Development of the Speech Intelligibility Index (SII) for Korean

by

In-Ki Jin

B.A., Hallym University, 2007

M.A., Hallym University of Graduate Studies, 2009

A dissertation submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of Speech, Language, and Hearing Sciences
2014

This dissertation entitled:
Development of the Speech Intelligibility Index (SII) for Korean

____
Kathryn H. Arehart, Committee chair

____
James M. Kates, Co-advisor

____
Pui-Fong Kan

____
Phillip Gilley

____
Rebecca Scarborough

Date______________

The final copy of this dissertation has been examined by the signatories, and we
Find that both the content and the form meet acceptable presentation standards
Of scholarly work in the above mentioned discipline.

IRB protocol # __13-0547__

ABSTRACT

Jin, In-Ki (Ph. D., Speech, Language, and Hearing Sciences)

Development of the Speech Intelligibility Index (SII) for Korean

Dissertation directed by Professor Kathryn H. Arehart

The purpose of this dissertation is to provide a baseline for developing the Speech Intelligibility Index (SII) for Korean. In the first study, the dynamic range (DR) of Korean speech was measured to analyze differences in DR across Korean, English, and Mandarin. Recorded sentence-level speech materials were used as stimuli. DR was quantified using different definitions of DR (defined as the range in dB from the highest to the lowest signal intensities), for several integration times (from 1 to 512 ms), and in different frequency bands (center frequencies (CFs) ranging from 150 to 8600 Hz). Across-language differences in DR were evident when considering frequency-band effects. Specifically, the DR for Korean was smaller than the English DR and the Mandarin DR in low-frequency bands (less than a CF of 455 Hz). Compared to Korean and Mandarin, the DR for English was smallest in mid-frequency bands (between a CF of 455 Hz and 4050 Hz) and was greatest in high-frequency bands (above a CF of 4050 Hz).

In the second study, band-importance functions (BIFs) for Korean were derived using procedures from Studebaker and Sherbecoe (1991) and Kates (2013). Seventy-eight native Koreans with normal hearing were tested with twenty-one low-pass (LP) and high-pass (HP) filtering conditions at varying signal-to-noise ratios (SNRs). Stimuli included 250 standardized Korean sentences presented in speech-shaped noise. BIFs produced by the two different procedures were similar. The BIF for Korean sentences showed that the frequency bands contributing most to speech intelligibility were near a CF of 350 Hz. The importance weights of the frequency bands below a CF of 630 Hz were 37.6% for the Studebaker and Sherbecoe (1991) procedure and 36.1% for the Kates (2013) procedure. Compared to published English and Cantonese BIFs for sentence and discourse materials, the present results showed that low-frequency regions (less than a CF of 840 Hz) were more important in the BIF for Korean sentences.

The accuracy of the SII for Korean may be improved by using the BIF and DRs for Korean speech. The results will provide insights into how to better optimize hearing-aid fittings for speakers of the Korean language.

ACKNOWLEDGMENTS

I would like to thank my advisor, Kathryn H. Arehart, for encouraging and guiding me through the Ph.D process and for exposing me to valuable educational and research experiences along the way. I have learned what I need to do for people who have hearing loss through her.

I would also like to thank my co-advisor, James M. Kates, for advising me to build research designs and for teaching me about depth of knowledge of hearing aids;

My dissertation committee members, Kathryn H. Arehart, James M. Kates, Pui-Fong Kan, Phillip Gilley, and Rebecca Scarborough for their advice;

My parents, Ducksan Jin and Jungsook Kim for their full support. I will never forget their dedicated encouragement.

I would also like to thank my Boulder Korean church members, KGSA members, and all friends in Boulder.

Lastly, I would also like to thank my fiancé, Sinhye Bang, for her constant support and dedication through all these days.

CONTENTS

CHAPTER

TABLES

Table

FIGURES

Figure

CHAPTER I

INTRODUCTION

The goal of this dissertation is to develop evidence that will guide future protocols for the evaluation and fitting of hearing aids (HAs) for native speakers of Korean. HAs are a common treatment for people who have sensori-neural hearing loss (SNHL). Most current HA fitting protocols are based on the English language. However, evaluation methods for HAs should consider the acoustic and linguistic properties of a patient's native language because different characteristics (e.g., acoustic properties) of languages may require different HA settings (Chasin, 2011). Thus, modification of the tools used in HA evaluations may be necessary when considering other languages (e.g., Korean and Cantonese) which differ from English in terms of acoustic and linguistic properties.

