The Impact of Segmental Accuracy on Intelligibility

In Young Na

A Scholarly Paper
Submitted in Partial Fulfillment of
Master of Arts in Second Language Studies

Department of Second Language Studies
University of Hawai‘i at Mānoa

First Reader: Dr. Dustin Crowther
Second Reader: Dr. Daniel Isbell

Spring 2021
THE IMPACT OF SEGMENTAL ACCURACY ON INTELLIGIBILITY

Abstract

Intelligibility over nativeness has been increasingly argued as an appropriate goal for second language (L2) pronunciation teaching (Levis, 2005), yet relatively little is known about the phonological factors that make a nonnative speaker’s speech intelligible. Previous studies on the impact of nonnative English speakers’ segmental (i.e., consonants and vowels) pronunciation accuracy on listeners’ level of actual understanding (i.e., intelligibility) are relatively limited, compared to that of suprasegmental features such as stress and intonation. Therefore, the current study examined the relationship between segmental accuracy (in terms of target segments in minimal pairs) and listener-based intelligibility (i.e., rate of accurate word identification across listeners). Eight native-English listeners were assigned to complete a minimal-pairs forced-choice task recorded by twenty Korean EFL learners. The results showed substantial and negative correlations between segmental accuracy and listener-based intelligibility. Vowel errors were linked to lower intelligibility overall, but the negative correlation between consonant errors and intelligibility was stronger. Some sounds were not substituted/produced erroneously by the speakers, yet still posed intelligibility problems to native listeners. Descriptive analyses of individual contrasts indicate target phonemes most often misperceived by native listeners. These findings will help make specific pedagogical recommendations for the teaching of English pronunciation.

Keywords: intelligibility, foreign accent, segmental errors, Korean speakers of English

Introduction

The importance of pronunciation instruction that leads to successful communication has been gaining recognition, yet empirical studies on the phonological factors that contribute to the intelligibility of nonnative speakers of English are limited. Therefore, it is worthwhile to
examine the impact of phonological features on the intelligibility of second language (L2) speech. The current study investigates the relationship between segment production accuracy and overall intelligibility by analyzing the English productions of L2 Korean learners of English. The findings will contribute to existing knowledge about L2 English pronunciation by the identified identifying problematic speech sounds for accurate identification of words. Specifically, those sounds are more likely to hinder listeners’ understanding of an utterance. Such knowledge will help instructors to set priorities for pronunciation teaching.

**Literature Review**

**Nativeness and Intelligibility**

In line with trends in the L2 pronunciation research, the pedagogical goal for pronunciation teaching has shifted a focus away from nativeness (or foreign-accent reduction) to the more realistic goal of intelligibility. Many English language teachers have gradually come to realize that it is often time-consuming and *de facto* unrealistic to pursue acquisition of native-like accents with the abundant amount of evidence arguing for intelligibility over nativeness (Field, 2005; Levis, 2005). For one, there has been a general agreement among researchers on a critical period for acquiring pronunciation, that is, an ability to learn to produce speech accurately is biologically constrained, and beyond that critical period, it is either difficult or impossible to achieve native-like accents (Flege, 2003). Furthermore, communication can be remarkably successful, even when foreign accents are strong (Levis, 2005). Derwing and Munro (2009) after conducting numerous studies on the relationship between accentedness and intelligibility, came to a conclusion that a given speaker could have a strong foreign accent, yet still be highly intelligible. Thus, the presence of a foreign accent does not necessarily indicate reduced intelligibility.
Defining the Construct

While the notion of intelligibility has become central to pronunciation research, much discussion of defining the construct has failed to arrive at a clear consensus (Field, 2005; Nelson, 2008; Kang et al., 2018). Smith and Nelson (1985) observed that terms intelligibility and comprehensibility are often used interchangeably. They suggested limiting the first to the recognition of words and utterances and the second to the construction of meaning. Quite different definitions were proposed by Derwing and Munro (1997), who differentiated the two constructs on the basis of actual understanding versus perceived ease of understanding. Yet, Derwing and Munro suggested that their notion of intelligibility involving the recognition of words and sentences, corresponded fairly closely to that of Smith and Nelson (Derwing & Munro, 1997; Deterding, 2013). Munro and Derwing (2015) defined intelligibility as the “extent to which listeners’ perceptions match speakers’ intentions (actual understanding)” (p. 141). Nonetheless, as pointed by Nelson (2008), there seem to have been numerous terminological inconsistencies in pronunciation work over a 40-year period.

Derwing and Munro (2009) defined and operationalized the three partially related constructs – accentedness, comprehensibility, and intelligibility – in terms of listeners’ perceptions of speech but made it clear that intelligibility should be distinguished from the other two. Accentedness refers to how an L2 speaker’s pronunciation differs from that of a target speech variety. Comprehensibility denotes the listener’s perception of how easy or difficult it is to understand the utterance. Lastly, intelligibility is defined as the actual understanding of the utterance by the listener. As Field (2005) noted, intelligibility could be rather narrowly defined as “the extent to which acoustic-phonetic content of the message is recognizable by a listener” (p. 401). The current study examines the impact of strictly phonological factors on
understanding, disregarding any higher-level evidence originated outside the speech signal (e.g., world knowledge, lexical knowledge). Therefore, Field’s narrower construct of intelligibility – restricted to features of the speech signal only - is more relevant and utilized for the purposes of the present study.

