significantly higher nPVI in dialogue than in a monologue setting, but the pronunciation proficiency level used in this study did not lead to differences in the degree of stress timing or adaptation of timing. Therefore, the second research question was not confirmed in this study, which prompts further discussion on the characteristics of L2 speech, the use of rhythm metrics for L2 speech, and the nature of the reading materials of the study.

First of all, this result can be attributed to the characteristics of L2 speech. When analyzing L2 English speech and calculating nPVI, many issues emerged, raising questions about the validity of vocalic nPVI for the L2 speech model. Above all, frequent speech errors made by learners inevitably affected, and more precisely distorted, the nPVI values. Several participants in this study produced epenthetic vowels after a voiced or voiceless consonant at the end of words, such as "could," "need," and "has." The duration of these epenthetic vowels was generally very short compared to the stressed full vowels. These features of epenthetic vowels resulted in higher nPVI values. In other words, the more epenthetic vowels the participants produced, the higher the nPVI values became. This clearly contradicts our understanding of English speech rhythm, leading to the decision to exclude epenthetic vowels from the analysis. Nevertheless, learners with lower pronunciation proficiency produced more frequent speech errors, such as substitution, deletion, additions, and repetition of parts or entire words and phrases. The vowel durations in these errors could be very short or very long. The current study included these speech errors, except epenthetic vowels, in the analysis, considering them as natural aspects of L2 speech; however, this remains a limitation of the research.

Another aspect worth noting about L2 speech characteristics is that Utterance 1 showed a negative trend in stress-timing based on pronunciation proficiency. First, the length of the utterance and the familiarity with the text may have influenced the rhythm production. For example, participants produced Utterance 1, "I'm really in a rush," very accurately and fluently, without errors and with confidence. Additionally, individual differences, such as the tendency to engage in conversational contexts—which are not explained by the term proficiency—seemed to influence this. Some participants truly enjoyed reading with another interlocutor and tried to naturally perform their role in the dialogue. It's also important to note that the construct of pronunciation proficiency used as a reference in this study cannot be fully explained by nPVI alone. Perceived pronunciation proficiency involves a complex relationship among various phonetic properties (Kang, 2010; Trofimovich & Isaacs, 2012). Utterance 1 in this study, which shows a reversed timing trend in pronunciation proficiency, also exhibits considerable variation among speakers with similar proficiency levels, raising questions about the interpretive significance of this trend. Another possibility is that Utterance 1 is closer to a frequently used chunk expression, and learners are less likely to make errors when producing it. In a similar context, participants could also produce it very fast. Chung (2013b) observed that as Korean English speakers' speech rate increases, they tend to adopt a syllable-based rhythm rather than a stress-timed rhythm. The participants in this study also exhibited more syllable-based characteristics as their pronunciation proficiency increased, and this may have resulted from their faster speech.

Next, nPVI may not accurately represent the actual rhythm. According to Gibbon (2003), while nPVI is a useful metric, it relies solely on the relative changes in the lengths of neighboring elements and cannot fully capture the overall structure of linguistic rhythm or perceptual rhythmicity. For example, Gibbon (2003) points out the limitation that nPVI assigns the same value to a rhythm that alternates regularly with 2-4-2-4 and a rhythm that increases exponentially with 2-4-8-16. In Jang (2008), nPVI values of sentences with six to eight words showed significant differences among different proficiency levels, while another metric, such as Above-word-level measures, was effective for all sentence lengths from four to 13 words. In this regard, future research should systematically explore the relationship between sentence length and timing patterns.

Finally, to investigate possible systematic variations related to the experimental text, the Readability test score can be considered. In this study, the utterances included different texts presented in conversational and monologue conditions. To support a detailed discussion of the results, the Flesch Readability Test Score (Flesch, 1948) was used for each utterance. The Flesch readability test (Flesch Reading Ease) assesses a text's readability based on sentence length and word complexity. The Flesch Reading Ease score is calculated using the following formula:

$$206.835 - 1.015\left(\frac{total\ words}{total\ sentences}\right) - 84.6\left(\frac{total\ syllables}{total\ words}\right)$$

This score decreases as sentences become longer and as the number of polysyllabic words increases. A score of 90-100 is considered Very easy, while 0-30 is Very difficult. Scores can go above 100, but any score over 100 is displayed as 100. The data from this study resulted in an overall score of 80.4, which is classified as Easy. The calculated scores for each utterance were: Utterance 1 – 100, Utterance 2 – 100, Utterance 3 – 56.7, Utterance 4 – 91.7, Utterance 5 – 83, and Utterance 6 – 87.9. All except Utterance 3 were rated at the Easy or Very Easy level. Utterance 1, which showed a reversed trend line, scored 100 with high readability, while Utterance 3 had the lowest readability. Overall, it does not seem easy to determine that the readability level of the text had a consistent impact on the participants' timing production.