Signal audibility is a fundamental concept in amplification. One of the tools that has been applied clinically to assess audibility for selection and evaluation of HAs is the speech intelligibility index (SII) (Macrae & Brigden, 1973; Pavlovic, 1984; Wilde & Humes, 1990; RanKovic, 1991; ANSI, 1997/R2012; Beutelmann & Brand, 2006). The SII models the contribution of audible speech cues in given frequency bands to speech intelligibility (Amlani et al., 2002). The SII allows people to quantify the impact of hearing loss on the audibility of speech. Thus, the SII is able to predict how loss of audibility due to hearing loss affects speech intelligibility. The SII is calculated as the proportion of speech information that is audible across a specific number of frequency bands. The first step is to measure the audible dynamic ranges

(DRs), which is defined as the dB range from a speech peak to auditory threshold in frequency-specific bands for people who have hearing loss and for speech presented without added noise. Then, the proportion of audible speech in each frequency region (band-audibility function, BAF) is multiplied by the relative importance of that frequency region (band-importance function, BIF). Finally, the resulting values are summed across the total number of frequency bands. The result is a number between 0 and 1, with 1 indicating perfect intelligibility and 0 indicating that none of the speech was understood.

Since the SII can quantify the impact of hearing loss, the SII allows us to predict speech understanding in specific test settings (e.g., unaided and aided conditions) and compare the predicted and measured performance (intelligibility scores) of people with hearing loss. For example, the SII can be used for prediction of HA outcomes (Mueller and Killion, 1990; Rankovic, 1991; Dillon, 1993; Amlani et al., 2002; Hornsby, 2004; Killion and Mueller, 2010). By comparing the aided condition with the unaided condition of SII value, the improvement in intelligibility due to the aided gain can be predicted. Also, the SII can be used as a basis for a prescription of frequency-gain responses (Dillon, 1999; Keidser et al., 2012). For example, NAL-NL1 (National Acoustics Labs, Non-Linear, version 1) is a prescriptive procedure that uses the SII model to derive optimal frequency-gain characteristics (Dillon, 1999). NAL-NL1 is designed to amplify the full range of speech at each frequency band using the SII for listeners who have hearing loss. In addition, NAL-NL2 (National Acoustics Labs, Non-Linear, version 2), which is an advanced version of the NAL-NL1, aims to maximize speech intelligibility while keeping overall loudness no greater than that perceived by a normal-hearing listener (Keidser et al., 2012). Briefly, in order to maximize speech intelligibility and keep the loudness to normal levels, two loops are used. These loops are based on the SII and the auditory threshold and are designed to provide maximum audibility at a comfortable level. The role of the first loop is to maximize intelligibility by turning up and down individual gains in each frequency band. In this process, the frequency-gain is based on input sound levels. For example, if the input sound level is too soft, high-level gain is provided. The role of the second loop is to turn down the gain so that the overall loudness is equal to or less than normal-hearing loudness. Thus, the NAL-NL2 provides overall gain adjustment based on the normal-hearing loudness.

Components of the SII (BIF and BAF) may vary across languages. First, the BIF may depend on the language. Wong et al. (2007) examined the SII for Cantonese sentences. Intelligibility (percent keywords correct) was measured in various filtering and signal-to-noise ratio (SNR) conditions. The authors reported that low-frequency information was more important for comprehension of Cantonese speech, in comparison to English (DePaolis et al., 1996). In the case of Cantonese sentences, the most important frequencies were around 1600 Hz and the band importance weight for frequencies less than 355 Hz was 0.136 in one-third octave bands. In the case of English sentences, however, the most important frequencies were around 2000 Hz, and the band importance weight for frequencies less than 400 Hz was 0.0464 for near-octave bands (DePaolis et al., 1996).

