THE JOURNAL OF ASIA TEFL
Vol. 19, No. 1, Spring, 2022, 197-215
http://dx.doi.org/10.18823/asiatefl.2022.19.1.12.197

The Journal of Asia TEFL

http://journal.asiatefl.org/

e-ISSN 2466-1511 © 2004 AsiaTEFL.org. All rights reserved.

A Study on the Intelligibility of Korean-Accented English: Possibilities of Implementing AI Applications in English Education

Bohyon Chung
Hanbat National University, South Korea

Hyun Kyung Miki Bong
Ritsumeikan University, Japan

The current study investigates the generalizability of the Lingua Franca Core (LFC), conducting a Korean-accented English (KoE) intelligibility test with human subjects and with Artificial Intelligence mobile applications (AI apps). The data of KoE transcriptions collected from 30 native English speakers (NSs) and two AI apps were examined. The NSs group identified the words related to three LFC features significantly better than the AI apps, while with no significant difference, both groups recognized least frequently and with most difficulty the words that had non-LFC features. Despite NSs often performing better in the KoE intelligibility test, AI apps can capture speech realization at a level similar to the NSs (a moderate correlation found). The current results support the variability of both LFC and non-LFC feature sets and the novel application of the LFC analysis. Identifying difficult (less/least intelligible) English phonemes of KoE to NSs and to AI apps, we recommend that the variability of the LFC feature set should be taken into account in English language education and that learners should aim to communicate intelligibly with humans with various accents, by practicing/communicating with current available AI apps to make ready for the AI-Intervened/Mediated Circle.

Keywords: intelligibility, Korean-accented English, pronunciation teaching, computer-assisted language learning, mobile-assisted language learning

Introduction

In South Korea, where digital literacy is stressed, the applicability of computer-assisted language learning (CALL) and mobile-assisted language learning methods to the teaching and learning of second or foreign language (L2) pronunciation has been explored continuously. A wide range of methods have been applied and researched, from the earlier method of drills (A. Kim, 2012; Kim & Moon, 2013; Ko & Jung 2021) to communicative approaches that use computerized feedback (Carey, 2004; H. Kim, 2012; Kim & Lee, 2015). More recently, many studies have attempted to incorporate automatic speech recognition (ASR) technology into the teaching and learning of L2 pronunciation: Validation studies of ASR have been conducted by measuring recognition accuracy levels (Kim et al., 2011; Park et al., 2016): learner performances were compared after pronunciation training with ASR-powered devices (Inceoglu, Lim, & Chen, 2020; Kim, 2006), and existing ASR-based programs for L2 teaching and learning in Korea were reviewed (Kim & Jung, 2018). In respect that positive attitudes are often reported as having a close relation to learning outcomes (Gardner, 1985), the number of previous studies suggest that demands

The Journal of Asia TEFL
Vol. 19, No. 1, Spring 2022, 197-215

for implementation of CALL approaches are legitimate in Korea. It was indicated that Korean-speaking learners (KSLs) of English hold generally positive attitudes toward CALL instruction using an ASR-featured software called “Dr. Speaking” (Sung, 2008), using flipped classroom (Copeland, 2019), or gathering pronunciation feedback from multimedia sources (Yoon & Lee, 2009). The fact that KSLs have an affinity with CALL methods thus seems to be making South Korea an optimal place to employ the methodology.

English has become a lingua franca (ELF), being widely used by multinational speakers. Considering that the number of fluent English speakers in the Expanding Circle exceeds those in the other two circles put together (Crystal, 2003), ELF interactions are characterized by being “fluid, hybrid, and evolving in nature” (Dewey, 2007). Within the vast body of ELF research, pronunciation issues have received significant attention (Deterding, 2013). Jennifer Jenkins (2000) has analyzed pronunciations causing misunderstandings in the ELF conversations between Japanese and Swiss-German. By organizing what appeared to be required for retaining intelligibility in ELF interactions, Jenkins (2000) specified these Lingua Franca Core (LFC) features as a list of pronunciation features to promote mutual intelligibility in the ELF context in which two or more non-native English speakers communicate each other. There were also certain features that Jenkins (2000) distinguished as having no influence on intelligibility (hereafter known as non-LFC features). According to Jenkins (2000), the LFC features that affect intelligibility are worth teaching to language learners, while non-LFC features, which are less related to intelligibility in ELF interactions, may not be worth learning. It seems worth pursuing to what extent Jenkins' suggestion can be considered to explain the intelligibility of Korean-accented English (KoE), defined as part of the Expanding circle under Kachru’s influential Three-Circle model (Kachru, 1985). Is it true that LFC features are worth teaching for KSLs? Or is it true that non-LFC features are not noteworthy to stress in L2 pronunciation learning and teaching in South Korea?

Indeed, although Kachru’s (1985) Three-Circle model played a crucial role in accounting for world English varieties, it is questionable whether the model reflects the current status of CALL integration in English language teaching and learning. To provide evidence on the feasibility of implementing state-of-the-art technology, this study explores whether cross-agent interactions between humans and AI-powered mobile applications (AI apps) have the potential to add a new dimension to the Three Circles of World Englishes. AI apps are implemented to test the intelligibility[1] of native Koreans who lack chances to encounter the Inner Circle varieties of English. Based on the LFC features proposed by Jenkins (2000), the current study compared the intelligibility test results between the AI apps and 30 native English speakers (NSs) to highlight the feasibility of AI apps in L2 pronunciation practice in South Korea and promote the intelligibility of KoE.

Literature Review

Tackling Challenges in Teaching and Learning Pronunciation Using ASR

L2 pronunciation teaching and learning via CALL programs has been studied to address the subjectivity of human raters and the efficiency issues of transcribing speech by utilizing ASR technology (Bahari, 2021), and to provide clear evidence of pedagogical advantages. For example, Delmonte (2002) developed a prototype of an interactive multimedia self-learning linguistic software called SLIM, which relies on ASR features to generate feedback on both writing and speaking proficiency. In addition, Wang and Young (2014) established a program that combines ASR technology with multiple layers of corrective feedback, known as intelligent computer-assisted speaking learning (iCASL). Despite the generally positive effects found in prior literature, several limitations have been pointed out. According to

[1] In this study, we follow the definition of intelligibility of Smith and Nelson (2006), referring to the recognition of the word and utterance-level.

The Journal of Asia TEFL
Vol. 19, No. 1, Spring 2022, 197-215

Kim and Jung (2018), these ASR-based programs that were regarded as providing unbalanced language models (either general American or Received Pronunciation English), offer access to a limited range of content for learners to practice, and require high maintenance fees to pay for upgrades or support. Indeed, while ASR-based software offers personalized feedback (e.g., Dr. Speaking, iCASL, SLIM), it is often challenging for learners to receive timely feedback, due to the low accessibility of programs, which are offered for a limited time and place only. In turn, those programs may not be entirely effective at developing learners' agency or autonomy. In addition, pre-selected English pronunciation models in the programs can quantify and measure learners' pronunciation (Delmonte, 2002; Sung, 2018); however, without considering learners' preferences and social contexts, the goal for pronunciation training has been to emulate the pronunciation of NSs (Kim & Jung, 2018) rather than to achieve intelligible pronunciation for effective communication.

Given the challenges, the ASR technology implemented in AI apps (such as Google Assistance and Apple Siri) seem to be a viable option for L2 pronunciation training. While ASR-based software would require students to repeat minimal pairs that are chosen in advance (Delmonte, 2002), AI apps are capable of understanding speech from the sentence-level utterance, and then respond accordingly (i.e., recognition accuracy of voice commands for Google Assistant was 93% and Apple Siri was 83%; see Munster & Thompson, 2019). This promotes similar collaborative interactions between human beings. Furthermore, several studies revealed the effectiveness of using AI apps with ASR capabilities in L2 pronunciation teaching and learning. Guskaroska (2020), for example, compared the transcriptions created by AI apps on mobile devices and human listeners for vowel sounds. The findings in the study (Guskaroska, 2020) suggest that the transcriptions from the AI apps have closely matched human raters' judgments, which legitimates the use of transcriptions as feedback. Additionally, Inceoglu et al. (2020) found that three weeks of receiving implicit feedback from AI apps helped KSLs to improve frontness of the vowel /i/. Also, when applying ASR and text-to-speech resources together, Mroz (2018) observed that students consider the feedback from the ASR-based technologies useful to improve intelligibility of targeted phonemes. In this light, AI apps take advantage of the fact that students born between 1997 and 2012 (Generation Z, Gen Z, Dimock, 2021) are already familiar with their platforms at no additional cost. Thus, the advantages offered by AI apps are likely to offset the downside of existing ASR-based pronunciation training software.