Regarding the operationalization of the three constructs, the first two constructs – accentedness and comprehensibility – are generally measured by listeners’ impressionistic judgments on a Likert scale, whereas there is much more variation in how intelligibility is measured. While intelligibility has most commonly been measured using listeners’ transcription, a clear comparison across studies on intelligibility is rather difficult due to still widely disparate methods that have been used for assessing intelligibility (Kang et al., 2018). Methods for measuring intelligibility vary in terms of the materials (e.g., words, sentences, passages), the type of L2 speech (i.e., read speech or spontaneous speech), and the tasks of the listeners (e.g., scalar ratings of intelligibility, orthographic transcription, nonsense sentences, filtered sentences, cloze test, comprehension questions, summaries, and true/false judgment; Bent et al., 2007; Kang et al., 2018; Munro & Derwing, 2015).

**Factors Influencing Intelligibility**

Continuous efforts have been made to identify specific features of L2 speech that contribute to each of the three constructs – accentedness, comprehensibility, and intelligibility – yet studies on intelligibility are less frequent. In addition, studies that do target intelligibility have been divided as to the relative contribution of segmental and suprasegmental features to speakers’ intelligibility. In the field of second language research, many have emphasized suprasegmentals over segmentals in promoting intelligibility. However, while many studies involved pronunciation and intelligibility judgments, only a few have investigated the actual
influence of prosodic features measuring functional intelligibility. Hahn (2004) reported a negative impact of incorrectly placed primary stress (sentence stress) on intelligibility. In her study, native listeners listened to one of three different versions of lectures (i.e., appropriate primary stress, incorrect primary stress, and no primary stress) given by a Korean speaker of English. Intelligibility was measured by asking the listeners to complete a recall task (i.e., write down as much as they could recall from the lecture) and short-answer comprehension questions. The findings showed that the listeners understood and retained more information from the speech with appropriate sentence stress. Field (2005) examined the role of word stress on intelligibility, as measured by (both native and nonnative) listeners’ orthographic transcription, of words manipulated to have accurate or inaccurate placement of stress. He found that word stress errors as well as stress-related changes in vowel quality negatively influenced intelligibility.

Mixed findings have been found regarding the relationship between segmental features and intelligibility, yet research has generally indicated that “some portion of variability in overall intelligibility can be accounted for by variability in production accuracy at the segmental level” (Bent et al., 2007, p. 334). Caspers and Horloza (2012) assessed the relative contribution of stress and segmental errors (phoneme substitution) to the perception of intelligibility of words produced by French and Chinese speakers of Dutch. Intelligibility of the stimulus words was tested either in an offline (orthographic transcription task) or online (word naming task) manner. For the offline intelligibility test, native listeners were auditorily presented with Dutch words produced by native and non-native speakers and asked to write down what they heard. For the online intelligibility test, listeners (different from the ones who completed the transcription task) were asked to listen to Dutch words spoken by native and non-native speakers and repeat each word as fast as they could. This test was conducted to measure additional reaction times to
THE IMPACT OF SEGMENTAL ACCURACY ON INTELLIGIBILITY

The impact of segmental accuracy on intelligibility, and thereby measuring the influence of (supra)segmental errors on processing times. The authors reported that both stress and segmental errors influenced intelligibility, while a combination of stress and segmental error influenced intelligibility and processing time the most. The authors also found that segmental errors were as harmful to intelligibility as suprasegmental ones, contrary to the general belief that suprasegmentals play a more important role than segmentals. Rogers (1997) examined the relationship between segmental accuracy and sentence intelligibility of Mandarin-accented English. Native listeners were asked to complete a minimal pair identification task, in which native listeners had to choose the word between two minimal pair alternatives corresponding to the one spoken by the nonnative speakers. For assessing sentence intelligibility, listeners transcribed sentences presented to them one at a time. Intelligibility scores were calculated according to the percentage of content words correctly transcribed by the listeners. The results indicated a significant positive correlation between segmental accuracy (as measured by minimal pair identification task) and intelligibility of sentences read by Mandarin speakers of English. Moreover, speakers’ sentence intelligibility was correlated positively with vowel production accuracy, but not with consonant accuracy. Bent et al. (2007) examined the relationship between segment production accuracy in different word positions and the overall intelligibility of Chinese-accented English. Native English listeners were assigned to transcribe sentences read by the speakers to assess their intelligibility, which was then calculated based on the number of keywords correctly transcribed by the listeners. Results showed that speakers' vowel production accuracy, but not consonant accuracy, correlates with their overall intelligibility. Furthermore, word-initial segmental errors were found to be more detrimental to intelligibility than errors in word-final positions and errors in other positions, suggesting the greater importance of word-initial information both for lexical access and word
THE IMPACT OF SEGMENTAL ACCURACY ON INTELLIGIBILITY

recognition. To recap, previous studies examining monologic speech have found the effect of segmental accuracy on intelligibility, indicating vowels as having a greater effect on listeners’ perception of intelligibility than consonants in general.

