



Prosodic Manifestations of L2 English Read Speech and Pronunciation Proficiency: Bridging the Gap between Research and Practice

In Young Yang (Catholic Kwandong University)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: September 23, 2023

Revised: November 4, 2023

Accepted: November 25, 2023

In Young Yang

Assistant Professor, Dept. of English Education, Catholic Kwandong University

Tel: 033-649-7762

E-mail: inyoung@cku.ac.kr

ABSTRACT

Yang, In Young. 2023. Prosodic manifestations of L2 English read speech and pronunciation proficiency: Bridging the gap between research and practice. *Korean Journal of English Language and Linguistics* 23, 1094-1110.

This study aims to explore ways of bridging the gap between pronunciation research and teaching practices by evaluating learners' pronunciation proficiency through the assessment of the efficacy of two established pronunciation rating dimensions: nativelikeness and comprehensibility, with the goal of applying these dimensions in language classrooms. Speech samples from 35 Korean speakers of English were rated by 10 native English listeners in terms of pronunciation nativelikeness and comprehensibility. These samples were then analyzed for various prosodic speech characteristics, including prominence occurrence, pitch variation, pause occurrence, speech planning size, and speech rate. Subsequently, the speakers were clustered using *k*-means, and statistical tests were conducted using *R* to reveal the adequacy of evaluating learners' pronunciation proficiency based on the two pronunciation rating dimensions. The results showed that each speech factor correlated with pronunciation ratings to varying degrees, with pitch variation being statistically insignificant. Speech planning size demonstrated a stronger correlation with comprehensibility than with pronunciation nativelikeness. Pronunciation proficiency levels clustered based on the two rating dimensions revealed distinctive speech characteristics among the clusters, suggesting the effective implementation of these dimensions in pronunciation teaching practice.

KEYWORDS

second language pronunciation, L2 prosody, comprehensibility, nativelikeness, pronunciation proficiency

1. Background

1.1. Pronunciation in Second Language Acquisition

In English teaching in the EFL context, interest in speaking/listening education is increasing both formally (i.e., language classrooms) and informally (i.e., self-directed exposure to spoken forms). Pronunciation is an essential part of the speaking and listening process, forming important decoding processes of spoken language (Field 2009); however, in many cases, the importance of pronunciation is neglected (Levis 2022). Interest in pronunciation teaching has greatly decreased in the communicative language-teaching paradigm (CLT) (Celce-Murcia, Brinton and Goodwin 1996), and relatively few linguistic studies have discussed communicative value in terms of pronunciation.

In particular, one aspect of pronunciation that is neglected in CLT is segmental sounds in English, and as a result, interest in prosody, which shows more global characteristics, has tended to increase (Celce-Murcia, Brinton and Goodwin 1996, Menne and Leeuw 2014). Accordingly, some studies have addressed the prosodic characteristics of L2 speech and emphasized the importance of the influence of prosody on language use and comprehension (Baker 2010, Hahn 2004, Kang 2010, Trofimovich and Baker 2007, Wennerstrom 1994, 1998). Additionally, research on English pronunciation at the discourse level has begun (Celce-Murcia, Brinton and Goodwin 1996).

Another issue related to pronunciation in second language acquisition (SLA) is the gap between phonetic research and teaching practice (Lieberman 2008). Unlike other areas of SLA, second language (L2) phonetic research and practice are not closely related (Miller 2018). Despite a considerable amount of research, pronunciation teaching practice remains limited to listening and repeating. Furthermore, the assessments of pronunciation proficiency are considered challenging (Yates, Zielinski and Pryor 2011). Consequently, this study was motivated by the lack of attention to overall learner pronunciation and pronunciation issues in SLA research and teaching practice, despite the growing interest in speaking and listening processes.

1.2. Pronunciation Rating Dimensions in Research and Practice

In the communicative language teaching paradigm, fluency is considered more important than accuracy, and the emphasis on accurate pronunciation has gradually faded with the increased emphasis on those with relatively high communicative values. The recognition that a foreign accent in learner speech does not necessarily hinder communication has shifted attention towards the intelligibility and comprehensibility of speech (Munro and Derwing 1995), and the question of how similar the pronunciation is to that of native speakers (i.e., foreign-accentedness/nativeness) has gradually lost importance in the educational field. Accordingly, many studies have been conducted in the field of L2 phonetics/pronunciation to uncover the factors that affect or predict nativeness (the perceived level of accentness by English native listeners)¹ and comprehensibility (listeners' judgment of the ease or difficulty of understanding the presented speech) (Isaacs and Trofimovich 2012, Saito, Trofimovich and Isaacs 2017, Trofimovich and Isaacs 2012).

Given that native listeners assess nativeness and comprehensibility on an intuitive, global, and holistic basis, numerous intertwined factors may influence these evaluations. Several studies have sought to identify the factors contributing to enhanced comprehensibility in learner speech and establish their distinct associations with each rating dimension (Isaacs and Trofimovich 2012, Kang 2010, Trofimovich and Isaacs 2012). Nevertheless, the inherent variability of speech, combined with diverse speaker variables such as native language background and

¹ Recently, accentness is occasionally referred to as linguistic nativeness, with the two terms being interchangeably used (i.e., Saito, Trofimovich and Isaacs 2017). In this study, the higher the degree of accent in pronunciation, the lower the score, whereas a more nativelike pronunciation results in a higher score. Therefore, for the sake of better readability, the term 'nativeness' is maintained throughout.

language proficiency levels, frequently leads to conflicting findings.²

Despite the research interest in pronunciation rating dimensions, identifying applicable and effective pronunciation dimensions and rating methods for teaching purposes remains challenging. Moreover, pronunciation does not necessarily progress in tandem with learners' overall language proficiency. As a result, it becomes necessary to explore how to define L2 pronunciation proficiency through efficient and effective rating processes for practitioners.