Achieving Intelligibility in Lingua Franca Communication

Exploring what aspects of L2 pronunciations are important to produce and understand "correctly" or "accurately" and what causes misunderstandings in ELF conversations, Jenkins (2000) has analyzed pronunciations of Japanese and of Swiss-Germans derived from 40 tokens, and specified LFC features as a list of pronunciation features that are required to retain "intelligibility," distinguishing them from non-LFC features that have no influence on intelligibility. Jenkins outlined the LFC features (2000) and reaffirmed them in the following studies and books (Jenkins, 2002, 2006, 2007). LFC and non-LFC features are illustrated under similar categories in Table 1.

TABLE 1
LFC and non-LFC features

Category	LFC Features	Non-LFC Features
Consonant inventory	Most consonants	Dental fricatives and dark /l/
Consonant cluster	Word-initial and -medial clusters	Word-final clusters
Vowel	One vowel sound /ɜː/ and vowel lengths	Vowel quality
Prosody	Tonic (nuclear) stress placement	Word stress

Note. Adapted from Jenkins (2000)

Various researchers have confirmed Jenkins' LFC proposals and suggested that the LFC features are essential to assure intelligibility among Outer and Expanding Circle speakers. For example, Deterding

Vol. 19, No. 1, Spring 2022, 197-215

and Kirkpatrick (2006), described phonological characteristics that may be associated with English in 10 countries in southeast Asia. Drawing upon a 20-minute conversation in six groups of three to four speakers from different countries, they identified commonalities of the pronunciation features among the speakers including the replacement of dental fricative /θ/ by [t], reduced initial aspiration, lack of reduced vowels, and heavy final-stress. Interestingly, Deterding and Kirkpatrick (2006) claimed that the shared features in the emerging Asian Englishes did not contribute to communication breakdown, but the unfamiliar pronunciations, all of which were LFC features did (e.g., initial consonant sounds, cluster reduction). In a similar vein, Chung (2013) investigated natural speech by paired speakers who differed in their language backgrounds. By focusing on the features of KoE, Chung (2013) identified the segmental substitution mistakes that may lead to reducing intelligibility, largely in line with the LFC features. Chung et al. (2016) further stipulated the problematic features of KoE, indicating that a significant decrease in KoE intelligibility occurred when the coda voicing effects on the preceding vowels were rare, and obstruent sounds were replaced with glides and liquids, which were also in line with the LFC features.

While the reviewed studies provide evidence that the LFC features put forward by Jenkins (2000) are necessary to achieve intelligibility in different varieties of English, there are also a number of studies that seemingly undermine the fixed set of the LFC features. Deterding and Mohamad (2016) outlined five varieties of English in Asia, and evaluated the potential for applications of LFC in each context. Their findings indicated that certain features (e.g., vowel quality, voicing dental fricatives) may have been under-recognized among LFC features for east and southeast Asian countries, as those features reduced intelligibility in ELF settings (Deterding & Mohamad, 2016). Further, Field (2005) indicated that intelligibility was largely determined by the direction of stress movement and vowel quality, which contrast with the LFC proposals. In addition, Low (2016) investigated the monophthongs (vowels) and rhythm of speakers from five Asian countries and pointed out that certain vowel sounds may hamper intelligibility, which disagrees with the LFC features. Similarly, Chung and Bong (2021) indicated that the vowel qualities (intelligible front and central vowel sounds) would improve the overall KoE intelligibility when testing four listener groups with different degrees of KoE accent familiarity. In short, these studies claimed that problematic phonetic features that cause “misunderstanding” may vary depending on the varieties of English, suggesting a modification of the LFC features.

Recapitulation and Research Questions

We have discussed that LFC features are crucial when Asian English speakers are to produce and understand intelligible utterances by speakers of Asian English (Chung, 2013; Chung et al., 2016; Deterding & Kirkpatrick, 2006), and this is also true of non-LFC features to some extent (Chung & Bong, 2021; Deterding & Mohamad, 2016; Field, 2005; Low, 2016). For example, the word stress and vowel qualities in the non-LFC inventory are important for maintaining intelligibility, a finding that is inconsistent with the LFC proposals (cf. Chung & Bong, 2021). This line of argument leads to tentatively postulating the variability of the inventory of both LFC and non-LFC. It seems worthwhile to test the validity of the arguments, examining Asian Englishes, such as KoE. Indeed, previous studies on which of the KoE pronunciation features should be prioritized when trying to enhance intelligibility have shown confusing results. On the one hand, a couple of studies (Chung, 2013; Chung et al., 2016) found that most LFC features in KoE cause problems, and suggested that therefore they should be trained for improving KoE intelligibility. On the other hand, certain non-LFC features in KoE (i.e., vowel qualities) could be recognized only with difficulty by speakers from different first language (L1) backgrounds, which resulted in a loss of intelligibility: that is necessary to gain intelligibility (Barrass et al., 2020; Chung & Bong, 2021; Lim et al., 2016). Thus, it seems reasonable to pursue further studies on KoE intelligibility to see whether the inventory of the LFC and that of the non-LFC vary or not.

Furthermore, based on this previously discussed research on the use of ASR-powered AI apps, we reasoned that the use of ASR-powered AI apps instruction provides evidence to support L2 pronunciation teaching and learning. By providing useful feedback in response to learners’ speech, all these endeavors contribute toward positive impacts, both on performances (Guskaroska, 2020; Inceoglu et al., 2020) and on effective filters for L2 language learners (Mroz, 2018). However, these recent studies often rely on the acoustic analysis of specific vowels, or depend on data from a students' perception survey, which will likely yield little evidence of the extent to which KoE features affect overall intelligibility. The current study, therefore, attempts to expand a novel approach, evaluating and validating ASR-powered AI apps by applying LFC analysis (compare and contrast) on the KoE intelligibility test results obtained from human listeners and those from AI apps given the learners' affinity with digital devices in South Korea, where the importance of digital literacy is emphasized. In such a context, the comparisons would give an indication of the legitimacy of incorporating AI apps into L2 pronunciation teaching and learning.

The current experimental study, therefore, was designed to measure the overall intelligibility level of KoE by utilizing AI apps that appear to be more suitable for Gen Z, and by human subjects, namely NSs. In accordance with the LFC and non-LFC features, the transcriptions created by AI apps are compared to those from NSs. If LFC features are to ensure the mutual intelligibility between human speaker-listeners, would they still be required when communicating with AI apps? The specific research questions addressed in this study are as follows:

1. Do NSs and AI apps show meaningful differences in the extent to which they find KoE intelligible?
   A. Is there any difference in recognizing the LFC features related words?
   B. Is there any difference in recognizing the non-LFC features related words?
2. Which of the phonological features involved in LFC or non-LFC features contribute to the level of KoE intelligibility?
   A. For NSs (human listeners)
   B. For AI apps

Methods

Participants

A total of 30 college students attending a university in North Carolina in the US were recruited to take the intelligibility test. These students were from China (3.3%), India (6.7%), Ireland (3.3%), Jamaica (6.7%), the UK (6.7%), the US (70%), and Venezuela (3.3%). Some students' first language was not English, but they were equally considered NSs on account of their near-native-like English proficiency; aside from the 80% native English-speaking students, other foreign-born students (20%) immigrated to America in their childhood. Female students made up 60% of the sample (N = 18) and the other 40% of students (N = 12) were male. Their ages ranged from 18 to 54 years old (Mean age = 22). None of these participants displayed hearing disorders.

KoE Stimuli

The stimuli were digital recordings of 100 sentences that were adopted from a test battery, the Oxford Placement Test (Allan, 2004). Retesting was approved by the author. Each of the 100 items consisted of a sentence that included a target word. The following is an example item with an underlined target word: "I gather you've been having trouble with your hearing." The average sentence length was 10.56 words with an average reading level of 4.35 on the Flesch-Kincaid Grade level, meaning that they were "easy enough to read by nine- to ten-year-olds." On average, the Common European Framework of Reference (CEFR) levels of all target words in the sentences were at intermediate level, A2.