The aforementioned studies were all examining monologic speech based on lab-based elicitation and analysis. However, interactive research should be of interest to L2 pronunciation research since L2 usage in spontaneous communication often serves as an overarching goal of second language acquisition (Loewen, 2015). In fact, research on spoken interaction provides even greater support for segmentals in successful communication between language users (Jenkins, 2002; Kennedy et al., 2015; Loewen & Isbell, 2017). In an effort to identify sources of communicative breakdowns in interaction, Kennedy et al. (2015) found that pronunciation, primarily segmental accuracy, accounted for 18% of comprehension problems. These findings are congruent with Loewen and Isbell (2017), who also found that 16% of communication problems were related to pronunciation, and 90% of which came from segmental rather than suprasegmental concerns. Furthermore, Jenkins (2002) put a primary emphasis on segmental over suprasegmental productions by proposing the Lingua Franca Core (LFC), a theoretically motivated set of priorities for what should be taught specifically in the EIL (English as an International Language) context. The LFC has been criticized for its limited data set, in terms of both interlocutors and speech material. However, analyses of communicative breakdowns in her study generally indicated segmental sources of communication breakdowns. Accordingly, she made a distinction between “core” and “non-core” features of pronunciation that contribute to mutual intelligibility among nonnative users of English. While the core features of the LFC, those crucial to intelligibility, largely include segmental production (e.g., consonants and vowels), suprasegmental productions (e.g., word stress, pitch range, and rhythm) were relegated
to non-core features. The focal emphasis on segmental accuracy is of particular concern when considering the aforementioned laboratory research findings. Therefore, the present study focuses on segmentals over suprasegmentals.

Indeed, teachers have long resorted to L1-based textbook recommendations, or often just their intuition, in determining the relative importance of speech features rather than based on the principled application of research findings (Levis, 2005; McAndrews & Thomson, 2017; Munro, 2018). L1-based textbooks could be of practical use for teachers by giving some generic descriptions of problematic sounds largely predicted by the learners’ L1-specific pronunciation difficulties (McAndrews & Thomson, 2017; Munro, 2018); for example, learners in whose L1s the vowel /i/ does not exist will struggle with pronouncing that sound. However, L1-based pedagogical recommendations often lack empirical foundations, not providing any actual instances of pronunciation divergence in learners’ speech on which their predictions are based. Also, it should be noted that those recommendations do not give such information on pronunciation features’ relative importance for intelligibility (McAndrews & Thomson, 2017).

Given the limited portion of class time usually allotted to pronunciation instruction (Foote et al., 2012), it is useful to know which sounds of L2 speech have the highest effect on speakers’ intelligibility and thus should receive the most attention in class (Munro & Derwing, 2006; Foote et al., 2012). One concept that is useful in deciding which sounds to teach is that of functional load (FL). Catford (1987) developed a hierarchical list of phonemic contrasts based upon the frequency with which each phoneme contrasts with other phonemes in distinguishing different numbers of words. This relative weight then in turn allowed for each phoneme to be ranked from high to low on an FL continuum. For example, speech errors involving high FL phonemes (e.g., /l/ and /n/) theoretically may pose more problem for intelligibility than low FL
errors (e.g., /ð/ and /d/). However, this prediction needs to be tested empirically in order to validate its claims (Munro & Derwing, 2006; Bent et al., 2007; Foote et al., 2012). To conclude, more research should be conducted on identifying problematic speech sounds to set priorities for pronunciation teaching, centering on those speech features that directly feed into the intelligibility of the speakers.

**Situating the Current Study**

*Korean Learners of English*

There has been an overwhelming desire for learning English in Korea, where “children as young as five years as well as school-age students are studying English in cramming schools (called hagwon)” (Park, 2009, p. 50). English education in Korea is now gradually shifting away from its reliance on the Grammar-Translation Method to Communicative Language Teaching, recognizing the importance of English as a means of international communication (Hwang, 2008). Despite general agreement on the value of pronunciation as a critical element in successful communication, English pronunciation instruction has been neglected in K-12 classroom settings in Korea (Hwang, 2008). Hwang noted among several reasons, this may be attributed to Korean teachers’ insufficient knowledge of what and how to teach pronunciation, which should clearly be guided by research findings. However, compared to research targeting Chinese, Japanese, or Spanish learners of English (e.g., Derwing & Munro, 1997; Saito & Lyster, 2012; Xie & Fowler, 2013), there are fewer studies conducted on Korean learners of English. As these learners are those who will be most impacted by the shift in teaching methods, understanding the extent of their intelligibility is necessary for subsequent pronunciation instruction. Therefore, the present study begins to address these concerns by analyzing segmental
errors produced by Korean learners of English, a group underrepresented in current L2 speech intelligibility.

**Research Questions**

The review of the literature showed previous research on the relationship between segment production accuracy and the overall intelligibility of L2 speech. Relatively disparate research methods and contradictory results hardly allow for firm conclusions to be made. Several gaps in the research on the construct of intelligibility emerge from the review of the literature, suggesting research opportunities in this area. First, only a small number of studies have examined segmental features of L2 speech in relation to how a listener actually understood at a word-level. Existing literature tended to have paid more attention to suprasegmental features or listeners' judgments of comprehensibility. Second, limited research has been conducted on Korean learners of English. Third, while some segment distinctions have been proposed to have a greater impact on listeners' perceptions of intelligibility, this has yet to be tested empirically. To address these concerns, the current study was guided by the following research questions:

1. What is the relationship between segmental errors and Korean learners’ English intelligibility as perceived by native English listeners?
2. Which segmental sounds are most difficult for Korean learners to produce intelligibly as perceived by native English listeners?