1.3. Research Questions

In the pronunciation rating rubric of some current speaking tests such as the IELTS Band score (2023), detailed descriptions of the subareas of pronunciation are provided for scoring. For example, on Band score 6, the descriptions include: *Uses a range of phonological features, but control is variable, Chunking is generally appropriate, but rhythm is affected by a lack of stress-timing and/or a rapid speech rate, Some effective use of intonation and stress, but this is not sustained, Individual words or phonemes may be mispronounced but this causes only occasional lack of clarity, Can generally be understood throughout without much effort.* However, unless a speech analysis tool such as *Praat* software is used, raters have no choice but to rely on impressionistic evaluation, which inevitably introduces subjectivity into the assessment. In addition, the sub-criteria for the pronunciation assessments are unlikely to be exhaustive.

On the other hand, when assessing learners' pronunciation, utilizing both dimensions of nativelikeness and comprehensibility ratings, although also based on the listener's intuition, offers the advantage of evaluating pronunciation more holistically and easily from a communicative perspective. Based on the results of these pronunciation rating dimensions, we examine how effectively students' pronunciation proficiency can be distinguished in terms of their prosodic aspects when divided into three groups.

Based on the research background, this study set the following two research questions:

1. How is speech prosody manifested in the English read speech of Korean learners?
2. To what extent do the two rating dimensions (nativelikeness and comprehensibility) differentiate learner prosody?

To answer these questions, this study examined five factors related to learner speech prosody: prominence occurrence, use of pitch variation, pause occurrence, size of speech planning, and speech rate. These factors were analyzed based on two dimensions of pronunciation rating: nativelikeness and comprehensibility. Subsequently, learners' pronunciation proficiency was clustered based on the pronunciation rating results to verify significant differences among the groups with a specific focus on prosodic aspects. Specifically, the recorded speech of 35 people was clustered into three levels, and the prosodic characteristics of each group were examined to determine their suitability for use in language classrooms. The study also provides implications regarding learners' prosodic characteristics, pronunciation proficiency assessment, and pronunciation teaching practices.

² This study diverges from a conventional research perspective by shifting its focus away from examining the factors influencing the two pronunciation rating dimensions. Instead, its primary aim is to investigate the practical applicability of established pronunciation rating dimensions in real language classrooms. Consequently, this paper deliberately avoids an in-depth discussion of the factors that influence these two pronunciation rating dimensions.

The following results were predicted based on the research questions and procedures outlined above:

1. The five speech variables being measured will generally have ‘better’ characteristics as the pronunciation rating results increase.
2. Among the measured speech characteristics, those with more linguistic (phonological) characteristics (prominence, pitch range) and fluency aspects (pause, speech planning, speech rate) may exhibit different patterns of change.
3. Due to the nature of the read speech employed in this study, there may be no significant differences in some aspects of speech among learners for scores above a certain level.
4. When the learners are categorized into three levels based on the pronunciation rating results, their prosody patterns are expected to reveal distinct characteristics to some extent.

2. Methodology

2.1. Participants and Recording

A total of 35 voluntary participants were recruited at a university in Seoul, including both graduate and undergraduate students with diverse backgrounds in English and majors, to secure the basic to full-fledged realization of L2 speech prosody. All participants were female speakers of L2 English to avoid confounding variables that were likely to affect the acoustic analysis. The average age of the speaker participants was 24.3 (*SD* 5.21, Min 19, Max 39) at the time of recording. Of the 35 participants, 17 reported some length of residence in English-speaking countries, and the period of their stay varied from 1 to 228 months. The age at arrival in the country ranged from 1 to 35 years. One native speaker of North American English from Canada was also included in the recording to confirm the reliability of the native English listeners’ ratings of L2 pronunciation.

The speakers were individually placed in a recording booth with a soundproof facility. They read a short paragraph from a textbook for discourse-level pronunciation practice containing six short sentences using the built-in microphone of a SONY PCM-M10 recorder. Considering that the text was on linguistics, albeit a very basic level, and containing a technical term ‘morpheme,’ participants were encouraged to seek clarification on the accurate pronunciation of any unfamiliar words. Importantly, such inquiries did not bear significance in this study, as the primary focus was on assessing the production of suprasegmentals. The text is as follows:

A morpheme is the basic unit of meaning in a language. There are two major types of morphemes. First, let’s look at free morphemes. “Group” is a free morpheme. “Child” is also a free morpheme. So free morphemes are independent words. (Hahn and Dickerson 1999)

2.2. Ratings

In all, 10 native speakers (6 male and 4 female) of North American English (7 from the USA and 3 from Canada) participated as listeners and rated speech samples from 35 Korean speakers of English. The rating was conducted

individually in a quiet room, and the recorded speeches were allowed to be replayed if needed, but in fact, were rarely replayed. Each speech sample was scored on a seven-point Likert scale for pronunciation nativelikeness and comprehensibility, respectively (1 = *very strongly accented/incomprehensible* to 7 = *nativelike/comprehensible*). Five speech data samples were provided before the actual rating to familiarize the participants with the task. They varied in their length of residence in Korea from 4 months to 10 years, and nine of them reported that they frequently heard Korean-accented English at work or school.

2.3. Prosody Measurements

Five prosodic characteristics were measured to examine the prosodic characteristics of L2 read speech: prominence occurrence, use of pitch variation, pause occurrence, speech planning size, and speech rate. The following describes the measurements of each factor, and all measurements were conducted using *Praat* (version 5.3.82).