Six audio speakers from Korea recorded 16 to 17 sentences each, creating an audio with a total of 100 sentences. The students were from similar regions, such as Daejeon and Chungcheong, all located within the central areas of South Korea. The speakers were considered to have a homogeneous level in their English proficiency test performances on the Oxford Placement Test (Allan, 2004) [M = 111.7 (out of 200), SD = 15.055], which is equivalent to the level A2 in the CEFR (i.e., an intermediate level). To improve the demographic balance of the recording, female speakers recorded half of the items and the other half were recorded by male speakers. The recording was made in a quiet office using a Zoom H1n recorder, with settings of 44,100 kHz and 16-bit resolution. Each of the 100 sentences was edited to be repeated twice, yielding a total playing time of 31 minutes and 11 seconds. The KoE stimuli (the recorded audio) was uploaded to YouTube and a quick response (QR) code was created to direct participants to the YouTube audio.

Procedures

Experiment 1: Human listener intelligibility testing

In this section, the KoE intelligibility test was carried out to 30 NSs. At one of the six different time slots scheduled beforehand, 28 students who signed up in advance came to an allocated lecture room at a university in the US. Two additional KoE intelligibility tests were done by students enrolled at the same school who spontaneously wanted to contribute to the study. The test takers were provided with a pen and test paper which consists of a QR code directing to KoE listening audio and 100 sentences with blank spaces. A cloze test (a dictation test) was employed to measure the KoE stimuli while the test takers were wearing their personal headphone or earphone. During the test, test takers could not play back the audio, but they were allowed to pause the audio at their own pace, which was to minimize the negative effects of participants' fatigue on the test. Once data were collected, the target words in the test items were coded in due consideration of handwriting ambiguities. All test takers and speakers completed consent forms that were approved by the Institutional Review Board and were given a gift card for their participation.

Experiment 2: AI app intelligibility testing

In this experiment, an intelligibility test was administered to a voice-based artificial intelligence system with the same test audio. Google Assistant, the most widely used AI app (virtual assistant mobile application) among Koreans, was used to test the intelligibility of KoE; more than 73% of people in South Korea reported that they use smartphones with an Android operating system that has Google Assistant pre-installed, followed by 27% of people who use an Apple iPhone (iOS; StatCounter, 2021). In addition to its wide usage, the superior compatibility of Google Assistant over Apple Siri was taken into consideration. Google Assistant can not only be used on any Android system, but can also be run on iOS devices as a separate app. Moreover, superior performance was counted when choosing the Google Assistant. It was confirmed that Google Assistant understood 100% of voice commands and answered them correctly with 93% accuracy. However, Apple Siri only scored 99.8% for recognizing voice commands and answered them 83% correctly (Munster & Thompson, 2019).

The AI app intelligibility test was conducted twice with Google Assistant. The authors used two operating systems: Android on a Galaxy Note 9 and iOS on an iPad Mini 4. This reflected differences in the transcriptions between the multiple AI app operating devices. In a quiet room, the AI app devices were played in sequence, and were placed approximately 15cm from the speaker of a desktop PC (LG DESKTOP-V5HSKMB) that played the audio recording. Ten independent blocks were screen recorded during each AI app testing process. Each block included 10 sentences, and the overall experiment for each device lasted approximately 40 minutes in total. The transcriptions of each target word in the 100 items generated by the AI apps were coded for analysis by computer (see Figure 1).

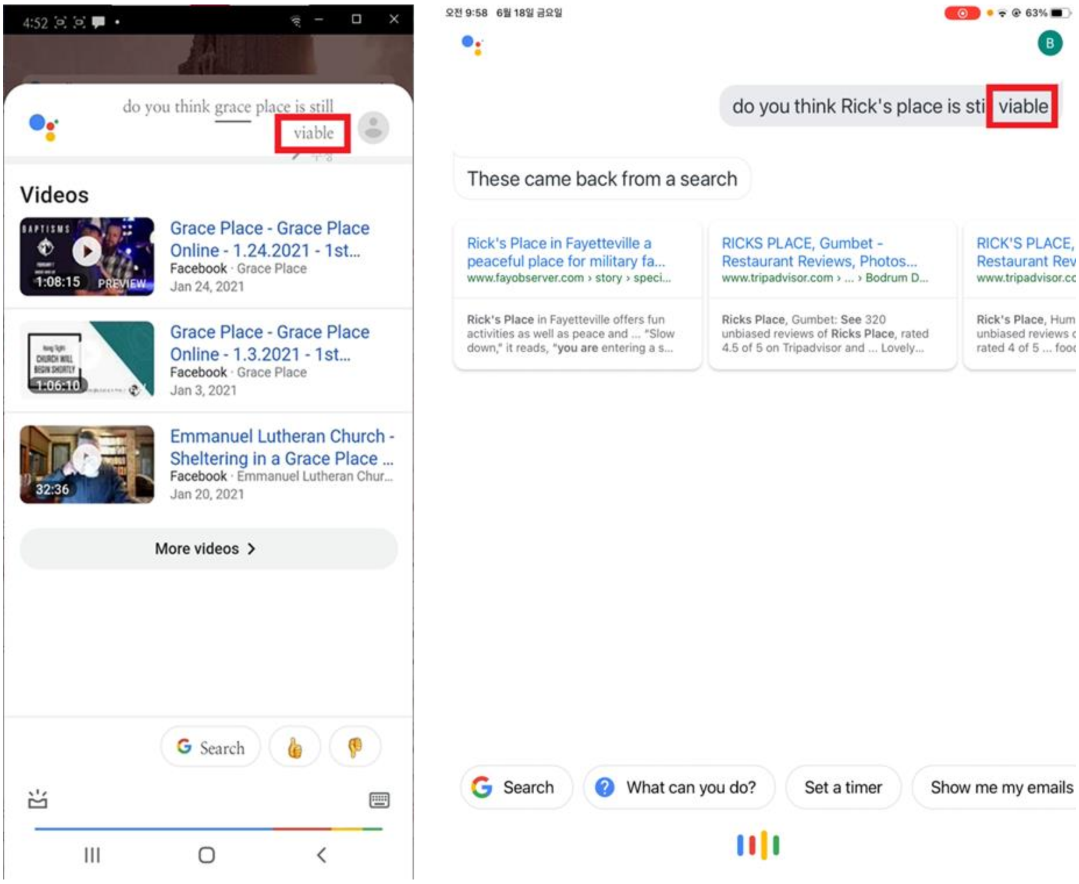

Figure 1. Screenshots of a test item (No. 41) with the target word emphasis on both devices.

Data Analysis

In each data set from both experiments, one using NSs and the other AI apps, the scripts of 100 target words were analyzed quantitatively and qualitatively. First, the target words were coded as either correct (1) or incorrect (0) answers. Then the correct percentage of items in each LFC and non-LFC feature was calculated to analyze the "intelligibility" manifested in the data. The KoE intelligibility was what we would expect given the mean scores from the two testings: (1) human listener intelligibility testing to find out how "intelligible" is KoE to NSs, and (2) AI app intelligibility testing to explore how well KoE is "recognized" with AI apps. With this respect, the terms "intelligibility" and "recognition" are used interchangeably in this study.

Thereafter, the most confused word for each item was identified to reveal the source of the problems based on the distinctive feature of the pair. To shed light on which phonological features help determine the intelligibility level of KoE, each distinctive feature was then grouped into the two categories described by Jenkins (2000). Here, our research extends the concept of LFC (Jenkins, 2000) to include non-LFC lists in order to further examine the realization of less highlighted LFC features in KoE intelligibility: [1] LFC features that are considered critical in communication between speakers of different language backgrounds, and [2] others that appeared to have relatively less importance, referring to non-LFC features in the current study.

Apparently, categorization into a single attribute was not always clear; it resulted in 35% of the items being excluded from the analysis. Items could not be categorized into one of the features for the following reasons: No common incorrect alternative was found for perfect score items or in highly incorrect answer rate items, as there were too many distractors. In addition, there were cases where the misrecognized word was different between listener groups, so that a consistent characteristic for classification was not identified. For instance, "40p" in item 40 was falsely recognized as one of four others "40, 40pee, 40page, 40 feet," which made it difficult to tag it as one problematic source. In addition, an acronym "HA" in item 81 was found to be highly intelligible, with only one inaccurate transcription, "AJ." Such cases were not included in the analysis, as they would be less likely to represent the factor. The reference variety in this study is general American.