The reason for the differences in BIFs across languages is related to differences in acoustic and linguistic characteristics. Wong et al. (2007) attributed the differences in the BIF between Cantonese and American English to be due to different phonological characteristics including the tonal nature of Cantonese. Cantonese tones carry different lexical meanings with pitch variations (Dodd and So, 1994). For example, a syllable of ‘si’ with high pitch means ‘poem,’ the same syllable with mid pitch means ‘examination,’ and the same syllable with low pitch means ‘matter’ (Browning, 1974). Because pitch variations caused by changes in fundamental frequency (F0) provide the main cues for tone perception, low-frequency information might be more important in tonal languages like Cantonese compared to non-tonal languages like English (Cheung, 1992; Wong et al., 2007).

In addition, the DR of speech, which is a key component for deriving the BAF, may also be different across languages. Boothroyd et al. (1994) showed that English phonemes vary in their frequency and level characteristics. For example, the consonant /m/ has a frequency emphasis around 250 Hz with a level variation across subjects of 15 dB. However, the consonant /s/ has a frequency emphasis around 4000 Hz with 10 dB variation in level (Boothroyd et al., 1994). If the acoustic (frequency and level) characteristics of specific phonemes differ across languages, then the speech DR may vary from one language to another. Byrne et al. (1994) showed that the DR for English was approximately 8 dB smaller than Mandarin DR at 4000 Hz (1/3 octave band). Different phoneme frequency distributions between Mandarin and English might contribute to this difference. For example, Li et al. (2007) showed the English /s/ has a higher energy concentration than the Mandarin /s/.

In summary, if linguistic and/or acoustic characteristics are different between two languages, the BIF and/or the BAF between two languages may also differ. In other words, because spoken languages have different acoustic and linguistic characteristics, the SII predictions for each language may differ. Thus, the development of an SII for individual languages is important for more adequate prescription of frequency-gain response and prediction of HA outcomes.

The SII for Korean may be different from SIIs for other languages due to across-language differences (Halle et at., 1957; Tu, 1991; Lee et al., 2005). Lee et al. (2005) reported that the frequency distributions of Korean vowels are similar to English vowels, but the frequency distributions of several Korean consonants (e.g., /b/, /p/, and /g/) are different from English consonants. For example, the Korean /b / showed a frequency range from 1072 Hz to 2543 Hz, but the English /b/ showed a frequency range from 500 Hz to 1500 Hz (Halle et al., 1957). The frequency distributions of Korean phonemes are also different from Mandarin (a tonal language). For instance, the Korean /s/ shows a frequency range from 4044 Hz to 6461 Hz (Lee et al., 2005), but the Mandarin /s/ shows a frequency range from 5000 Hz to 9000 Hz (Tu, 1991). These differences of frequency distributions of phonemes in Korean indicate that the DR of Korean speech may have different characteristics compared to other languages.

The Korean language also differs in syllable structures compared to English. In this dissertation, consonants are expressed as (C), and vowels are expressed as (V). In the case of English, the smallest syllable is (V), and the longest syllable is (CCCVCCCC) like ‘strength’ (McLeod, 2007; Smit, 2007). While the Korean syllables always contain similar numbers of consonants and vowels (e.g., (V), (CV), (VC), and (CVC), English syllables sometimes have more consonants than vowels. These different syllable structures indicate that low-frequency information may have a relatively more important role in Korean in terms of the BIF compared to English because there are more overall consonants in English and that consonants are mostly high frequency (Mines et al., 1978).

The purpose of this dissertation is to provide a baseline for developing the SII for Korean. In the first study, the DR of Korean speech was measured to analyze differences in DR across Korean and other languages including English and Mandarin. In the second study, the BIFs were derived using an established protocol (Studebaker and Sherbecoe, 1991) and a recently published method (Kates, 2013). The accuracy of the SII for Korean may be improved by using the BIF and DRs for Korean speech. Thus, more accurate prediction for speech intelligibility will be feasible in hearing-aid fittings for Koreans.