**Methodology**

**Participants**

**Speakers**

The speech samples were recorded from 20 Korean speakers of English at a high school in Seoul, South Korea; all speakers were in their final year at the time of study (Female = 16, Male = 4; Age = 18). Speakers’ language background information is summarized in Table 1.
Speakers had studied English for 8.76 years (SD = 3.22); 5 reported study abroad experience, among whom three participants had their experience in English-speaking countries and two in non-English-speaking countries (China = 2). All but two Speakers reported English as their L2 (2 reported Japanese). None of the Speakers reported any history of speech or hearing disorders.

Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of English study</td>
<td>8.76</td>
<td>3.22</td>
</tr>
<tr>
<td>Duration of study abroad experience</td>
<td>0.57</td>
<td>2.23</td>
</tr>
<tr>
<td>Self-rated English proficiency (on a 9-point scale)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaking</td>
<td>5.15</td>
<td>2.11</td>
</tr>
<tr>
<td>Listening</td>
<td>6.8</td>
<td>1.44</td>
</tr>
<tr>
<td>Reading</td>
<td>6.16</td>
<td>1.43</td>
</tr>
<tr>
<td>Writing</td>
<td>4.58</td>
<td>1.87</td>
</tr>
<tr>
<td>English use at home/school (on 4-point scale)</td>
<td>2.85</td>
<td>0.75</td>
</tr>
<tr>
<td>English exposure at home/school (on 4-point scale)</td>
<td>2.90</td>
<td>0.96</td>
</tr>
</tbody>
</table>

*a1 = extremely poor, 9 = extremely fluent. b1 = none, 4 = often

Listeners

Eight native English-speaking listeners (Mage = 27.25, SD = 8.66, Range = 19-45; Female = 7, Male = 1) were recruited from the University of Hawai‘i at Mānoa. Listeners included both undergraduate students and graduate students (undergraduate = 2, graduate = 6), who were primarily from the Department of Second Language Studies (n = 5; all were graduate students); the remaining 3 listeners represented different academics disciplines (Political Science = 1, Communications = 1, Korean Studies = 1). All of the participants reported knowledge of an L2 (Korean = 3, Japanese = 2, Chinese = 1, Tagalog = 1, Spanish = 1). Listeners' ratings on a 5-point Likert scale (1 = Not familiar at all, 5 = Very familiar) for their familiarity with accented-English, in general, were on average 4 (SD = 0.76) and specifically with Korean-accented
English were 4 (SD = 0.74). None of the native listeners reported any history of speech or hearing impairment. Listeners were paid an honorarium for their participation.

**Speech Materials**

In order to explore the effects of segmental errors, a list of ten minimal pairs was created selected from Avery and Ehrlich’s (1992) chapters on pronunciation problems of different L1 groups (presented in Table 2). The list contained problematic phonemes specifically targeting Korean learners of English as well as those that are common to English learners from most language backgrounds. Each item in the list consists of one target word and one foil word. The target word contains the target (intended) phoneme, and the foil word differs from it by only a single phoneme (as in the target-foil pair *rhyme-lime*).

**Table 2**

*Minimal-pairs word list*

<table>
<thead>
<tr>
<th>No.</th>
<th>Target</th>
<th>Foil</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>bit</td>
<td>/ɪ/</td>
</tr>
<tr>
<td>2</td>
<td>fan</td>
<td>/ʊ/</td>
</tr>
<tr>
<td>3</td>
<td>lagging</td>
<td>/g/</td>
</tr>
<tr>
<td>4</td>
<td>rhyme</td>
<td>/ɹ/</td>
</tr>
<tr>
<td>5</td>
<td>ripping</td>
<td>/p/</td>
</tr>
<tr>
<td>6</td>
<td>sand</td>
<td>/æ/</td>
</tr>
<tr>
<td>7</td>
<td>sheet</td>
<td>/ʃ/</td>
</tr>
<tr>
<td>8</td>
<td>song</td>
<td>/ɔ/</td>
</tr>
<tr>
<td>9</td>
<td>taste</td>
<td>/ɛɪ/</td>
</tr>
<tr>
<td>10</td>
<td>vote</td>
<td>/v/</td>
</tr>
</tbody>
</table>
Data Collection Procedure

Recording

Speech data collection was conducted at a high school in Seoul, Korea, in a private and silent setting. A known and trusted contact at the school collected data under instructions provided by the researcher during the whole session. Twenty Korean speakers of English were asked to complete a word-reading task. Before the task, they received oral and written task instructions. They were then given the list of ten target words and were asked to indicate any unfamiliar words, in which case a definition and a pronunciation model were provided. Each speaker then recorded their speech individually using the mobile app, Extempore (https://extemporeapp.com) under the supervision of the research assistant. They read aloud the ten target English items that were displayed one by one on the app. Their speech data were automatically recorded on the app and sent to the researcher. Recordings were edited using the audio software program Audacity (https://www.audacityteam.org); peak amplitude was normalized to 0.0 across samples and sound files were all converted to a .wav file format. The edited sound files were embedded into a survey, after randomization, using the online experiment platform Gorilla (https://gorilla.sc).