2.3.1. Prominence Occurrence

It has been known that Korean learners do not produce sentence-level stress appropriately and frequently (Yang 2021). Assuming that the better the pronunciation proficiency, the more frequent the prominence occurrence, the prominence occurrence frequency was calculated by the number of prominences per minute (*pace* in Vanderplank 1993). For the speech data, four English native-speaker graduate students who have educational backgrounds in phonetics and phonology marked words with prominence. They listened to sentences with the discursial context removed and marked words with prominence and boundaries, and those words were considered prominent if more than three of the four native listeners marked them as prominent.

2.3.2. Use of Pitch Variation

Korean learners of English tend to have relatively compressed pitch ranges (Kang, Guion-Anderson, Rhee and Ahn 2012). A narrow pitch range may indicate an insufficient realization of the prosody necessary for English speech, such as word- and sentence-level stress and boundary tones. As pronunciation proficiency increases, it can be predicted that word- and sentence-level stress will be better used, and as a result, the range of pitch changes is likely to gradually increase. Kang, Guion-Anderson, Rhee and Ahn (2012) observed that the pitch range of Korean English learners with considerable experience living in English-speaking countries was similar to that of native speakers. The authors interpreted these results as indicating that the overall pitch range used in the target language can be learned. Therefore, in this study, the pitch range of each speaker was calculated by measuring the lowest and highest fundamental frequencies appearing in the speech data.

2.3.3. Pause Occurrence

According to previous research, the appearance of pauses is significantly correlated with the nativelikeness and/or comprehensibility of learner speech (Kang 2010, Trofimovich and Baker 2006, 2007) as well as fluency (Derwing 2017, Kahng 2018). High pronunciation proficiency also suggests that pauses occur less frequently. In this study, the total duration of pauses was calculated by adding the lengths of all pauses that appeared throughout the utterance. Following the definition of previous research (Kang 2010, Trofimovich and Baker 2006), silence

longer than 0.1 s was considered a meaningful pause.

2.3.4. Size of Speech Planning

It is expected that the more proficient a speaker is in language and pronunciation skills, the longer their simultaneously planned speech will be. In this study, we calculated the mean length of run to measure the size of the speech planning for each speaker. The mean length of run is defined as the number of syllables uttered without a pause (defined as 0.1 s and longer). Therefore, we counted and averaged the number of syllables between all pauses for each speaker. Research varies in its interpretation of the mean length of run, with some considering it a measure of speech rate (Kang 2010), while others view it as an indicator of speech planning capability (Tseng, Su, Huang and Visceglia 2010). In this study, we posit that the characteristics of the MLR measurement method are more intuitively related to the size of speech planning and distinguish it from speech rate.

2.3.5. Speech Rate

To measure the speech rate factor, which is widely known to affect the nativelikeness and comprehensibility of pronunciation as well as fluency (Derwing 2017, Kang 2010), the number of syllables per second was calculated. Speakers with a higher pronunciation proficiency are expected to exhibit faster speech rates. However, no significant difference is expected after reaching the appropriate or moderate speech rate at a certain level. Furthermore, considering the characteristics of the data (read speech) used in this study, this tendency may become more noticeable.

2.4. Statistical Treatment

All statistical analyses were conducted using *R* (4.3.1), and the following statistical analyses were performed according to the flow of the study. First, intraclass correlation coefficients were calculated to determine the inter-rater reliability of native listeners' evaluation results for the L2 read speech data. Subsequently, a correlation analysis was conducted between each pronunciation rating dimension (i.e., pronunciation nativelikeness and comprehensibility of L2 read speech as evaluated by native speakers) and the prosodic features of speech. Next, clustering analysis using *k*-means was performed to classify learners' pronunciation proficiency based on the holistic ratings of native speakers. Finally, this process was followed by an ANOVA test and Scheffé's post-hoc analysis to generalize the differences in prosodic characteristics between the groups.

3. Results

3.1. Rater Reliability

First, for the speech data of one native speaker included in this study, all 10 native listeners gave 7 points for both the nativelikeness and comprehensibility of the given speech. This means that the native listeners who participated in this study evaluated the pronunciation of English utterances appropriately and based on the norms of native speaker pronunciation, and that their evaluations were reliable.

An intraclass correlation coefficient analysis was performed to confirm the rater reliability of the 10 native

listeners' ratings of nativelikeness and comprehensibility of the learners' read speech. The reliability of the comprehensibility score of the 10 raters was 0.92 (intraclass correlation coefficients – average random raters), and the reliability analysis result for nativelikeness ratings was 0.94.

3.2. Prosodic Manifestations

Table 1 shows the rating scores for the two dimensions of learners' pronunciation and the descriptive statistics of the five prosodic factors measured.

Table 1. Descriptive Statistics of the Two Rating Scores and Five Prosodic Measures

	Mean	<i>SD</i>	Min	Median	Max	<i>N</i>
Nativelikeness	4.494	1.328852	2.3	4.4	6.9	35
Comprehensibility	4.949	1.245078	2.5	5	6.8	35
Prominence frequency	14.27	7.17325	0	14.112	31.339	35
Pitch range	195.3	44.83903	113.7	195.8	270.9	35
Total duration of pauses	3.778	1.58046	1.483	3.269	7.577	35
Mean length of run	5.894	1.262112	3.5	5.7	8.143	35
<i>N</i> of syllables per second	3.24	0.538163	2.221	3.314	4.323	35

The average nativelikeness and comprehensibility scores of the Korean English speakers who participated in this study were 4.494 and 4.949, respectively. On average, comprehensibility was higher than nativelikeness of pronunciation. The read speech of the participants in this study was perceived by native speakers to have an average of 14 prominences per minute and showed an average pitch range of 195.3 Hz. The total length of the pauses in all utterances was 3.778 seconds, and, on average, speakers uttered approximately six syllables at a time without a pause. The average number of syllables uttered per second was 3.24 syllables.