CHAPTER II

STUDY 1

The dynamic range for speech materials in Korean, English, and Mandarin: A cross-language comparison

ABSTRACT

The purpose of this study was to identify whether differences in dynamic range (DR) are evident across the spoken languages of Korean, English, and Mandarin. Recorded sentence-level speech materials were used as stimuli. DR was quantified using different definitions of DR (defined as the range in dB from the highest to the lowest signal intensities), for several integration times (from 1 to 512 ms) and in different frequency bands (center frequencies (CFs) ranging from 150 to 8600 Hz). Across the three languages, DR was affected in similar ways with changes in DR definition and integration time. In contrast, across-language differences in DR were evident when considering frequency-band effects. Specifically, the DR for Korean was smaller than the English DR and the Mandarin DR in low-frequency bands (less than the CF of 455 Hz). Compared to Korean and Mandarin, the DR for English was smallest in mid-frequency bands (between the CF of 455 Hz and 4050 Hz) and was greatest in high-frequency bands (above the CF of 4050 Hz). The observed differences in DR across languages suggest that the best-fit DR for Korean and Mandarin may be different than the best-fit for English.

I. INTRODUCTION

Dynamic range (DR or level distribution) of speech refers to the difference between the minimum and maximum speech levels. The DR of speech is important for intelligibility. For example, if the entire DR of speech is above a listener's hearing threshold, the listener can hear all speech sounds. On the other hand, if the entire DR of the speech is below the listener's hearing threshold, the listener cannot hear any speech sounds. Currently, DR is commonly used in the computation of the band-audibility function for the Speech Intelligibility Index (SII) (French and Steinberg, 1947; Amlani et al., 2002; ANSI, 1997/R2012)). The SII models the contribution of audible speech cues in given frequency bands to speech intelligibility (Amlani et al., 2002). For instance, the SII allows audiologists to predict unaided and aided speech intelligibility in listeners who have hearing loss (Studebaker and Sherbecoe, 1993; Humes, 2002). In the SII, the band-audibility function refers to the proportion of audible speech energy that is above the listener's hearing threshold in a given frequency band (ANSI, 1997/R2012)). In the calculation of the band-audibility function, the proportion of audible speech energy that contributes to speech intelligibility may be affected by the DR of speech. For example, if the entire speech DR is considered to be 30 dB and a listener's threshold is 15 dB above the minimum level of the entire speech DR, then half (50%) of the speech is audible within that frequency band. Thus, the DR of the speech may play an important role in the band-audibility function of the SII.

The DR of speech may be affected by several factors including the DR definition, integration time and frequency band. The DR varies with how the minimum and maximum levels are defined. The maximum level of speech is determined by the level (dB Sound Pressure Level (SPL)) above which a certain percentage of speech signals is not exceeded. For example, when a 99% criterion is used to define the peak DR, then 99% of the speech signals are at or below the peak level. Similarly, the minimum level of speech is determined by the dB SPL below which a certain percentage of speech signals fall. If a 20% criterion is used for the minimum, then 20% of the speech signals fall at or below the minimum level. Different criteria for establishing maximum and minimum levels have been used in previous studies. Based on the data from Dunn and White (1940), Beranek (1947) proposed a 30 dB DR that used the difference between the maximum defined by 99% and the minimum defined by 20% (99%-20%). However, Lobdell and Allen (2007) reported a DR that exceeds 40 dB for a 99%-1% criterion.

The integration time may also be an important factor when measuring the DR. When the DR of speech is measured, the speech is divided into segments. The speech level is averaged over each segment (the segment length defines the integration time) and the distribution of the segment levels is then used to determine the DR. The level distribution depends on the integration time. Consider a short spike in the signal level. If the segment duration is also short, this spike will completely fill the segment, resulting in a segment having a very high level. On the other hand, if the segment is much longer than the spike duration, the level of the spike will be averaged over the entire segment and its effect on the segment level will be much smaller. Similarly, a short reduction in the signal level will have a greater impact on the level measured using a short segment than for a long segment. Thus, the DR estimated using short segments would be expected to be greater than the DR estimated using long segments. For example, Rhebergen et al. (2009) reported that DRs decreased from 59.1 dB for the integration time of 1 ms to 4.9 dB at 8192 ms.

The DR may also depend on the frequency. Byrne et al. (1994) showed that the DR was different in lower frequency bands compared to higher frequency bands. Using a 99%-1%

definition, the DR for English in a band with a center frequency (CF) of 400 Hz was approximately 7 dB greater than a band centered at 4000 Hz.