Results

KoE Intelligibility Comparison

Overall mean correct rates

The first research question concerns meaningful differences within the correct rates of LFC and non-LFC features for both AI apps and human listeners. When compared to the overall test results between groups, findings showed a statistical difference in the mean correct rates of LFC feature words. In LFC features, AI apps resulted in significantly lower intelligibility than with the other group (NSs). The mean rates of LFC features for NSs were 59.1% and those for AI apps were 25.6% (U = .000, p = .004). On the other hand, the correct answer rates of non-LFC features were found to indicate equal difficulty for both groups. For NSs, the mean correct rates were 32.8% and for AI apps they were 22.5 (U = 13.8, p = .226).

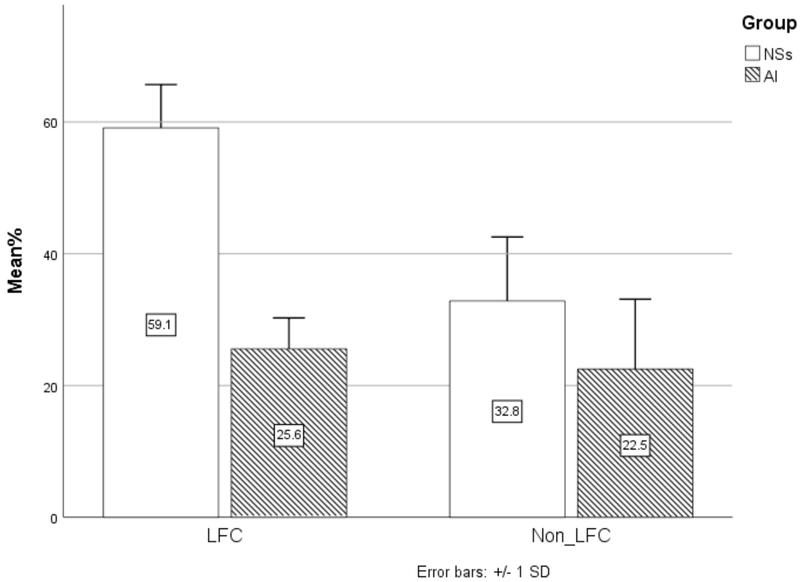

Figure 2. Mean rate differences by listener group for LFC and non-LFC features.

However, the results of the intelligibility test related to both groups showed a positive moderate correlation (r = .470, p = .000); hence, the harder some KoE features are for NSs, the more likely they would be to contribute to the poor intelligibility for AI apps, and the reverse would likely be true as well. As depicted in Figure 2, NSs generally performed better than AI apps.

LFC intelligibility

Subsequently, the test results were further broken down by four features and compared by groups of human listeners (NSs) versus AI apps. Of the 45 LFC labeled items, 29 tokens (64.4%) were included in the "Consonant" feature, as the target words had problematic consonants, except for dental fricatives and dark /l/, which were conflated with others (e.g., *buyable/viable*). Seven target words (15.6%) were grouped into the consonant "Clusters" feature, which represented misrecognitions due to word-initial or -medial consonant clusters in the tokens (e.g., *glass/class*). The other seven tokens (15.6) were categorized as "Vowel length," where confusion was created by vowel length, including the coda voicing effect (e.g., *set/said*). Regarding the last sub-feature of LFC, known as "Tonic stress," there were two target words (4.4%) that generated misrecognized pairs by placing the primary stress incorrectly (e.g., *two days'/today's*).

Based on the mean rates for each LFC feature, the tokens that had consonant "Clusters" as a predominant issue led KoE intelligibility, at 69.6%, with those of "Vowel length" coming in second at 60.3% and "Tonic stress" at 54.7%. Most misrecognitions occurred because of the "Consonant" feature, which accounted for 53.3% in the mean correct rates among the four LFC features.

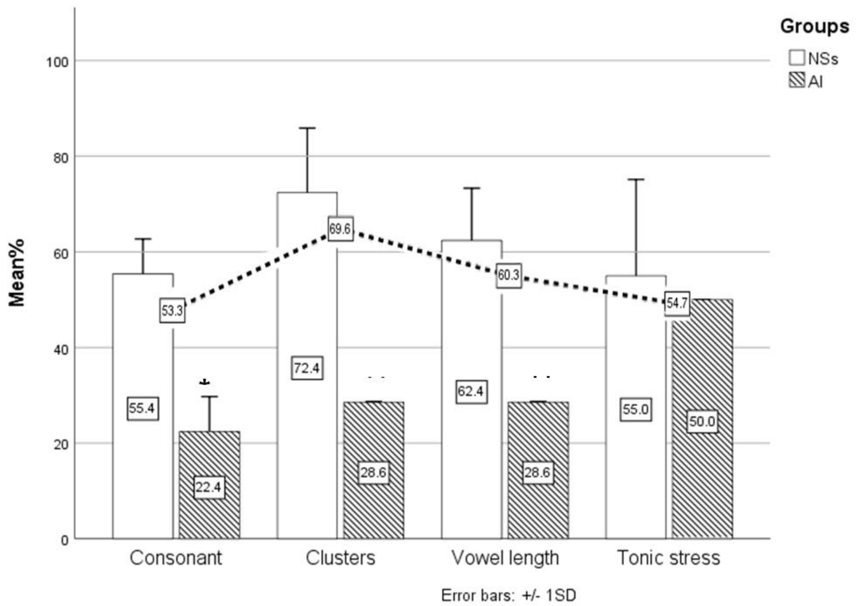

Figure 3. Mean rates for LFC features by listener group. The dotted line represents the total mean value. Asterisks (*) are used to indicate significant differences with the counterpart group. ** $p < .001$

Figure 3 summarizes the mean intelligibility rates of LFC features by testing group. As shown in Figure 3, the mean rates in response to LFC sub-features were largely varied between the two test groups. The mean correct rates of items, including consonant features, were 55.4% for the NSs and 22.4% for AI apps (U = .000, p = .004). The average rates of cluster items reached 72.4% for NSs and 28.6% for AI apps (U = .000, p = .004). In addition, the average corrections of items with vowel length differences were 62.4% for NSs and 28.6% for AI apps (U = 1.000, p = .008). For the "Tonic stress" feature, the NSs group achieved correct mean rates of 55%, while AI apps showed 50% without a statistical difference (U = 27.000, p = .847). Aside from the last feature, "Tonic stress," the mean differences of each feature between the two testing groups were found to be significant.

Non-LFC intelligibility

Next we analyze non-LFC features for each testing group. Five misrecognitions (25%) arose because of two consonant features, dental fricatives and dark /l/ (e.g., trials/triers), and three items caused misrecognitions predominantly because of consonant clusters in word-final position by being labeled with word-final clusters (e.g., clocks/clothes). As for the vowel suprasegmental and prosodic features, the vowel quality feature involved 10 target words (50%) that produced misrecognition tokens with vowel confusions (e.g., mini/many). Lastly, the word stress feature included two target words (10%) that produced erroneous examples when different syllables were stressed (e.g., personnel/personal).

The total mean rates showed that the misunderstanding occurred owing to the "Word stress" feature, which showed the highest average correct rate of 54.7% among the four non-LFC features. The "Word-final cluster" feature reached 50% of the correct mean rates, while the other consonant feature, "Dental fricative and dark /l/," had lower intelligibility rates (41.2%). Less than 3 out of 10 (27.5%) recognized target items involving erroneous vowel quality in KoE. In Figure 4, the mean rates of each non LFC feature by testing groups are outlined.

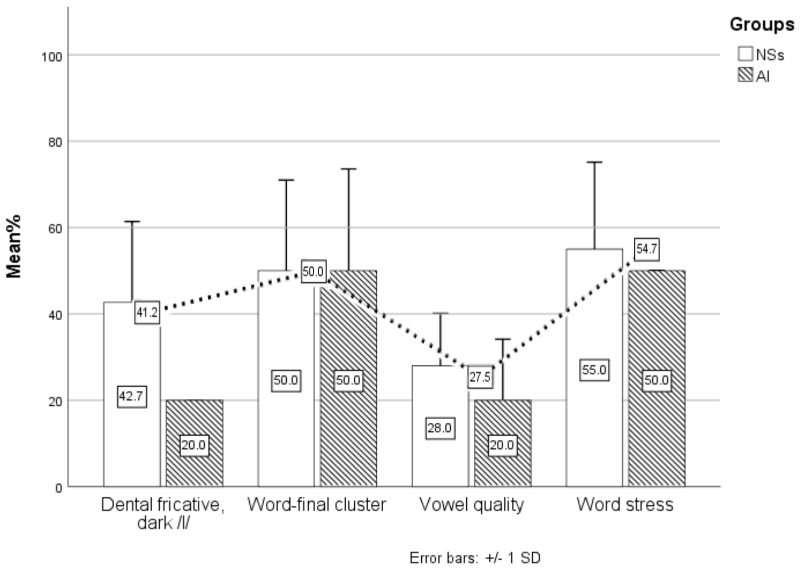

Figure 4. Mean rates for non-LFC features by listener group. The dotted line represents the total mean value.