The following point graph using ggplot shows the values of each prosodic element that appeared in the utterances of speakers who participated in the study. The *x*-axis lists individual speakers, from left to right, with lower to higher mean scores of the two rating results, and the *y*-axis represents the value of each prosody measure. The circles in each graph represent the measured prosodic values for each speaker. The larger the circle, the higher the speaker's comprehensibility. The darker the circle, the higher the degree of accentedness; the lighter the color, the lower the degree of accentedness, which increases nativelikeness. In the middle of the point graph, a trend line is displayed, showing the trend of each speaker's prosody measurement according to pronunciation proficiency.

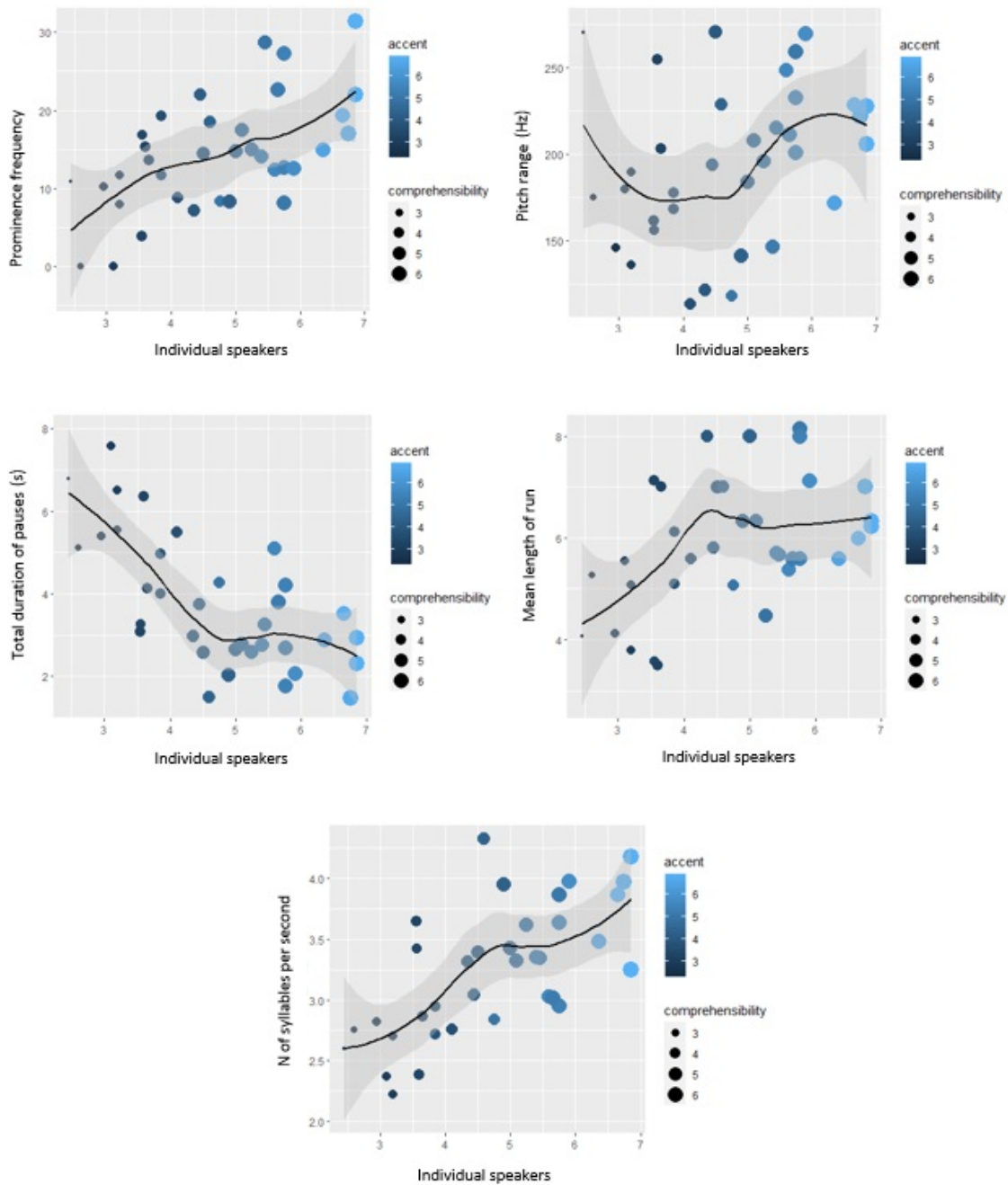


Figure 1. Point Graphs of Each Prosodic Measure with Trend Lines by Individual Speakers

1. Prominence frequency: It was found that the more nativelike and comprehensible the pronunciation was perceived to be, the more sentence-prominence was perceived to be present in the speaker’s speech. It can be seen that prominence appears more frequently through a line that continuously rises from the lowest evaluation score to the highest score. However, an interesting observation was made when the slope of the line was examined. Compared to the slope of the line shown by speakers with mean evaluation scores of four or less and six or more,

the trend of speakers with mean scores between four and six appears to have a somewhat smaller slope. This suggests that for learners with pronunciation proficiency in this range, development related to prominence may be slower and require more time to learn.