Finally, the DR of speech may vary across languages. Boothroyd et al. (1994) showed that English phonemes vary in their frequency and level characteristics. For example, the consonant /m/ has a frequency emphasis around 250 Hz with a level variation of 15 dB. However, the consonant /s/ has a frequency emphasis around 4000 Hz with 10 dB variation in level (Boothroyd et al., 1994). If the acoustic (frequency and level) characteristics of specific phonemes differ across languages, then the speech DR may vary from one language to another. Using a 99%-1% definition, Byrne et al. (1994) showed that the DR for English was approximately 8 dB smaller than Mandarin DR at 4000 Hz (1/3 octave band). Different phoneme distributions between Mandarin and English might contribute to this difference. For example, Li et al. (2007) showed the English /s/ has a higher energy concentration than the Mandarin /s/.

Although DRs for several languages were compared by Byrne et al. (1994), the DR for Korean was not considered. The Korean language is the official language of South Korea and North Korea and is spoken by more than 75 million people in the world as a first language (Kim & Pae, 2007). Thus, the measurement of the DR for Korean may be important for the development of the SII or hearing aid fittings for Korean listeners. We hypothesize that the DR for Korean may be different from English and Mandarin because of cross-language phoneme differences (Halle et at., 1957; Tu, 1991; Lee et al., 2005).

Lee et al. (2005) reported that the frequency distributions of Korean vowels are similar to English vowels but the frequency distributions of several Korean consonants (e.g., /b/, /p/, and /g/) are different from English consonants. For example, the Korean /b / showed a frequency range from 1072 Hz to 2543 Hz but the English /b/ showed a frequency range from 500 Hz to

1500 Hz (Halle et al., 1957). The frequency distributions of Korean phonemes were also different from Mandarin. For instance, the Korean /s/ showed a frequency range from 4044 Hz to 6461 Hz (Lee et al., 2005) but the Mandarin /s/ showed a frequency range from 5000 Hz to 9000 Hz (Tu, 1991).

The SII plays an important role in hearing-aid fittings (e.g., prescription of frequency-gain response, evaluation of amplitude strategies, and prediction of hearing aid outcomes) (Ching et al., 2010; Stiles et al., 2012; Rajkumar et al., 2013). However, SII predictions will vary depending upon the acoustic and linguistic characteristics of each language. Wong et al. (2007) reported that the Cantonese-SII is different when compared to the English-SII due to language differences, including the tonal nature of Cantonese. Therefore, it is important to consider possible language-specific differences in SIIs for hearing-aid fittings for different linguistic groups. The development of new SII functions for a specific language (e.g., Korean) requires first establishing the DR of that spoken language, because the DR is a basic building block of the SII.

The purpose of this study was to analyze differences in DR across Korean, English, and Mandarin while considering the factors of DR definition, integration time, and frequency band. In order to quantify the DR of the three languages, we considered three DR definitions including 99%-20% (used by French and Steinberg, 1947), 99%-10% (used by Byrne et al., 1994), and 99%-1% (used by Lobdell and Allen, 2007). Similar to Rhebergen et al. (2009), we also considered a wide range of integration times (10 values from 1 ms to 512 ms) including integration times similar to those used in a number of other studies (cf. Dunn and White, 1940; Byrne et al., 1994; Lobdell and Allen, 2007). Finally, we considered frequency-dependent effects by computing the DR in each of 21 auditory filters having CFs ranging from 150 to 8600 Hz (ANSI, 1997/R2012)).

II. METHODS

A. Speech materials

Recorded sentence-level speech materials spoken by male talkers were used as stimuli. For Mandarin and English, Hearing-In-Noise Test (HINT) sentences were used (Nilsson et al., 1994; Wong et al., 2007). For Korean, sentences selected and recorded as part of the development of the Korean HINT were used (Soli, personal communication) (cf. Moon et al., 2008).