4.3.3. Limitations of the study and directions for future research

The findings of this study are limited to controlled reading tasks and cannot be generalized to spontaneous speech and authentic conversational situations. Additionally, the sample size—35 female participants—was insufficient to examine the potential influence of individual factors such as gender or the amount of English exposure. Future research should involve learner groups with varying levels of English proficiency and more natural interaction settings that incorporate spontaneous speech. More importantly, instructional design should be tested in actual classrooms through long-term studies. Specifically, future research should determine the optimal length of rhythm-based training required to internalize the native-like rhythm of English.

To better understand how reading context influences speech, it is crucial to control for language factors beyond the reading environment, such as vocabulary and structure. This is a limitation of the study, and future research should address it. When the same participants read the same sentences multiple times, familiarity with the text might affect the timing and rate of speech. Designing a more careful experiment would help address this problem.

Imbalanced sampling is also a limitation of this study and may have affected the statistical power of the tests. Since groups with larger sample sizes tend to have a greater influence on mean estimates, the dialogue setting, which has more observations, may disproportionately shape the overall patterns. At the same time, the statistical power to detect the positive relationship between pronunciation proficiency and stress-timing may be lower in the monologue setting, where this relationship is more clearly observed, beacause of its smaller sample size. These methodological limitations restrict the interpretation of the results. They do not invalidate the observed trends, but when the sample size is small, outliers can have a more pronounced impact. Future research should definitely address and improve these issues by using more balanced designs and larger samples.

Recent advances in speech technology, especially AI-based systems, are increasingly being used in L2 speaking and listening classes. More importantly, the comparison between interaction-based language learning and traditional teaching methods has provided limited empirical evidence, particularly on how specific language features appear differently. The present study employed a human interlocutor, and therefore, its findings cannot be directly applied to learning situations involving AI interlocutors. Human speakers are known to align with their interlocutors at multiple linguistic levels and are often emotionally motivated during interaction, which can influence mutual speech production and perception. It remains uncertain whether AI-generated speech can adequately adapt to the speech characteristics of human learners or whether learners are sufficiently motivated to engage in sustained verbal interaction with AI systems to facilitate acquisition.

5. Conclusion

This study examined L2 rhythm production among Korean learners of English, using nPVI as a speech timing measure. The research included two reading situations: one with a native speaker interlocutor and the other without such interaction. Results showed that learners produced more stress-timed speech when reading a dialogue with a native speaker compared to reading a monologue alone, strongly indicating that reading context influences how English rhythm is realized in L2 learners. These findings highlight the importance of rhythm instruction through dialogue-based tasks. Incorporating AI-driven speech tools could further aid in internalizing stress-timed patterns, provided that such approaches are supported by research-backed theories. Longitudinal studies could offer more detailed and formative insights into how the frequency and duration of interaction influence pronunciation development. Such research would contribute to moving L2 pronunciation instruction and research beyond conventional approaches, toward a more nuanced understanding of the context of pronunciation development.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*(1-2), 46-63.

Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, *38*(4), 365-391.

Bock, J. K., & Mazzella, J. R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory and Cognition*, *11*(1), 64-76.

Boersma, P., & Weenink, D. (2025). Praat: Doing phonetics by computer (version 6.4.39) [Computer program]. Retrieved from https://www.praat.org/

Choe, W. K. (2019). The realization of English rhythm by Busan Korean speakers. *Phonetics and Speech Sciences*, *11*(4), 81-87.

Choe, W. K. (2022). The effect of pronunciation teaching on the realization of English rhythm by Korean learners of English. *Phonetics and Speech Sciences*, *14*(2), 19-28.

Chung, H. (2013a). Phonological accommodation and convergence of English consonants in interlanguage talk by Korean learners of English in the UK. *Language and Linguistics*, *59*, 241-266.

Chung, H. (2013b). A study on the rhythm of Korean English learners' interlanguage talk. *Phonetics and Speech Sciences*, *5*(3), 3-10.

Cutler, A. (1976). Beyond parsing and lexical look-up. In R. J. Wales & E. C. T. Walkers (Eds.), *New approaches to language mechanisms: A collection of psycholinguistic studies* (pp. 133-149). Amsterdam: North Holland.

Cutler, A. (1984). Stress and accent in language production and understanding. In D. Gibbon & H. Ritcher (Eds.), *Intonation, accent and rhythm: Studies in discourse phonology* (Vol. 8, pp. 76-90). New York: Walter de Gruyter.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*(1), 49-59.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, *20*(1), 1-10.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research based approach. *TESOL Quarterly*, *39*(3), 379-397.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221-233.

Gibbon, D. (2003, August). Computational modelling of rhythm as alternation, iteration and hierarchy. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2489-2492). Barcelona, Spain.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 515-546). Berlin: Mouton de Gruyter.