Listener Task

Eight listeners completed a 200-item (10 words produced by 20 speakers) minimal-pairs forced-choice task administered using Gorilla. Prior to the actual task, listeners received online written instructions through the Gorilla interface (see Appendix for instructions). Listeners then completed a 5-item practice session, where they heard speech samples produced by an additional nonnative speaker of English (a Vietnamese speaker of English to avoid any initial familiarization effect). The same words were used as speech stimuli for the practice session as in
the actual task. In the actual 200-item task, listeners listened to each speech sample recorded by the speakers on each minimal-pairs forced-choice trial. The corresponding target-foil pair was then simultaneously displayed on the computer monitor, and listeners chose which word they heard. Listeners were allowed as long as they needed to respond. They pressed the play button when they were ready to hear the next speech sample (see Figure 1). Listeners were given two short breaks throughout the task, which was completed in 25 minutes or less. Responses were scored automatically for the minimal-pairs task.

**Figure 1**

*Display used in minimal-pairs task*

Press the play button when you are ready to hear the audio clip.

Choose the closest word you heard.

**Data Analysis**

**Listening Task**

Following Rogers (1997), the overall intelligibility scores were calculated by the mean percentage of words identified correctly by the native listeners (henceforth, percentage correct) for each of the 20 Korean speakers of English. For each of the ten target words, the mean percentage of items correctly identified by the listeners was also computed across all speakers (out of 160 items possible for each target word; 8 intelligibility ratings for the same target word produced by 20 speakers). The ten target words were then divided into those containing consonant targets (n = 6) and vowel targets (n = 4), and speakers’ mean intelligibility scores for
each category were also calculated. This lead to a total of four types of intelligibility scores that were considered during analyses: overall intelligibility (i.e., mean percentage correct for each speaker), item intelligibility (i.e., mean percentage correct for each of the ten target item), vowel intelligibility (i.e., mean percentage correct for items containing vowel targets), and consonant intelligibility (i.e., mean percentage correct for items containing consonant targets).

**Error Coding**

Error coding was conducted focusing on speakers’ deviations on target phonemes from General American (GA) English pronunciation. The GA English that was used as a basis for comparison was the English referenced in Avery and Ehrlich (2008). Each target phoneme (i.e., that distinguishes a minimal pair with the foil) in each word (n = 200) was coded as either “on target” or “not on target” by the researcher and a secondary coder with an L1 English background. Following previous studies (e.g., Caspers & Horloza, 2012; Julkowska & Cebrian, 2015), only those errors that were perceptually salient as indicating phoneme substitutions (e.g., /p/ for /f/ in ‘faint’) were considered as “deviations”. Phonetic errors that did not deviate substantially from GA English were disregarded (e.g., differences in aspiration, vowel quality variations that did not impede identification). The analysis allowed making an inference that if a word is problematic to the listeners, it would be because the target phoneme was creating the difficulty, and not due to the other phonemes in the word being mispronounced. For example, in the word “lagging (/lægɪŋ/)”, where the target phoneme is /g/, we coded it as “on target” if we heard /g/, but “not on target” if we heard any other phoneme than /g/, such as /k/ (e.g., /lækɪŋ). Accordingly, if “lagging” was difficult for listeners to identify accurately and coded as “having a deviation”, it is the /g/ sound likely creating the difficulty. The initial agreement between the two coders was 89.5 %. Cases of discrepancy were reevaluated by a closer visual (inspecting
spectrograms) and auditory examination until an agreement on all errors was reached. We used the software program Praat (https://www.fon.hum.uva.nl/praat/) for this acoustic analysis (Boersma & Weenink, 2021). The number of segmental errors made by each speaker was listed as well as in terms of each of the ten segment targets across the speakers. The total number of errors was then divided into consonant and vowel errors. To avoid any confusion, the term “error(s)” is consistently used throughout this paper to refer to analytically coded errors in learners’ production of target phonemes, rather than errors in identification by the listeners.

Results

The results of the Cronbach alpha analyses confirmed the acceptable inter-rater reliability for the 8 listeners’ intelligibility judgments ($\alpha = 0.87$) in line with previous intelligibility studies (e.g., Munro et al., 2006 for $\alpha = 0.825-0.922$). Therefore, their judgments were averaged to compute overall intelligibility scores for each speaker. For statistical analyses, alpha was set at .05, and Cohen’s $d$ effect sizes were interpreted following Plonsky and Oswald (2014).