2. Pitch range: Contrary to the expectation that pitch variation would be greater for ‘better’ pronunciation, there was no clear, consistent trend. The pitch range of speakers with low pronunciation rating scores was quite large, and the pitch range of speakers with mid-level pronunciation scores was small on average but showed a very wide distribution.

3. Total duration of pauses: As predicted, the sum of the pause lengths was shorter for students who received higher evaluation scores. In addition, because of the characteristics of read speech, the slope of the trend line appears to become smaller for learners who received an average score of 5 or higher, and the differences in the sums of pause lengths for these speakers appear to be small.

4. Mean length of run: In the case of MLR, which indicates the number of syllables uttered without pauses, there was no significant difference when the average pronunciation evaluation score was 4 or higher. This appears to be a characteristic of read speech. The sentence length of the text used in this study was short, indicating that speakers with a certain level of English proficiency had no problem uttering sentences as a single unit.

5. *N* of syllables per second: The number of syllables uttered per second appears to increase rapidly from students with the lowest proficiency to those with intermediate proficiency, but there does not seem to be much change among speakers with an average pronunciation score of 5 or higher. As expected, this measure appeared to be highly correlated with the total duration of pauses, and the line graph shows an almost completely negative correlation.

Table 2 presents the results of the correlation analysis between the two pronunciation rating dimensions and the prosodic measures.

**Table 2. Results of Correlation Analysis
between the Pronunciation Rating Dimension and Prosodic Measures**

	Nativeness	Comprehensibility
Prominence frequency	.572**	.583**
Pitch range	.291	.316
Total duration of pauses	-.625**	-.725**
Mean length of run	.370*	.509**
N of syllables per second	.631**	.685**

The correlation analysis between each of the two pronunciation rating dimensions and the five prosodic measures showed that pitch range difference had no significant correlation with nativeness or comprehensibility. The other four prosodic measures exhibited overall moderate-to-high correlation coefficients, but the mean length of run with weak correlation with pronunciation nativeness. Meanwhile, the mean length

of run was found to have a greater correlation with comprehensibility than with nativelikeness. Table 3 shows the results of correlation analysis among the prosodic measures.

Table 3. Results of Correlation Analysis among the Prosodic Measures

	Prominence	Pitch range	Total duration of pauses	Mean length of run
Pitch range	.319			
Total duration of pauses	-.450**	-.11		
Mean length of run	.106	.118	-.660**	
N of syllables per second	.321	.224	-.905**	.604**

The results were generally as predicted; the total duration of pauses and the number of syllables per second showed a very strong negative correlation ($-.905^{**}$). Meanwhile, the number of syllables per second and mean length of run showed a moderate correlation of $.604^{**}$, and the total duration of pauses and mean length of run also showed a moderate correlation of $.660^{**}$. Other prosodic measures showing significant correlation were prominence frequency and total duration of pauses, with a correlation coefficient of $-.450^{**}$. No other significant correlations among the prosodic measures were found.

3.3. Clustering (K-Means) and Group Differences

To determine whether students' pronunciation prosody proficiency could be graded based on the two dimensions of pronunciation rating, namely nativelikeness and comprehensibility, a *k*-means clustering analysis was performed, and the prosodic characteristics of each group were examined.

To perform the *k*-means clustering analysis, the nativelikeness and comprehensibility rating results were set as reference variables. Using *R*, among all indices, seven indicated two clusters as the best number of clusters and five indicated three. Based on the basic and frequently used assumptions of English language education assessment levels, three groups were set as the number of clusters. In addition, when the number of clusters was set to two, the description of the prosodic measures was very simplified, so it had little value, not only for research but also for pronunciation teaching practice. Therefore, three groups were constructed to observe the learners' prosodic patterns in more detail by pronunciation proficiency.

Upon standardizing the pronunciation rating scores and performing *k*-means clustering with three clusters, the *Z*-values of the following cluster centers were derived (Table 4). Among the 35 learners, 12, 12, and 11 were in Clusters 1, 2, and 3, respectively.

Table 4. Z-values of Cluster Centers

	Nativelikeness	Comprehensibility	N
Cluster 1	-0.008	0.168	12
Cluster 2	-1.056	-1.163	12
Cluster 3	1.160	1.085	11

The clustering results were visualized in Figure 2 using gplot: The x-axis represents nativelikeness (1–7) and the y-axis comprehensibility (1–7).

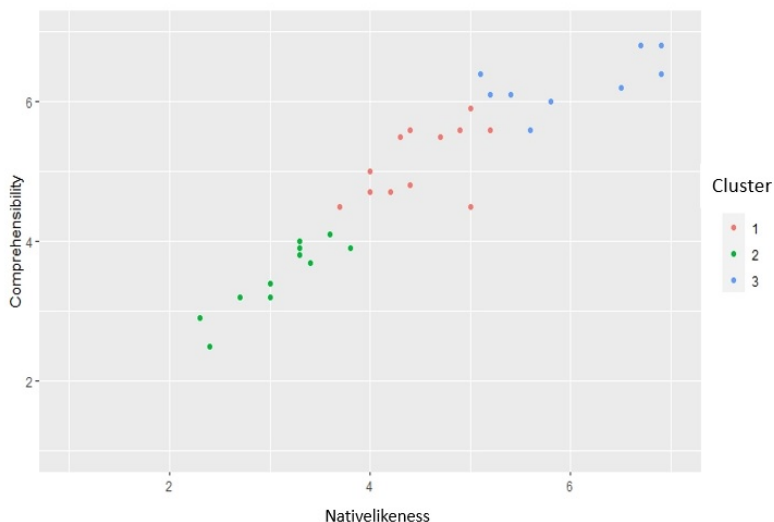


Figure 2. Clusters in a Scatter Plot by Comprehensibility and Nativelikeness

Cluster 2 was labeled Low, Cluster 1 Mid, and Cluster 3 High, and group information was then described (Table 5). The table below summarizes the average comprehensibility and nativelikeness scores of each group and the length of residence of the speakers in each group in English-speaking countries.