For the non-LFC features, the mean correct rates from both testing groups suggest that they were similarly difficult to recognize. All non-LFC features were less intelligible to both NSs and AI apps. The mean rates for all sub-features were found to be less than 60%, and the differences were not significant between groups. The mean correct rates of Dental fricative, and dark /l/ were 42.7% for NSs and 20.0% for AI apps (U = 8.000, p = .101). Moreover, half the NSs and half the AI apps recognized word-final clusters correctly (U = 30.000, p = 1.000). Further, the percentage of correct rates for items that included "Vowel quality" problems was 28.0% for NSs and 20.0% for AI apps (U = 19.500, p = .444). For the final non-LFC feature, "Word stress," 55.0% of NSs and 50.0% of AI apps answered these items correctly (U = 27.000, p = .847).

Identifying KoE Phonological Features Causing Misrecognitions

Human listeners

The second research question concerns which of the phonological features involved in LFC or non-LFC features contribute to the level of KoE intelligibility.

TABLE 2
Means and Standard Deviations per Features by NSs

Features	SF	Conflation	*F*	*M%*	*SD*	*Example*
LFC [a] *Consonant*	/p/	➔ [pr]	1	93.3	.254	/ˌ**p**ɜrsəˈkjuʃən/ ➔ [ˌ**pr**isəˈkjuʃən]
	/t/	➔ [r, tr]	2	91.7	.264	/**t**aɪt/ ➔ [**r**aɪt]
	/z/	➔ [s, θs]	2	83.3	.324	/groʊ**z**/ ➔ [groʊ**θs**]
	/b/	➔ [v]	1	83.3	.379	/ˈ**b**aɪəbl/ ➔ [ˈ**v**aɪəbəl]
	/k/	➔ [g, t]	2	73.3	.435	/ˈdɛlə**k**ət/ ➔ [ˈdɛlə**g**ət]
	/s/	➔ [t, z, θ]	4	57.5	.389	/feɪ**s**/ ➔ [feɪ**θ**]
	/r/	➔ [b, l]	3	54.4	.333	/**r**ɪˈgeɪn/ ➔ [**b**ɪˈgɪnrɪ**l**i]
	/l/	➔ [r]	2	45.0	.153	/**l**æʃ/ ➔ [**r**æʃ]
	/ʤ/	➔ [ʧ]	1	36.7	.490	/**ʤ**ɔɪz/ ➔ [**ʧ**ɔɪs]
	/h/	➔ [k]	2	26.7	.254	/**h**æʧ/ ➔ [**k**æʧ]
	/d/	➔ [t]	3	17.8	.290	/ˈ**d**ɛnɪs/ ➔ [ˈ**t**ɛnɪs]
	/v/	➔ [f]	1	6.7	.254	/sti**v**/ ➔ [stɪ**f**]
	/ʒ/	➔ [z]	1	0.0	.000	/ˈfriːʒn/ ➔ [ˈfrizɪŋ]
LFC [b] *Cluster*	/st/	➔ [s]	1	90.0	.305	/***st***ɑks / ➔ [***s***ɑks]
	/bl/	➔ [spl]	1	80.0	.407	/ˈ**bl**ɛndəd / ➔ [ˈ**spl**ɛndəd]
	/fr/	➔ [θ]	2	76.7	.343	/**fr**i/ ➔ [**θri**]
	/gl/	➔ [kl]	2	46.7	.507	/**gl**æs/ ➔ [**kl**æs]
	/sl/	➔ [l]	1	43.3	.504	/**sl**ɑt/ ➔ [**l**ɑt]
LFC [c] *Vowel length*	/eɪ/	➔ [i]	1	96.7	.183	/ʤ**eɪ**/ ➔ [ʤ**i**]
	Devoiced coda	➔ Voiced coda	2	78.3	.372	/s**ɛt**/ ➔ [s**ɛd**]
	Voiced coda	➔ Devoiced coda	4	45.8	.271	/ʃ**ɑd**/ ➔ [ʃ**ɑt**]
LFC [d] *Tonic stress*	Function words	➔ Content words	1	96.7	.183	/ʃiz/ ➔ [ʃi s**ɔ**]
	Content words	➔ Function words	1	13.3	.346	/t**u** deɪz / ➔ [t**ə**ˈdeɪz]
N-LFC [a] *Dental fricative, Dark /l/*	/ɫ/	➔[r, ɔ]	2	53.3	.458	/ˈtraɪə**l**z/➔[ˈtraɪə**r**z]
	/ð/	➔[v]	1	0.0	.000	/loʊ**ð**d/ ➔ [lʌ**v**]
N-LFC [b] *Word-final cluster*	/kst/	➔ [st]	1	96.7	.183	/ˈkɑntɛ**kst**/ ➔ [ˈkɑntɛ**st**]
	/ks/	➔ [z]	1	50.0	.509	/klɑ**ks**/ ➔ [kloʊ**z**]
	/ksts/	➔ [sts]	1	3.3	.183	/tɛ**ksts**/ ➔ [tɛ**sts**]
N-LFC [c] *Vowel quality*	/ɪ/	➔ [ɛ]	1	50.0	.509	/ˈm**ɪ**ni/ ➔ [ˈm**ɛ**ni]
	/ɛ/	➔ [æ, ɪ]	2	41.7	.464	/ʧ**ɛ**rd/ ➔ [ʧ**ɪ**rd]
	/ɔ/	➔ [ʌ]	1	10.0	.305	/l**ɔ**nʧ/ ➔ [l**ʌ**nʧ]
	/i/	➔ [æ]	1	10	.305	/ˈm**i**niəl/ ➔ [ˈm**æ**njuəl]
	/æ/	➔ [ɛ, ɜr, ɪ]	3	8.9	.228	/m**æ**s/ ➔ [m**ɛ**s]
	/ɑ/	➔ [ɛ]	1	6.7	.254	/sk**ɑ**rd/ ➔ [sk**ɛ**rd]
N-LFC [d] *Word stress*	Other than first syllable	➔ The first syllable	2	6.7	.254	/ˌpɜrsəˈ**nɛl** / ➔ [ˈ**pɜr**sɪnɪl]

Note. The dark shaded cells indicate less intelligible features. *SF* = segment features; *F* = frequency; *M%* = mean percentage correct; *SD* = standard deviation

In Table 2, the segmental sounds that cause errors in each of the LFC and non-LFC features were displayed in descending order of mean correct rates. For example, the problematic segmental sounds with the highest mean rates are regarded as intelligible, hence are shown first, and those with the lowest average rates are shown last.

Table 2 illustrates the list of distinctive features caused by misrecognition involved in the LFC and non-LFC features in the NSs test group. The table shows that the most complex instances of misunderstanding occurred in the "Consonant" feature, as there were 13 sounds that caused confusions. Otherwise, the segmental variation included in each feature ranged from one in "Word stress," to six in the "Vowel quality" feature. In the "Consonant" feature, five of the items including the aspirated stops /p, t, k/ and two of the voiced obstruents /b, z/ produced by KSLs were found to be relatively intelligible, as the mean rates exceeded 60%. On the other hand, eight consonant features were likely to hamper KoE intelligibility (M < 60%). In the consonant "Cluster" feature, /st, bl, fr/ sounds were not misheard more than /gl, sl/. Meanwhile, "Vowel length" confusions occurred when the diphthong /eɪ/ in KoE was misheard as a monophthong, /i/. Coda voicing gave rise to misrecognitions, and voiced consonants in coda positions were more likely to cause vowel length problems than voiceless codas. Regarding the "Tonic stress" feature, the mean rates tended to be reduced more when content words were not properly

Vol. 19, No. 1, Spring 2022, 197-215

stressed than when function words were unnecessarily stressed.