Details about the development of HINT speech materials are described in the literature (Nilsson et al., 1994; Wong et al., 2007; Moon et al., 2008). Basically, recordings of HINT sentences were made by a male professional voice actor. The voice actor was instructed to maintain clarity, normal pace, and normal effort. In each list, sentences were equally difficult and phonetically balanced. Generally, the sentence structure for English and Mandarin was Subject-Verb-Object (e.g., “A boy fell from the window.” and “他买了一大包五香瓜子.”) and the structure for Korean was Subject-Object-Verb (e.g., “저는 다음 달에 이사해요.”).

B. Analysis procedure

The DR was evaluated by the following procedure. The sentences for each of the three languages were concatenated without any separating pauses and then were digitized at a 44.1 kHz sampling rate using Adobe Audition (version 3.0; Adobe Systems Inc., San Jose, CA, USA). Matlab (version R2013b; MathWorks Inc., Natick, MA, USA) was used to calculate the cumulative histogram levels. Each sentence was normalized to a root-mean-squared (RMS) level of 65 dB SPL. The concatenated set of sentences was then passed through a filter bank to perform the frequency analysis. The filter bank had 21 frequency bands, with the bandwidths and center frequencies chosen to be consistent with the SII critical-band calculation procedure (ANSI, 1997/R2012)). The band center frequencies ranged from 150 to 8600 Hz. Linear-phase finite impulse response (FIR) filters were used, with each filter having an impulse response length of 16 ms (706 samples) at the 44.1 kHz sampling rate. The output signal in each frequency band was multiplied by a complex exponential at the band center frequency to modulate the output down to baseband, and the envelope was extracted by computing the magnitude of the baseband signal. The envelope in each band was then low-pass filtered by convolving the envelope signal with a raised cosine window function. The cosine window length determined the integration time, which varied from 1 to 512 ms. The smoothed envelopes were then subsampled at a rate corresponding to 50 percent overlap of the windows, and the samples were then converted to dB SPL.

The cumulative envelope distribution levels were then computed from the dB envelope histogram. Each line in the plot (Fig. 1) represents the sound level (dB SPL) at which a certain percentage of speech signals is not exceeded within each frequency band, plotted as a function of filter CF. For example, the 99% cumulative histogram level indicates that 99% of the measured the speech signals are at or below the peak level. Thus, the DR between the level of 99% and 1% can be obtained by subtracting the cumulative histogram level at 1% from the 99% level (99%-1%). The DR was quantified using three DR definitions (99%-20%, 99%-10%, and 99%-1%) and for various integration times (1, 2, 4, 8, 16, 32, 64, 128, 256, and 512 ms).

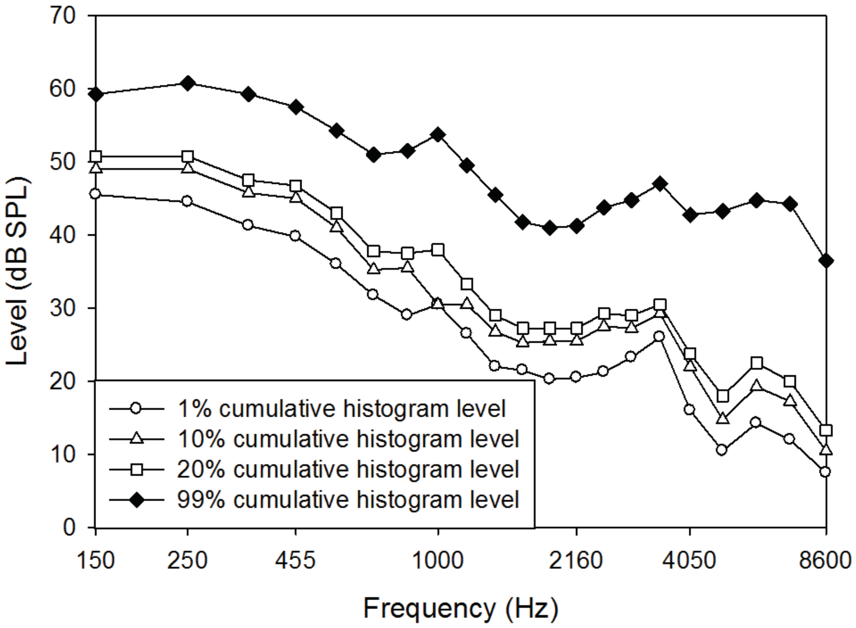


FIG. 1. Cumulative histogram levels for English sentences set using the integration time of 512 ms. Four cumulative histogram levels are shown: 1% (circle), 10% (triangle), 20% (square), and 99% (diamond).