Table 5. Descriptive Statistics of Comprehensibility, Nativelikeness, and LOR by Three Clusters

Clustering	Nativelikeness		Comprehensibility		Length of residence (month)		
	Mean	SD	Mean	SD	Mean	SD	N
Low	3.0917	0.45419	3.5	0.48804	1.75	3.59608	12
Mid	4.4833	0.4745	5.1583	0.50535	14.2917	35.16872	12
High	6.0364	0.74333	6.3	0.38471	120	88.34478	11
Total	4.4943	1.32885	4.9486	1.24508	43.2143	74.22574	35

Learners clustered at the low level (Low) were found to have an average nativelikeness score of 3.0917 points and a comprehensibility score of 3.5 points, and their period of residence in an English-speaking country was 1.75 months, which is interpreted as almost no experience. For those clustered at the intermediate level (Mid), the average nativelikeness score was 4.4833, the average comprehensibility score was 5.1583, and the average period of residence in English-speaking countries was 14 months. High-level learners (High) showed an average nativelikeness score of 6.0364, a comprehensibility score of 6.3, and a very long period of residence in English-speaking countries, with an average of 120 months. Descriptive statistics of prosodic measures for each group are presented in Table 6 and Figure 3.

Table 6. Descriptive Statistics of Each Prosodic Measure by Three Clusters

Clusters	Low (<i>n</i> = 12)		Mid (<i>n</i> = 12)		High (<i>n</i> = 11)	
Prosodic measures	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>
Prominence frequency	10.11932	6.20037	14.81208	6.358536	18.20629	7.079226
Pitch range	184.817	40.85229	178.1128	49.88367	225.3998	27.92608
Total duration of pauses	5.22742	1.421227	3.056366	1.054763	2.98382	1.098805
Mean length of run	5.0261	1.259151	6.248184	1.090927	6.453368	0.981934
N of syllables per second	2.787057	0.413087	3.391186	0.433852	3.567809	0.445114

First, as pronunciation proficiency increases from Low to Mid to High, the average number of prominence occurrences tends to increase, and the difference in standard deviation within each group was not large. In pitch range, there were considerable differences among the three groups, with the Mid group having the largest standard deviation, and the High group showing a small standard deviation and large pitch variation. Next, the prosodic measures related to pauses, speech planning size, and speech rate showed different overall patterns for Low versus the other speakers, whereas Mid and High showed similar measurements.

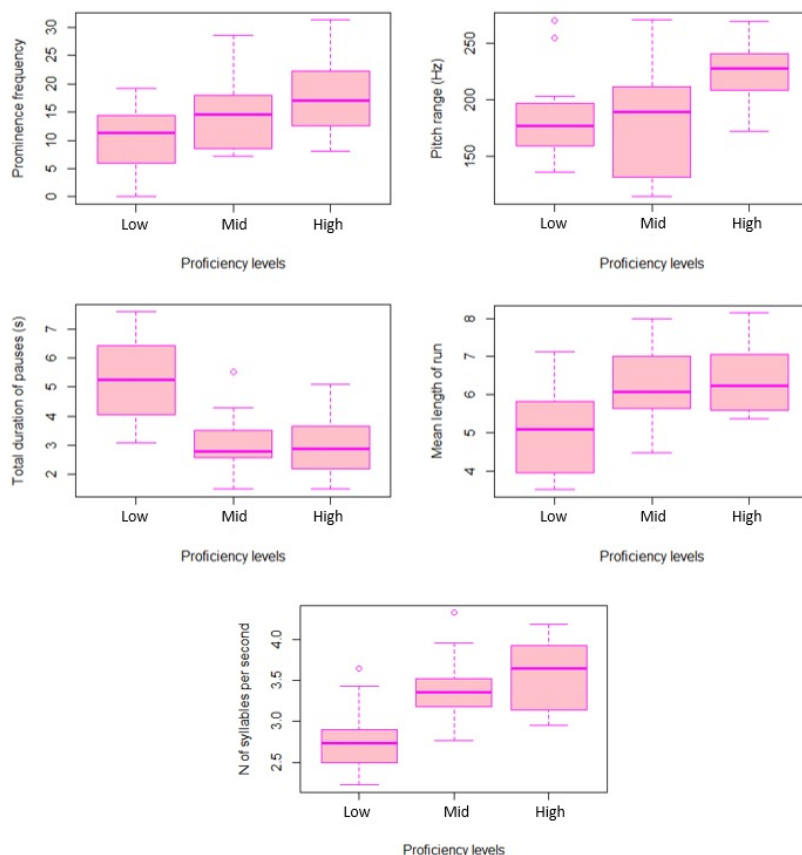


Figure 3. Boxplots of Each Prosodic Measure by Three Clusters

For the prominence frequency, as shown in Figure 3, the median value for each group gradually increased. The difference in the median values of the pitch range between Low and Mid was slightly smaller than that between Mid and High. As for pitch range, the difference between the maximum and minimum values in the Mid-level group was much larger than in the other groups, and the values of the first and third quartiles also showed a fairly wide distribution. It can also be observed that it is difficult to distinguish between Mid and High in total duration of pauses and mean length of run. Finally, regarding the number of syllables per second, the difference between the 1st and 3rd quartiles of the speech rates of High-level speakers was larger than that of the other groups. The median values gradually increased with the number of syllables uttered per second, from Low to Mid to High.