Gut, U. (2003, August). Non-native speech rhythm in German. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2437-2440). Barcelona, Spain.

Hahn, L. D., & Dickerson, W. B. (1999). *Speech craft: Discourse pronunciation for advanced learners*. Ann Arbor: The University of Michigan Press.

Jang, T. Y. (2008). Speech rhythm metrics for automatic scoring of English speech by Korean EFL learners. *Malsori*, *1*(66), 41-59.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, *38*(2), 301-315.

Kennedy, S., & Trofimovich, P. (2017). Pronunciation acquisition. In M. Sato & S. Loewen (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 260-279). New York: Routledge.

Kim, J. M., Flynn, S., & Oh, M. (2007). Non-native speech rhythm: A large-scale study of English pronunciation by Korean learners: A large-scale study of English pronunciation by Korean learners. *Studies in Phonetics, Phonology and Morphology*, *13*(2), 219-250.

Kim, S.-A. (2021). Foreign accentedness rating and rhythm in L2 English. *The Mirae Journal of English Language and Literature*, *26*(4), 101-120.

Kim, S.-A. (2016). Quantitative approach to L2 rhythm: With specific references to Korean, Vietnamese and Chinese learners of English. *Studies in Linguistics*, *39*, 49-65.

Kim, Y., & Michel, M. (2023). Linguistic alignment in second language acquisition: A methodological review. *System*, *115*, 103007.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369-377.

Lewandowski, N. (2012). *Talent in foreign language learning: The phonetic talent hypothesis from an experimental phonetic perspective* (Doctoral dissertation). Humboldt University of Berlin, Berlin.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017, August). Montreal forced aligner: Trainable text-speech alignment using Kaldi. In

Interspeech 2017 (pp. 498-502). Stockholm, Sweden.

Ordin, M., & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America*, *138*(2), 533-544.

Park, M., & Lee, J. (2022). A study on the comparison of AI Pengtalk's and English native speakers' assessment of elementary students' English pronunciation. *Journal of the Korean English Education Society*, *22*(4), 67-90.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169-190.

R Core Team. (2024). R: A language and environment for statistical computing (version 4.4.2) [Computer software]. Retrieved from https://www.R-project.org.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265-292.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(4), 905-916.

Trofimovich, P., & Kennedy, S. (2014). Interactive alignment between bilingual interlocutors: Evidence from two information-exchange tasks. *Bilingualism: Language and Cognition*, *17*(4), 822-836.

Trofimovich, P., McDonough, K., & Foote, J. A. (2014). Interactive alignment of multisyllabic stress patterns in a second language classroom. *Tesol Quarterly*, *48*(4), 815-832.

Trofimovich, P., McDonough, K., & Kennedy, S. (2014, August). Interactive alignment: What's in it for language teachers? *TESL Ontario Contact Magazine*, 15-21.

Van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy and authenticity*. London: Longman.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First and second language studies. *Journal of Phonetics*, *35*(4), 501-522.

Xue, W., & Chung, H. (2021). A study of the rhythm of native, Chinese, and Korean speakers of English in read speech. *Journal of the Korean English Education Society*, *20*(3), 29-47.

Yang, I. Y. (2014). Acoustic cues in Korean high school students' L2 phoneme perception and speech segmentation. *Korean Journal of Applied Linguistics*, *30*(4), 235-264.

Yang, I. Y. (2017). Instructional effects on voice onset time perception in L2 speech segmentation. *Language Research*, *53*(3), 561-585.

Yang, I. Y. (2021). Differential contribution of English suprasegmentals to L2 foreign-accentedness and speech comprehensibility: Implications for teaching EFL pronunciation, speaking, and listening. *Korean Journal of English Language and Linguistics*, *21*, 818-836.

Yoo, S., & Ahn, H. (2024). The effects of prosody training with AI chatbot on the English pronunciation improvement of Korean EFL learners. *Korean Journal of English Language and Linguistics*, *24*, 1300-1317.

In Young Yang
Associate Professor
Department of English Language and Literature
Korea National University of Transportation
50, Daehak-ro, Chungju 27469, Korea
E-mail: inyoung@ut.ac.kr

Hyunkee Ahn
Professor
Department of English Language Education
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
E-mail: ahnhk@snu.ac.kr

Received: 30 June, 2025
Revised version received: 21 November, 2025
Accepted: 25 November, 2025

Appendix

Dialogue (Hahn & Dickerson, 1999, p. 11)

Male: What's the matter?

Female: I'm really in a rush.

Male: Do you need any help?

Female: Let's see. Maybe you could help me finish these graphs.

Male: No problem. Are you always so busy?

Female: Today I'm especially busy. This morning, I'm finishing a review of an article, and chairing a staff-meeting. Then I need to get ready to present a seminar.

Male: What's the topic of the seminar?

Female: You'll never believe it. It's "how to manage your time."