The non-LFC features that remained are largely less intelligible to NSs. First, the voiced dental fricative /ð/ and dark /l/ in KoE, which were suggested as being non-critical to intelligibility in ELF context (Jenkins, 2000), were difficult to recognize by NSs. For example, the target word Trails was heard as [ˈtraɪərz] with a 53.5% correct rate, and dark /l/ seemed to constitute a contributory factor. Moreover, the "Word-final cluster" feature was often not detected by NSs. Except for /kst/ as in context, /ks/ and /ksts/ in coda positions had less than 50% average correct percentage. In addition, the "Vowel quality" feature showed diverse variation of segmental features, which were all found to be less intelligible to NSs. The substitutions of three lax vowels /ɪ, ɛ, æ/, and three tense vowels /ɔ, i, ɑ/ in KoE greatly impaired intelligibility, as the mean rates were less than half. Lastly, there were two tokens for which inappropriate lexical stress placement seemed to contribute considerably to misrecognition in KoE. In the "Word stress" feature, words with stress on the last syllable or the second-to-last syllable were likely to be mispronounced as having the first syllable stress, with a 6.7% correct rate.

AI apps

Problematic phonological features for AI apps categorized by each of the LFC and non-LFC features are shown in Table 3. In the first LFC feature (the "Consonant" feature), there were 13 consonant sounds in which conflation with other sounds may have triggered misunderstanding, 6 involving voiceless consonants /ʧ, t, k, h, ʃ, s/, 4 voiced consonants /d, z, v, b/, and 2 involving liquid consonants. Moreover, the second LFC feature (the "Cluster" feature), contains four consonant clusters /gl, sl, bl, fr/ in word-initial or word-medial position, causing instances of misrecognition. In the next LFC feature ("Vowel length") the target words with either voiced or voiceless codas showed that vowel length was not properly distinguished, which led to difficulties in KoE recognition. For the last LFC feature ("Tonic stress"), content words could not be recognized by AI apps, similar to NSs.

For all the non-LFC features, AI apps are likely to misrecognize voiced dental fricative /ð/ and dark /l/ in KoE, with a 0% correct rate. Additionally, two of the "Word-final cluster" features (/kst/ and /ksts/) could not be recognized when they were simplified. In terms of the "Vowel quality" feature, realizations of three lax vowels (/ɪ, ɛ, æ/), and two tense vowels (/ɑ, i/) in KoE seem to have created problems. In the final non-LFC feature (the "Word stress" feature), regardless of the lexical stress, all polysyllabic words in KoE received primary stress on the first syllable, which resulted in 0% average correct rates.

Vol. 19, No. 1, Spring 2022, 197-215

TABLE 3
Means and Standard Deviations per Features by AI Apps

Features	SF	Conflation	*F*	*M%*	*SD*	*Example*
LFC [a] *Consonant*	/ʧ/	➔ [ʤ]	1	50.0	.707	/**ʧ**ɛs/ ➔ [**ʤ**ʌst]
	/t/	➔ [f]	1	50.0	.707	/**t**aɪt/ ➔ [**f**aɪt]
	/k/	➔ [t]	1	50.0	.707	/**k**aɪndz/ ➔ [**t**aɪmz]
	/h/	➔ [æ]	2	25.0	.354	/**h**æʧ/ ➔ [ˈækʧuəli]
	/d/	➔ [ŋ, t]	3	16.7	.236	/ˈ**d**ɛnɪs/ ➔ [ˈ**t**ɛnɪs]
	/ʃ/	➔ [s]	1	0.0	.000	/læ**ʃ**/ ➔ [leɪ**s**]
	/z/	➔ [s]	1	0.0	.000	/ʤɔɪ**z**/ ➔ [ʤɔɪ**s**]
	/v/	➔ [b, l]	2	0.0	.000	/sti**v**/ ➔ [stɪ**l**]
	/l/	➔ [d, r]	3	0.0	.000	/ˈdɛ**l**əkət/ ➔ [ˈdɛ**d**əˌkeɪt]
	/ʒ/	➔ [z]	1	0.0	.000	/ˈfriː**ʒ**n/ ➔ [ˈfri**z**ər]
	/r/	➔ [b, ju, l]	4	0.0	.000	/ˈ**r**aɪdɪŋ/ ➔ [ˈ**l**aɪtɪŋ]
	/b/	➔ [v]	1	0.0	.000	/ˈ**b**aɪəbl/ ➔ [ˈ**v**aɪəbəl]
	/s/	➔ [ŋ,z]	3	0.0	.000	/ˈprɛ**s**ɪdənt/ ➔ [ˈprɛ**z**əˌdɛnt]
LFC [b] *Cluster*	/gl/	➔ [kl]	1	50.0	.707	/ˈ**bl**ɛndəd / ➔ [ˈ**spl**ɛndəd]
	/sl/	➔ [l]	2	25.0	.354	/**sl**ɑt/ ➔ [**l**ɑt]
	/bl/	➔ [spl]	1	0.0	.000	/ˈ**bl**ɛndəd / ➔ [ˈ**spl**ɛndəd]
	/fr/	➔ [f]	1	0.0	.000	/**fr**i/ ➔ [ˈ**f**eɪməs]
LFC [c] *Vowel length*	Voiced coda	➔ Devoiced coda	3	25.0	.000	/s**i**/ ➔ [s**ɪt**]
	Devoiced coda	➔ Voiced coda	2	0.0	.000	/s**ɛt**/ ➔ [s**ɛd**]
LFC [d] *Tonic stress*	Content words	➔ Function words	1	0.0	.000	/t**u** deɪz / ➔ [t**ə**ˈdeɪz]
N-LFC [a] *Dental fricative, Dark /l/*	/ɫ/	➔[r, ɔ]	2	0.0	.000	/ˈtraɪə**l**z/➔[θri jɪ**r**z]
	/ð/	➔[v]	1	0.0	.000	/loʊ**ð**d/ ➔ [lʌ**v**d]
N-LFC [b] *Word-final cluster*	/kst/	➔ [kt]	1	50.0	.707	/ˈkɑntɛ**kst**/ ➔ [ˈkɑntæ**kt**]
	/ksts/	➔ [kst]	1	0.0	.000	/tɛ**ksts**/ ➔ [tɛ**kst**]
N-LFC [c] *Vowel quality*	/ɪ/	➔ [ɛ]	1	50.0	.707	/ˈm**ɪ**ni/ ➔ [ˈm**ɛ**ni]
	/ɛ/	➔ [ɪ, aɪ]	1	25.0	.354	/ ʧ**ɛ**rd / ➔ [ʧ**aɪ**ld]
	/æ/	➔ [ɛ, ɜr, ɪ]	1	0.0	.000	/m**æ**s/ ➔ [m**ɛ**s]
	/ɑ/	➔ [ɛ]	1	0.0	.000	/sk**ɑ**rd/ ➔ [sk**ɛ**rd]
	/i/	➔ [æ]	1	0.0	.000	/ˈm**i**niəl/ ➔ [ˈm**æ**njuəl]
N-LFC [d] *Word stress*	Other than first syllable	➔ The first syllable	2	0.0	.000	/ˌpɜrsəˈ**nɛl** / ➔ [ˈ**pɜr**sɪnɪl]

Note. The dark shaded cells indicate less intelligible features. *SF* = segment features; *F* = frequency; *M%* = mean percentage correct; *SD* = standard deviation

Discussion

Which Testing Group Found KoE More Intelligible of NSs and AI apps?

It is well-established that the AI apps used in the current study are more accurate than other types of ASR-based mobile assistance (Munster & Thompson, 2019), yet the results imply that AI apps experience more problems in recognizing KoE. Between two test groups, NSs found KoE more intelligible than did AI apps. A threshold for specifying intelligibility speech is considered 60%: When the percentage of words that were understood was less than 60%, the speaker was perceived to be unintelligible; when the range was 60% to 70%, the speaker was judged difficult to understand in the previous studies (Monsen, 1981; Smith & Nelson, 2006). Based on prior studies, the findings suggest that mean rates above the accepted threshold (60%) should be considered intelligible, given that relatively high intelligibility scores were achieved by NSs. In fact, a notable difference between the two test groups in the mean rates of LFC features was demonstrated. The findings indicate that NSs would have fewer problems than AI apps when recognizing consonant sounds, consonant clusters in word-initial and -

The Journal of Asia TEFL
Vol. 19, No. 1, Spring 2022, 197-215

medial positions, and vowel length features.