A set of ANOVAs was performed for each prosodic measure to confirm the statistical significance of the differences among the groups. In addition, a Scheffé test was performed as a post-hoc analysis to compare each group, allowing us to select truly significant differences in each prosody measure and identify the developmental stages of prosody at each level of pronunciation proficiency. Table 7 summarizes the results of the ANOVA and Scheffé's tests.

Table 7. Results of ANOVA for Each Prosodic Measure by Three Clusters

Prosodic measures	df	ANOVA				Post-hoc
		Sum Sq	Mean Sq	F-value	Pr(>F)	Scheffé
Prominence frequency	2	14829	7415	4.433	0.02*	Low<High,
Pitch range	2	380.7	190.35	4.45	0.0197*	Mid<High,
Pause	2	38.4	19.198	13.2	0.0000659***	Low<Mid=High
MLR	2	13.99	6.993	5.57	0.0084**	Low<Mid=High
Speech rate	2	3.918	1.9591	10.57	0.000298***	Low<Mid=High

Note. *** indicates $p < .001$; ** indicates $p < .01$; * indicates $p < .05$.

The results of the statistical analysis indicated that the differences between the groups were significant for all prosodic measures. However, the patterns of the groups that showed significant differences within each prosodic measure differed. As for Prominence frequency, there was a statistically significant difference between Low and High, but the Mid did not differ from either. There was a significant difference between Mid and High for pitch range, but Low did not differ from Mid or High. In the case of Pause, MLR, and Speech rate, the overall statistical significance was derived from Low versus the other two groups.

4. Discussion

4.1. Prosodic Characteristics of L2 English Read Speech of Korean Learners by Proficiency Level

Speakers with low pronunciation proficiency levels show less prominence in their speech, and pitch range might be interpreted as a characteristic that reflects individual speech style rather than linguistic development in this study. In addition, the total length of pauses in their speech was longer than in the other two groups; the length of utterance without a pause (MLR) was rather short, and the speech rate measured in syllables per second was also slow.

In the case of Mid-level speaker utterances, sentence-level prominence was found to exhibit the characteristics of an intermediate stage, being perceived less frequently on average than at a High-level and more frequently than at a Low-level. While there was a statistically significant difference between the Low and High levels, the Mid-level was not significantly different from either. Due to the complex nature of English prominence, acquiring it can be challenging (Baker 2010). Speakers with intermediate proficiency may not manifest noticeable differences from the other two groups, suggesting that a substantial improvement in pronunciation proficiency is necessary to acquire sentence-level prominence.

There was a substantial difference in the use of pitch variation between groups. A significant difference was observed between Mid and High, but Low did not differ significantly from either. Two interpretations are possible, but more carefully designed research is required with more participants: Speakers at mid-level or higher levels tend to use a wider range of pitch as their pronunciation proficiency increases, and low-level speakers may exhibit individual speech characteristics that are not related to linguistic factors. Therefore, it appears that learners at the Mid-level start using pitch variation more actively for linguistic purposes, but more refined research is needed. Another interpretation is that the significant *p*-value observed in this study may be attributable to chance at a very slight probability as no significant correlation was found between pronunciation proficiency and pitch range.

Regarding Pause, MLR, and Speech rate, overall statistical significance was observed between Low versus the other two groups. Mid-level speakers showed no significant difference from High-level speakers in terms of the total pause length, MLR, or speech rate of read speech.

In the utterances of high-level speakers, prominence was most frequently recognized, and the pitch range varied greatly. The total length of pauses that appeared during speech was the shortest, the number of syllables in one unit was the largest, and the speech speed was, on average, the fastest.

4.2. Utilizing the Two Pronunciation Rating Dimensions in Practice

This study used two dimensions of pronunciation rating that are frequently used in L2 pronunciation research to divide learners' pronunciation proficiency levels in order to determine whether the holistic and intuitive methods used in research could be utilized in educational settings and whether they could serve as a meaningful tool for any educational treatment. Many studies of learner pronunciation have been conducted on various aspects, but most phonetic research focuses on describing the characteristics of learner speech and does not necessarily pay attention to the implications for and implementations in teaching practice in the classroom.

Based on the results of this study, this holistic and intuitive rating effectively distinguishes learners' prosodic characteristics and holds implications for the development of learner prosody and adequate intervention that can be used in practice. Therefore, based on the prosodic characteristics of speakers and the experimental design of this study (i.e., read speech), the following guidelines for treatment related to prosody at each pronunciation proficiency level can be proposed. More valuable insights will become available through follow-up research based on this current exploratory study.

Low-level pronunciation proficiency: This stage necessitates both phonological language skills and fluency-related interventions. The learners in this category should be more aware of the phonological realization of word- and sentence prominence, utilizing suprasegmentals such as pitch, duration, and intensity more appropriately. Speech fluency, including quantitative aspects of pauses, speech rate, and speech planning size, may require, at the pronunciation level, more frequent and intensive intervention regarding phonetic information for learners to produce words without hesitation.

Mid-level pronunciation proficiency: The stage can be defined as where the development of prominence and active use of pitch variation begins but needs further improvements to be better understood in communicative context. The learners in this developmental stage may need to further refine not only their knowledge of the places of word-level prominence and phonological rules of placing sentence-level prominence but also the execution of their phonetics putting aside the interference from their native language characteristics.