The first research question asked about the relationship between segmental errors and speakers’ overall intelligibility scores. Figure 2 presents the distribution of each speakers’ overall intelligibility scores based on listener responses for the minimal-pairs task. Overall Intelligibility scores ranged from 58.75% to 91.25% ($M = 78.56$, $SD = 7.43$). When the overall intelligibility scores were compared with the total number of target segment errors made by each speaker ($M = 2.60$, $SD = 1.27$, Range = 0-4; out of a maximum of 10 errors, 1 for each target word), a strong negative association ($r = -.750$, $p < 0.001$) was obtained (presented in Figure 3). This negative correlation demonstrates that speakers who produced a higher number of segmental errors tended to be less intelligible than speakers who produced fewer segments incorrectly. While overall
segment production accuracy was correlated with overall intelligibility, this correlation was carried by a subset of the segments; vowel errors were moderately correlated with overall intelligibility ($r = -0.396, p < 0.05$), and consonant errors revealed a strong correlation ($r = -0.790, p < 0.001$) with overall intelligibility. Note that overall intelligibility here indicates the mean percentage of items (including both consonants and vowels) identified correctly by the listeners for each speaker.

**Figure 2**

*Mean intelligibility scores (percent words correct) for the 20 Korean speakers of English*

![Figure 2](image)

**Figure 3**

*The effects of segmental errors on 20 Speakers’ mean intelligibility scores*

![Figure 3](image)
Table 3

Absolute (and relative) number of words identified correctly for each target item in the listener task, absolute (and relative) number of segmental errors identified in error coding

<table>
<thead>
<tr>
<th>Target items</th>
<th>Intelligibility (listener task)</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consonant targets</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fan (/f/)</td>
<td>140 (87.50%)</td>
<td>0</td>
</tr>
<tr>
<td>lagging (/g/)</td>
<td>157 (98.13%)</td>
<td>0</td>
</tr>
<tr>
<td>rhyme (/r/)</td>
<td>109 (68.13%)</td>
<td>7 (35%)</td>
</tr>
<tr>
<td>ripping (/p/)</td>
<td>151 (94.38%)</td>
<td>0</td>
</tr>
<tr>
<td>sheet (/ʃ/)</td>
<td>158 (98.75%)</td>
<td>0</td>
</tr>
<tr>
<td>vote (/v/)</td>
<td>103 (64.38%)</td>
<td>5 (25%)</td>
</tr>
<tr>
<td><strong>Total number</strong></td>
<td>818 (85.21%)</td>
<td>12 (10%)</td>
</tr>
<tr>
<td><strong>Vowel targets</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bit (/ɪ/)</td>
<td>59 (36.88%)</td>
<td>14 (70%)</td>
</tr>
<tr>
<td>sand (/æ/)</td>
<td>81 (50.63%)</td>
<td>12 (60%)</td>
</tr>
<tr>
<td>song (/ɔ/)</td>
<td>143 (89.38%)</td>
<td>14 (70%)</td>
</tr>
<tr>
<td>taste (/eɪ/)</td>
<td>156 (97.50%)</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total number</strong></td>
<td>439 (68.59%)</td>
<td>40 (50%)</td>
</tr>
<tr>
<td><strong>Total counts</strong></td>
<td>1257 (78.56%)</td>
<td>52 (26%)</td>
</tr>
</tbody>
</table>

The second research question asked about the most difficult segmental sounds for speakers to produce intelligibly. The analyses by individual target phoneme illustrate the target phonemes that were most often identified accurately as intended and those that were more likely to be misidentified. To obtain intelligibility scores for specific phoneme targets, the number of correct responses on the listener task was tabulated across listeners for each speaker, and
percentage error (i.e., phoneme targets coded as “not on target” by error coding) was computed for each target phoneme (presented in Table 3). Overall percentage error across the different consonant targets was found to be 10% (out of 120 items produced by the speakers), whereas for vowel targets 50% (out of 80 words items produced by the speakers).

Concerning item responses on the listener task, the percentage correct ranged from 36.88% to 98.75% (M = 78.29, SD = 20.37) across the different items. Dividing first between items that contained consonant or vowel targets (Consonants = 6, Vowels = 4), intelligibility scores for words containing consonant targets ranged from 64.38% (/v/) to 98.75% (/ʃ/) with a mean of 85.21%; for vowels, intelligibility ranged from 36.88% (/i/) to 97.50% (/ey/) percent correct with a mean of 68.59%.

Overall, the three most intelligible target items sheet (98.75%), lagging (98.13%), and taste (97.50%), in which target segments were /ʃ/, /g/, and /ey/ (where the substitutions /s/, /k/, /e/ respectively might have been produced) showed 0% error. The item ripping (/p/ contrasted with /b/) also showed high intelligibility (94.38%) with 0% error. The three least intelligible items were bit (36.88%), sand (50.63%), and vote (64.38%), where target segments /ɪ/, /æ/, and /v/ (contrasting with /i/, /ɛ/, and /b/) were of rather higher percentage error: 70%, 60%, and 25%, respectively. This is closely followed by the item rhyme (target vowel /ɪ/ contrasting with /l/) which also showed lower intelligibility (68.13%) with a 35% error rate. It is worth noting that the target item song (/ɔ/ contrasting with /a/) obtained a rather higher intelligibility score (89.38%) despite the highest percentage error (70%). On the other hand, fan (/f/) obtained a lower intelligibility score (87.50%) despite being produced accurately (0% error) by the speakers.
As shown in Figure 4, 16 out of 20 speakers produced items containing target consonants more intelligibly than those with target vowels (equally intelligible for Speakers 4, 14, and 19). A paired t-test revealed significant differences in speakers’ mean intelligibility scores between consonant targets and vowel targets \((p < .001)\), with a strong effect size \((d = 1.146)\). Speakers’ intelligibility for each consonant target ranged from 60.42% to 100% percent correct \((M = 85.21\%)\); for vowels, intelligibility ranged from 56.25% to 93.75% percent correct \((M = 68.59\%)\) (presented in Table 4).
Table 4