Accordingly, it might be argued that human listeners (and NSs in particular), are deemed more capable of recognizing certain LFC features in KoE than AI apps, while non-LFC features do cause problems for both NSs and AI apps. This result is not surprising and is in line with the findings of a study highlighting better performance of NSs in KoE intelligibility scores than other Expanding Circle country speakers (Chung & Bong, 2021). Meanwhile, the results were partly contrary to the findings of previous studies in that AI apps or ASR-based pronunciation programs had a tolerable level of accuracy when recognizing learners' speech and were further comparable to human raters (Guskaroska, 2020; Kim et al, 2011; Park et al, 2016). Instead, the findings indicate that AI apps found KoE less intelligible than did human test takers in the context of the current study.

However, it should be noted that positive correlations with statistical significance were found between the two test groups, meaning that the more difficult KoE features are for human listeners (NSs), the more likely they are to create difficulties for AI app recognition, and vice versa. Furthermore, common incorrect responses from the two test groups were similarly observed, in that the problematic phonemes appeared to be the same. For example, the target word Dennis was misrecognized as tennis by both groups, so the /d/ was likely to be misrecognized as [t] for both groups. Also, the two test groups found /v/ difficult to recognize, as in the target word Steve. NSs heard it as /f/, as in stiff, while AI apps as dark /l/, as in still. Since the current study was only able to measure KoE intelligibility to NSs as human listeners, the findings imply that inviting more test takers from various language backgrounds may provide somewhat different results from the current study. Hence, if participants from more diverse language background had been involved in this study, the differences in the intelligibility scores from those of AI apps might have been reduced allowing the detection of a closer relationship between human listeners and AI apps, as found in previous research (Guskaroska, 2020; Kim, 2006; Mroz, 2018).

Which of the LFC and Non-LFC Features Affect KoE Intelligibility?

LFC features appear to be a frequent source of KoE unintelligibility. In the intelligibility test, 60 items that gave rise to misrecognition were categorized under one of the LFC or non-LFC features. Of these, 45 tokens accounting for 75% corresponded to misrecognition due to LFC features, while the rest of the items (25%) were associated with non-LFC features. In turn, the findings were partly in line with the LFC suggestions, in that the highest proportion of KoE misrecognition emerged due to consonant inventory. Similar to the findings of Deterding and Kirkpatrick (2006), Chung (2013), and Chung et al. (2016), such findings constitute evidence that LFC features are required to enhance intelligibility and showcase the demand for LFC features to be taught in Korea. Nonetheless, the findings suggest that the items containing LFC features can still be considered intelligible. When considering a useful point of intelligibility reference of 60% (Monsen, 1981; Smith & Nelson, 2006), two of the LFC features, initial and medial consonant clusters and vowel length, may contribute to increased KoE intelligibility. This is partially consistent with the findings of Deterding and Mohamad (2016), where some Asian varieties of English, such as Brunei, Malaysian, and Singapore English, caused misunderstandings due to consonant clusters only in the word final position.

Interestingly, the findings suggest that non-LFC features in KoE could be more difficult to recognize than LFC features, regardless of the test groups (human listeners or AI apps). Regarding the current study's observations, the voiced dental fricative and dark /l/, word-final clusters, realizations of individual vowels (i.e., three lax vowels /ɪ, ɛ, æ/ and two tense vowels /ɑ, i/), and leftward mis-stressing are often problematic sources of KoE intelligibility, as both testing groups showed lower mean scores (< 60%) for the featured tokens. As Figure 2 illustrates, the status of non-LFC features should be regarded as equally crucial as the LFC features in the context of Korea. This finding collaborates a previous study that showed that certain non-core features (e.g., vowel quality and voicing of interdental fricatives) cannot be conflated in east and southeast Asia (Deterding & Mohamad, 2016).

The Journal of Asia TEFL
Vol. 19, No. 1, Spring 2022, 197-215

Feasibility of AI Apps in L2 Pronunciation Training in Korea

One of the primary contributions of this study is its novel method of analysis by utilizing AI apps to measure intelligibility levels. It took advantage of AI apps to identify learners' difficulties in English communication by examining transcriptions generated by AI apps, and then compared those of human listeners. The findings showed that AI apps have the potential to serve as an alternative to human raters or teachers in English pronunciation education. AI apps, with combined AI and ASR technologies, could capture how speech is realized and are comparable to human listening. Although there were some features that showed statistical differences in mean intelligibility rates between NSs and AI apps, a modest level of correlation was found between them in KoE intelligibility scores. Even for the most confused responses from the two test groups, a similarity was repeatedly noticed. This finding suggests that language learners may find AI apps useful to immediately predict how their speech may be heard by other people by using their mobile devices (compared to YouTube, audio CD, or other learning materials). This interpretation agrees with that of Mroz (2018), highlighting that over 80% of participants in their experiment showed a high level of trustworthiness of ASR output, as its transcriptions proved to reflect conversation with native speakers as well. These findings were similar to those of Guskaroska (2020), where close relations were found between human judgements and ASR mobile applications. Thus, the use of AI apps on mobile devices is expected to provide useful insights into how speech may be heard by other people.

Conclusion

Applying AI apps to L2 teaching and learning could become a trend, when considering the increasing number of mobile device users worldwide. During the COVID-19 lockdown, demands for the incorporation of AI apps into L2 education have increased. The current study was carried out to determine the feasibility of AI apps in L2 pronunciation practices by comparing the results of KoE intelligibility tests taken by human listeners and AI apps. With the intelligibility test results demonstrating lower mean accuracy rates for non-LFC features than for LFC features, the findings indicate that the importance of non-LFC features should be highlighted equally with LFC features for KSLs. Another noticeable finding suggests the potential of AI apps as an effective tool in L2 pronunciation practice, as the transcriptions generated by AI apps were similar to the responses collected from human listeners. We propose key phonological features that are noteworthy when teaching pronunciation in Korea. That is, based on the observations in the previous section regarding KoE phonological features that cause misrecognitions, we explored the distribution of KoE segmental and suprasegmental features (Table 4). Unshaded phonemes reflect more vital features that may cause confusions in KoE, as both testing groups (NSs and AI apps) showed correct answer rates of less than 60% on these items on the intelligibility test. On the other hand, the shaded phonemes have equivalents or near-equivalents in Korean, and were often recognized without much difficulty.

First, as for the consonant category, the eight consonants of KoE that showed lower/reduced intelligibility (i.e., NSs and AI apps recognized only with difficulties) are illustrated in unfilled cells in Table 4: namely /s/, /h/, /d/, /v/, /ð/, /ʒ/, /r/, /l/. Except for /s/ and /h/, the other consonants, /d/, /v/, /ð/, /ʒ/, /r/, /l/, including /ɫ/ hampered KoE intelligibility to NSs as well as AI apps. Additionally, word-initial consonant cluster containing a liquid sound, /sl/, and word-final cluster simplification of /ksts/ seemed to present a problem. At the same time, there were several cases of words with different vowels from those originally intended that were misrecognized. Specifically, front vowels and an open back unrounded vowel were found to be difficult for both groups to identify in the intelligibility test, and for NSs, open-mid back rounded vowels were likely to impede intelligibility. Meanwhile, our findings suggest that long-short vowel contrasts, vowel epenthesis in word-final codas, and coda voicing effects should be secured to maintain KoE intelligibility. Lastly, for prosodic features, our findings indicate that KSLs are likely to

Vol. 19, No. 1, Spring 2022, 197-215

move the word stress leftwards, stressing initial syllables even on function words, which have been shown to lower KoE intelligibility; therefore, the features mentioned above should be prioritized in the L2 pronunciation teaching in Korea.

TABLE 4
KoE Pronunciation Targets

Consonants	p	b	t	d	ʧ	ʤ	k	g
	f	v	θ	ð	s	z	ʃ	ʒ
	m	n	ŋ	h	l (ɫ)	r	w	j
Vowels	i	ɪ	u	ʊ	eɪ			
	e	ɛ	o	ɔ	ɔɪ	oʊ		
	æ	ə	ʌ	a	aɪ	aʊ		
Additional features	· Both lexical and tonic stresses are important. · Appropriate vowel duration differences, depending on coda voicing, should be secured. · Vowel epenthesis in coda should be avoided.							

In the present study, the participation of limited human listener groups with a perfect command of the English language prevented us from considering various listeners from different language backgrounds. A previous study found that KoE was statistically more intelligible to NSs than to the shared L1 background speakers, or to those from an Expanding Circle country (Chung & Bong, 2021). This means that the participation of a variety of test takers from different language backgrounds might have reduced the average intelligibility scores and narrowed the gap with AI apps. Therefore, future studies of KoE intelligibility ought to acquire evidence from global test takers, both native and nonnative English speakers, who are experienced in ELF use.