High-level pronunciation proficiency: This stage represents a level where English speech is fluent, and prosodic properties are well developed. The speakers at this level had lived in English-speaking countries for an average of 10 years, and they were quite young when they arrived there. Therefore, it was predicted that this stage would not commonly be observed in typical EFL contexts.

5. Conclusion

This study examined the prosodic characteristics of the read speech of Korean speakers of English. In particular, we examined their prosodic characteristics by pronunciation proficiency based on nativelikeness and comprehensibility scores, two rating dimensions actively used in L2 pronunciation research. Furthermore, we attempted to derive specific implications that could be used in teaching practice by grouping the research participants' utterances into three levels.

This study is expected to contribute to L2 pronunciation research and practice. First, it will help us understand learners' interlanguage system. There have been many studies of learners' interlanguage, but pronunciation has not been in the mainstream of such research. In addition, pronunciation development often does not accompany the development of other language areas because of the nature of pronunciation acquisition (Kennedy and Trofimovich 2017). Therefore, systematicity in pronunciation development is often not observed when examined on the basis of general language proficiency. Therefore, it is more reliable to evaluate pronunciation proficiency separately and examine various speech characteristics. Second, based on the speech characteristics observed among the groups clustered using the results of the two holistic and intuitive rating dimensions, a foundation for more efficient and effective implementation of pronunciation assessment and intervention was laid. If follow-up research compares the results with those of non-native teacher expert raters and identifies the number of listener raters required, it will be of greater use in educational settings.

References

- Baker, R. E. 2010. *The Acquisition of English Focus Marking by Non-native Speakers*. Doctoral dissertation, Northwestern University, Evanston, IL, USA.
- Boersma, P. and D. Weenink. 2014. Praat: doing phonetics by computer (version 5.3.82) [Computer software]. Available online at <http://www.praat.org>
- Celce-Murcia, M., D. M. Brinton and J. M. Goodwin. 1996. *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Cambridge: Cambridge University Press.
- Derwing, T. M. 2017. L2 fluency development. In L. Shawn and S. Masatoshi, eds., *The Routledge Handbook of*

- Instructed Second Language Acquisition*, 246-259. NY: Routledge.
- Field, J. 2009. *Listening in the Language Classroom*. Cambridge: Cambridge University Press.
- Hahn, L. D. 2004. Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38(2), 201-223.
- Hahn, L. D. and W. B. Dickerson. 1999. *Speech Craft: Discourse Pronunciation for Advanced Learners*. Ann Arbor: The University of Michigan Press.
- IELTS. 2023. Scoring in detail. Available online at <https://www.ielts.org/for-organisations/ielts-scoring-in-detail>
- Isaacs, T. and P. Trofimovich. 2012. Deconstructing comprehensibility. *Studies in Second Language Acquisition* 34(3), 475-505.
- Kahng, J. 2018. The effect of pause location on perceived fluency. *Applied Psycholinguistics* 39(3), 569-591.
- Kang, O. 2010. Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System* 38(2), 301-315.
- Kang, S., S. Guion-Anderson, S.-C., Rhee and H. Ahn. 2012. The effect of language immersion on the acquisition of second language suprasegmentals. *Korean Journal of Applied Linguistics* 28(1), 179-207.
- Kennedy, S. and P. Trofimovich. 2017. Pronunciation acquisition. In M. Sato and S. Loewen, eds., *The Routledge Handbook of Instructed Second Language Acquisition*, 260-279. NY: Taylor & Francis.
- Levis, J. M. 2022. Teaching pronunciation: Truths and lies. In: C. Bardel, C. Hedman, K. Rejman, and E. Zetterholm, eds., *Exploring Language Education: Global and local perspectives*, 39-72. Stockholm: Stockholm University Press.
- Liberman, M. 2008. HVPT. [Blog post]. Available online at <http://languagelog.ldc.upenn.edu/nll/?p=328>
- Mennen, I. and E. de Leeuw. 2014. Beyond segments: Prosody in SLA. *Studies in Second Language Acquisition* 36(2), 183-194.
- Miller, J. 2018. Bridging the gap between pronunciation research and proficiency-based teaching. *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference*, 111-119.
- Munro, M. J. and T. M. Derwing. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45(1), 73-97.
- R Core Team. 2023. R: A language and environment for statistical computing (version 4.3.1) [Computer software]. Available online at <https://www.R-project.org>
- Saito, K., P. Trofimovich and T. Isaacs. 2017. Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics* 38(4), 439-462.
- Trofimovich, P. and W. Baker. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28(1), 1-30.
- Trofimovich, P. and W. Baker. 2007. Learning prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics* 28(2), 251-276.
- Trofimovich, P. and T. Isaacs. 2012. Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition* 15(4), 905-916.
- Tseng, C. Y., Z. Y. Su, C. F. Huang and T. Visceglia. 2010. An initial investigation of L1 and L2 discourse speech planning in English. *2010 7th International Symposium on Chinese Spoken Language Processing*, 55-59.
- Vanderplank, R. 1993. 'Pacing' and 'spacing' as predictors of difficulty in speaking and understanding English. *ELT Journal* 47(2), 117-125.
- Wennerstrom, A. 1994. Intonational meaning in English discourse: A study of non-native speakers. *Applied*

Linguistics 15(4), 399-420.

Wennerstrom, A. 1998. Intonation as cohesion in academic discourse. *Studies in Second Language Acquisition* 20(1), 1-25.

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.

Yates, L., B. Zielinski and E. Pryor. 2011. The assessment of pronunciation and the new IELTS Pronunciation Scale. *IELTS Research Reports* 12, 1-46.

Examples in: English

Applicable Languages: English

Applicable Level: Tertiary