Each speakers’ overall intelligibility scores, mean scores for consonant and vowel targets, number of segmental errors produced by speakers

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Listener-Based Intelligibility (%)</th>
<th>Target Phoneme Production errors (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Consonants</td>
<td>Vowels</td>
</tr>
<tr>
<td>1</td>
<td>60.42</td>
<td>56.25</td>
</tr>
<tr>
<td>2</td>
<td>68.75</td>
<td>65.63</td>
</tr>
<tr>
<td>3</td>
<td>72.92</td>
<td>71.88</td>
</tr>
<tr>
<td>4</td>
<td>75.00</td>
<td>75.00</td>
</tr>
<tr>
<td>5</td>
<td>83.33</td>
<td>62.50</td>
</tr>
<tr>
<td>6</td>
<td>83.33</td>
<td>65.63</td>
</tr>
<tr>
<td>7</td>
<td>83.33</td>
<td>65.63</td>
</tr>
<tr>
<td>8</td>
<td>89.58</td>
<td>59.38</td>
</tr>
<tr>
<td>9</td>
<td>91.67</td>
<td>56.25</td>
</tr>
<tr>
<td>10</td>
<td>91.67</td>
<td>56.25</td>
</tr>
<tr>
<td>11</td>
<td>91.67</td>
<td>56.25</td>
</tr>
<tr>
<td>12</td>
<td>79.17</td>
<td>75.00</td>
</tr>
<tr>
<td>13</td>
<td>89.58</td>
<td>65.63</td>
</tr>
<tr>
<td>14</td>
<td>81.25</td>
<td>81.25</td>
</tr>
<tr>
<td>15</td>
<td>97.92</td>
<td>62.50</td>
</tr>
<tr>
<td>16</td>
<td>93.75</td>
<td>75.00</td>
</tr>
<tr>
<td>17</td>
<td>100.00</td>
<td>65.63</td>
</tr>
<tr>
<td>18</td>
<td>93.75</td>
<td>75.00</td>
</tr>
<tr>
<td>19</td>
<td>87.50</td>
<td>87.50</td>
</tr>
<tr>
<td>20</td>
<td>89.58</td>
<td>93.75</td>
</tr>
</tbody>
</table>
The Impact of Segmental Accuracy on Intelligibility

Discussion

Up until now, previous intelligibility studies have provided little guidance to L2 pronunciation teachers on determining the most critical sounds to teach in the classroom. As a central goal of pronunciation teaching has changed to help students make themselves understood, findings of the present study may suggest pedagogical solutions for achieving this aim. This study was designed to help teachers make informed decisions for their pronunciation teaching by empirically identifying problematic speech sounds that are most likely to affect intelligibility for listeners. Specifically, the main goals of this study were (a) to examine the relationship between segmental errors and Korean learners’ English intelligibility and (b) to identify the phonemes that are most difficult for Korean learners to produce intelligibly based on the native listeners’ responses to target items.

Concerning the first research question, the correlation analyses found a strong relationship between the segmental errors produced by the speakers and their measured overall intelligibility scores calculated as listeners’ percentage of correct responses on the minimal-pairs task. The overall strength of the correlation with speakers’ intelligibility scores was strong ($r = -0.750$), suggesting that speakers with fewer segmental errors tend to be perceived as more intelligible by native-English listeners. While the number of vowel errors significantly but rather weakly correlated with intelligibility, consonant errors showed a significant and strong negative correlation with the overall intelligibility of the speakers ($r = -0.790$). The stronger negative correlation between the consonant errors and the intelligibility scores should be largely attributed to the fact that the Korean participants generally made a greater number of vowel errors, while consonant errors were less frequent (e.g., they made no error on four out of six target consonants). Accordingly, consonant errors may have led to a seemingly greater impact on the
overall intelligibility of the speakers. These results appear contradictory to previous studies (Rogers, 1997; Bent et al., 2007), where speakers’ vowel production accuracy, but not consonant accuracy, was correlated with their overall intelligibility at the sentence level. One possible explanation for this may be a result of the number and nature of vowels under examination in the current study. This study included only 4 vowels, and the selection of these vowels was largely based on their expected difficulties for L1 Korean speakers of English. The inclusion of more vowels would possibly lead to more variation of speakers in terms of errors and intelligibility. Another explanation may be attributed to the fact that this study was conducted at the word-level, which may not reflect all the differences that may occur, especially in connected speech. However, further investigation is needed to confirm or disconfirm these hypotheses by extending the current study to include sentence stimuli as well as more vowel targets.