Acknowledgements

We are grateful to Professor Hae-Young Kim, and Dr. Eun-Young Kim at Duke University for their warm support during the period of the experiment. We also would like to thank Dave Allan for permission to reuse the listening section of OPT1. A portion of this paper was presented at the World Congress of Applied Linguistics in Groningen, the Netherlands, 2021. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2020S1A5A8047381).

The Authors

Bohyon Chung is currently a lecturer teaching various English courses. Her research areas focus on the Pedagogy of English – the development of instructional design combining artificial intelligence applications, and on the areas of Applied Linguistics - Second/foreign language acquisition, English as a lingua franca, and cross-cultural discourse analysis. She is a two-time recipient of a young researcher grant for a three-year period awarded by the National Research Foundation of Korea.

Nomad College
Hanbat National University
Daejeon, 34158, South Korea
Tel: +82 42-828-8876
Email: bchung@hanbat.ac.kr

Hyun Kyung Miki Bong is Professor at Ritsumeikan University, Japan. She holds BA and MA from Nagoya University in English Linguistics and M.Phil. and Ph.D. from the University of Cambridge in English and Applied Linguistics. Her main research interests lie in practical applications of English language studies: e.g., English language learning and teaching as a foreign language, translation and interpretation, English as a lingua franca, and in the field of Applied Linguistics: i.e., first and second language acquisition.

College of International Relations
Ritsumeikan University
Kyoto-shi, 603-8577, Japan
Tel: +81-75-466-3529
Email: hkb22@fc.ritsumei.ac.jp

References

Allan, D. (2004). *Oxford placement test 1*. Oxford University Press.

Bahari, A. (2021). Computer-assisted language proficiency assessment tools and strategies. *Open Learning: The Journal of Open, Distance and E-Learning, 36*(1), 61-87.

Barrass, J. P., Baffoe-Djan, J. B., Rose, H., & Boggs, J. A. (2020). Intelligibility and comprehensibility of Korean English speakers' phonological features in lingua franca listening contexts. *The Journal of Asia TEFL, 17*(1), 124-142. http://dx.doi.org/10.18823/asiatefl.2020.17.1.8.124

Carey, M. (2004). CALL visual feedback for pronunciation of vowels: Kay Sona-Match. *CALICO Journal*, *21*(3), 1-31.

Chung, B., & Bong, H. K. M. (2021). Intelligibility of Korean-accented English: Effects of listener's familiarity. *English Teaching*, *76*(1), 33-56. https://doi.org/10.15858/engtea.76.1.202103.33

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

Chung, H., Kim, Y-K., & Lee, S-K. (2016). A study on the features of English as a lingua franca in Asian contexts: Segmental features. *Language and Linguistics*, *71*, 237-266.

Copeland, C. (2019). Using a video lecture and online discussion to flip an EFL pronunciation class. *STEM Journal, 20*(4), 121-145.

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge University Press.

Delmonte, R. (2002). Feedback generation and linguistic knowledge in 'SLIM' automatic tutor. *ReCALL 14*(2), 209-234.

Deterding, D. (2013). *Misunderstanding in English as a Lingua Franca: An analysis of ELF interactions in South-East Asia*. Mouton de Gruyter.

Deterding, D., & Kirkpatrick, A. (2006). Emerging South-East Asian Englishes and intelligibility. *World Englishes, 25*(3/4), 391-409.

Deterding, D., & Mohamad, N. R. (2016). The role of vowel quality in ELF misunderstandings. *Journal of English as a Lingua Franca, 5*(2), 291-307. https://doi.org/10.1515/jelf-2016-0021

Dewey, M. (2007). English as a lingua franca and globalization: An interconnected perspective. *International Journal of Applied Linguistics, 17*(3), 332-354.

Dimock, M. (2021, April 13). *Defining generations: Where Millennials end and Generation Z begins*. Pew Research Center. https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*(3), 399-423.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. Edward Arnold.

Guskaroska, A. (2020). Asr-dictation on smartphones for vowel pronunciation practice. *Journal of Contemporary Philology, 3*(2), 45-61.

Inceoglu, S., Lim, H., & Chen, W-H. (2020). ASR for EFL pronunciation practice: Segmental development and learners' belief. *The Journal of Asia TEFL, 17*(3), 824-840. http://dx.doi.org/10.18823/asiatefl.2020.17.3.5.824

Jenkins, J. (2000). *The phonology of English as an international language.* Oxford University Press.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics, 23*(1), 83-103.

Jenkins, J. (2006). Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly, 40*(1), 157-181.

Jenkins, J. (2007). *English as a lingua franca: Attitude and identity.* Oxford University Press.

Kim, A. (2012). Investigating the effectiveness of computer-assisted language learning (CALL) in improving pronunciation: A case study. *Multimedia-Assisted Language Learning, 15*(3), 11-33.

Kim, H-Y. (2012). The influence of pronunciation instruction through multimedia on speaking proficiency. *STEM Journal, 13*(2), 23-38.

Kim, I-S. (2006). Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation. *Educational Technology & Society, 9*(1), 322-334.

Kim, J. K., & Moon, E. J. (2013). The effect of online and coached online instructions on pronunciation and vocabulary in blended learning. *Multimedia-Assisted Language Learning, 16*(3), 39-67.

Kim, S., & Jung, H. (2018). A study on the utilization of speech recognition technology in foreign language learning applications: focusing on English and French Speech. *Journal of Digital Contents Society, 19*(4), 621-230. https://doi.org/10.9728/dcs.2018.19.4.621

Kim, S-D., Kim, W-S., & Woo, I-S. (2011). A study on the multilingual speech recognition using International Phonetic Language. *Korea Academy Industrial Cooperation Society, 12*(7), 3267-3274.

Kim, S. M., & Lee, C. H. (2015). A case study on online peer feedback for learning English pronunciation at university level. *Multimedia-Assisted Language Learning, 18*(4), 42-69.

Ko, M- H., & Jung, G. (2021). An examination of online English pronunciation programs in Korea and overseas. *STEM Journal, 22*(1), 119-153.

Lim, I., Han, J-I., Choi, T., & Lee, J-K. (2016). Speech intelligibility of Korean EFL learners: Role of proficiency, text type, and text length. *The Journal of Asia TEFL, 13*(4), 347-362. http://dx.doi.org/10.18823/asiatefl.2016.13.4.7.347

Low, E-L. (2016). Phonological patterning for English as a lingua franca in Asia: Implications for norms and practice in multilingual Asia. *Journal of English as a Lingua Franca, 5*(2), 309-332.

Monsen, R. B. (1981). A usable test for the speech intelligibility of deaf talkers. *American Annals of the Deaf, 126*(7), 845-852.

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals, 51*, 617-637. https://doi.org/10.1111/flan.12348

Munster, G., & Thompson, W. (2019, August 15). *Annual Digital Assistant IQ Test.* Loup Ventures. https://loupventures.com/annual-digital-assistant-iq-test/

Park, H-b., Kim, H. D., & Joung, J. (2016). An automatic pronunciation evaluation system using non-native teacher's speech model. *The Journal of the Institute of Internet, Broadcasting and Communication, 16*(2), 131-136.

Sung, E. (2008). The effects of computer-assisted pronunciation training in the production of English coda consonants and consonant clusters. *Multimedia-Assisted Language Learning, 11*(3), 46-68.

Smith, L. E., & Nelson, C. L. (2006). World Englishes and issues of intelligibility. In B. B. Kachru, Y. Kachru, & C. L. Nelson (Eds.), *The handbook of World Englishes* (pp. 428-445). Blackwell. https://doi.org/10.1002/9780470757598.ch24

StatCounter. (2021, July 28). *Mobile operating system market share in the Republic of Korea.* StatCounter. https://gs.statcounter.com/os-market-share/mobile/south-korea

The Journal of Asia TEFL
Vol. 19, No. 1, Spring 2022, 197-215

Wang, Y.-H., & Young, S. S.-C. (2014). A study of the design and implementation of the ASR-based iCASL System with corrective feedback to facilitate English learning. *Educational Technology & Society*, *17*(2), 219-233.

Yoon, S. Y., & Lee, C-H. (2009). A study on voice recordings and feedback through BBS in teaching and learning pronunciation. *Multimedia-Assisted Language Learning*, *12*(2), 187-216.

(Received October 11, 2021; Revised February 20, 2022; Accepted March 18, 2022)