Regarding the second research question, the finding of the present study that most Korean speakers showed difficulty in producing vowel sounds intelligibly compared to consonant sounds is in accord with the findings of previous studies (Rogers, 1997; Bent et al., 2007) and can be explained by language-specific factors. In terms of language-specific factors, a negative L1 transfer could account for speakers’ overall difficulty with vowel sounds, and particularly items targeting tenseness of the vowels – *bit* and *sand*, 36.88% and 50.63% correct respectively. Since the distinction between tense and lax vowels does not exist in Korean, Korean speakers tend to produce a long vowel sound for the tense vowels and a short vowel sound for the lax vowels, resulting in producing the two in a similar manner. In the same vein, low intelligibility scores for *vote* (64.38%) and *fan* (87.50%) could also be explained by L1 phonological transfer. As Korean does not have equivalent sounds of labiodentals as /v/ and /ṽ/,
speakers tend to pronounce them as /b/ and /p/, the sound they clearly have in their phonological repertoires.

However, it should be noted that *fan* was less intelligible than it was expected to be considering that it was not phonemically deviated from its’ intended target phoneme /f/ (0% error percentage) according to phonemic coding. This might be due to the recording quality of the gathered speech samples; the aspiration of /f/ might get washed out making it sound closer to /p/. Another possible explanation could be as was previously referenced that *sand* was only intelligible 50.63% of the time, the issue in intelligibility for *fan* was not actually the /f/, but the vowel /æ/, which is the same in both *sand* and *fan*.

On the other hand, as for the target item *song*, despite having a high error percentage (70%), it seems quite intelligible to the native listeners (89.38%). When 20 speakers’ production of this particular item was analyzed, both coders found that the speakers were actually pronouncing the target phoneme /ɔ/ as /o/ (a common mispronunciation among Korean speakers of English). Given the native listeners’ high familiarity with Korean-accented English (4 out of 5 on a 5-point Likert scale), they may have been already aware of Korean speakers’ common problem with pronouncing *song*, which may contribute to a somewhat high intelligibility score (89.38%) for the item. However, this hypothesis needs to be validated with further research possibly taking a qualitative approach to intelligibility studies (e.g., conducting a short informal interview with listeners after the task completion), so that a comprehensive picture can emerge about sounds that affect the intelligibility of Korean speakers of English. Additionally, it should be acknowledged that despite consulting spectrograms to settle coding disputes between the two coders, vowels are quite variable that even if they stood out as errors (i.e., phonemically substituted with another vowel; for example, /ɔ/ with /o/ in *song*), they might have been rendered
rather ambiguous (and hence somewhat intelligible on the perception of listeners) in such a way that approximates the L1 version or nearby L1 analog.

The pedagogical goal for pronunciation instruction has changed its focus away from nativeness to the more realistic and appropriate goal of intelligibility (Levis, 2005). Given that the presence of a foreign accent does not preclude speakers’ being highly intelligible, the primary concern of the pedagogic intervention should be to promote intelligibility in L2 speakers’ speech production. However, considering that a limited portion of class time has usually been allotted to pronunciation instruction (Foote et al., 2012), it is significant to know which sounds have the highest effect on speakers’ intelligibility and thus should receive the most attention in class (Munro & Derwing, 2005; Foote et al., 2012). The findings of the present study support a classroom emphasis on segmental instruction. It is hoped that this study can promote evidence-based teaching practices, utilized by language teachers to determine priorities for pronunciation instruction, and thereby enhancing English learners' communicative success.

However, the findings should be interpreted with caution given the limited number of speakers, listeners, L1 backgrounds, and particularly the target sounds represented. Considering that GA English has around 42 to 44 phonemes in total, the current study is limited including only 10 phonemes – problematic phonemes specifically targeting L1 Korean speakers of English – based on Avery and Ehrlich’s (1992) chapter. As such, it should be noted that the findings of the current study are likely most relevant to Korean learners of English, despite their potential generalizability to non-Korean speakers. Also, the findings suggest that while L1-based intuitive/pedagogical recommendations were moderately successful in predicting actual problematic sounds for L1 Korean speakers – particularly in terms of intelligibility problems they posed – not all of those sounds ended up being problems for most learners. More studies are
needed in order to guide language teachers to implement more principled approaches for pronunciation instruction, based on empirical findings of divergence in learners’ actual speech. Namely, future research is needed to validate the current findings with larger samples including more varied populations and target sounds. It is likely that some sounds considered not important in achieving intelligibility will turn out to be crucial for speakers and listeners from other L1 backgrounds.

Another limitation of the current study should be that the data were collected under experimental conditions and were limited to a word-reading task. A follow-up study is needed to investigate the relationship between the intelligibility at the word level and the sentence level, or how word intelligibility may contribute to sentence intelligibility. Further, it must be noted that these tasks are still essentially different from spontaneous communication that often serves as an overarching goal of L2 acquisition. Therefore, future research should be conducted on segmental features of nonnative English that impact intelligibility of conversational English. Overall, this current investigation suggests that there is a lot more to be uncovered in relation to speech intelligibility studies.
THE IMPACT OF SEGMENTAL ACCURACY ON INTELLIGIBILITY

References


Appendix

Minimal-pairs forced-choice task

Directions

In this task, you will listen to the audio and choose the word you heard. You will listen to the audio carefully only once. Next, you will see two words on the screen. Choose the word closest to what you heard in the audio. (There will be a total of 200 words, not including the first five practice words to get you familiar with the task.)