C. Statistical analysis

Because DR values did not show a normal distribution, the nonparametric Mann-Whitney U test was used to assess whether there were significant differences in across-language DRs, which was averaged across 21 frequency bands, with consideration of the DR definition and integration time.

III. RESULTS

The effects of integration time (from 1 ms to 512 ms) for three definitions of DR (99%-20%, 99%-10%, and 99%-1%) are shown in Figure 2. The levels shown are averaged across the 21 frequency bands. There were no significant differences in DRs among the three sets of speech materials at all definitions of the DR and integration times ($p > 0.05$). DR differences were less than 4 dB in all conditions. At the same integration time, DRs for all three sentence sets decreased as the DR definition changed from 99%-1% to 99%-20%. For example, the calculated DR at 4 ms for English sentences varied from 56.75 dB for 99%-1% to 40.70 dB for 99%-20%. The DRs for the three sets of speech materials were affected by integration times ranging from 1 ms to 512 ms: for the 99%-1% condition, DR ranges from 57.78 dB for a window of 1 ms to 20.29 dB at 512 ms for Korean sentences.

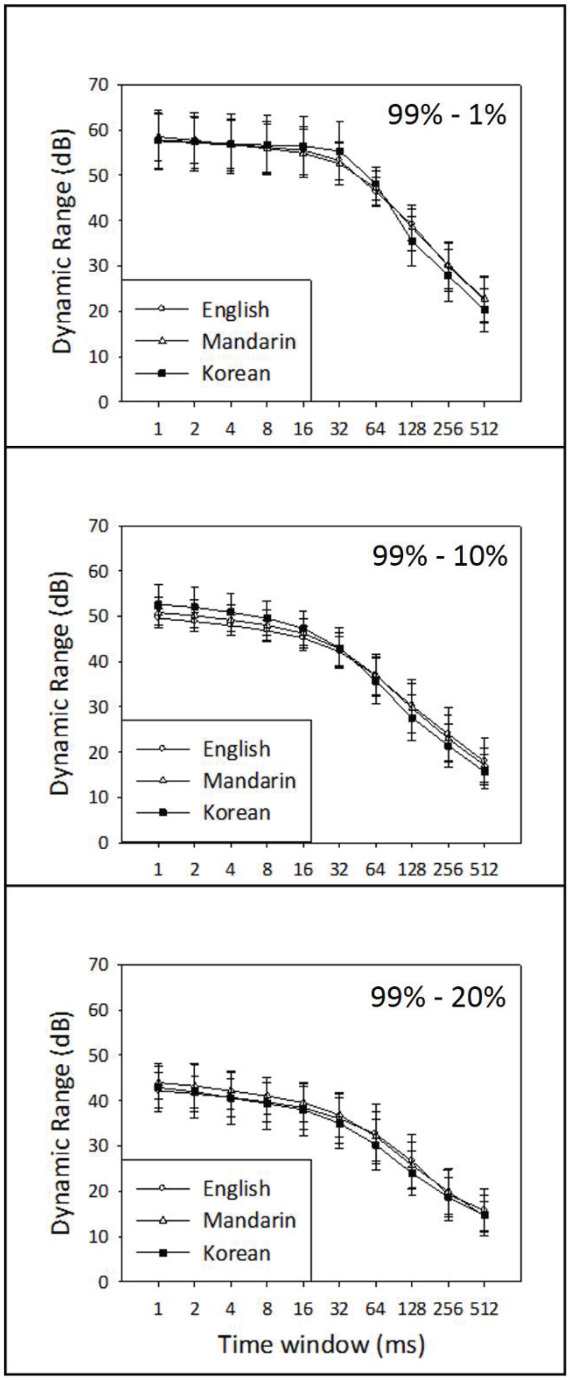


FIG. 2. The DR for three speech materials averaged across twenty-one frequency bands for ten integration times at three definitions of the DR. Three speech materials are shown: English (circle), Mandarin (triangle), and Korean (square).

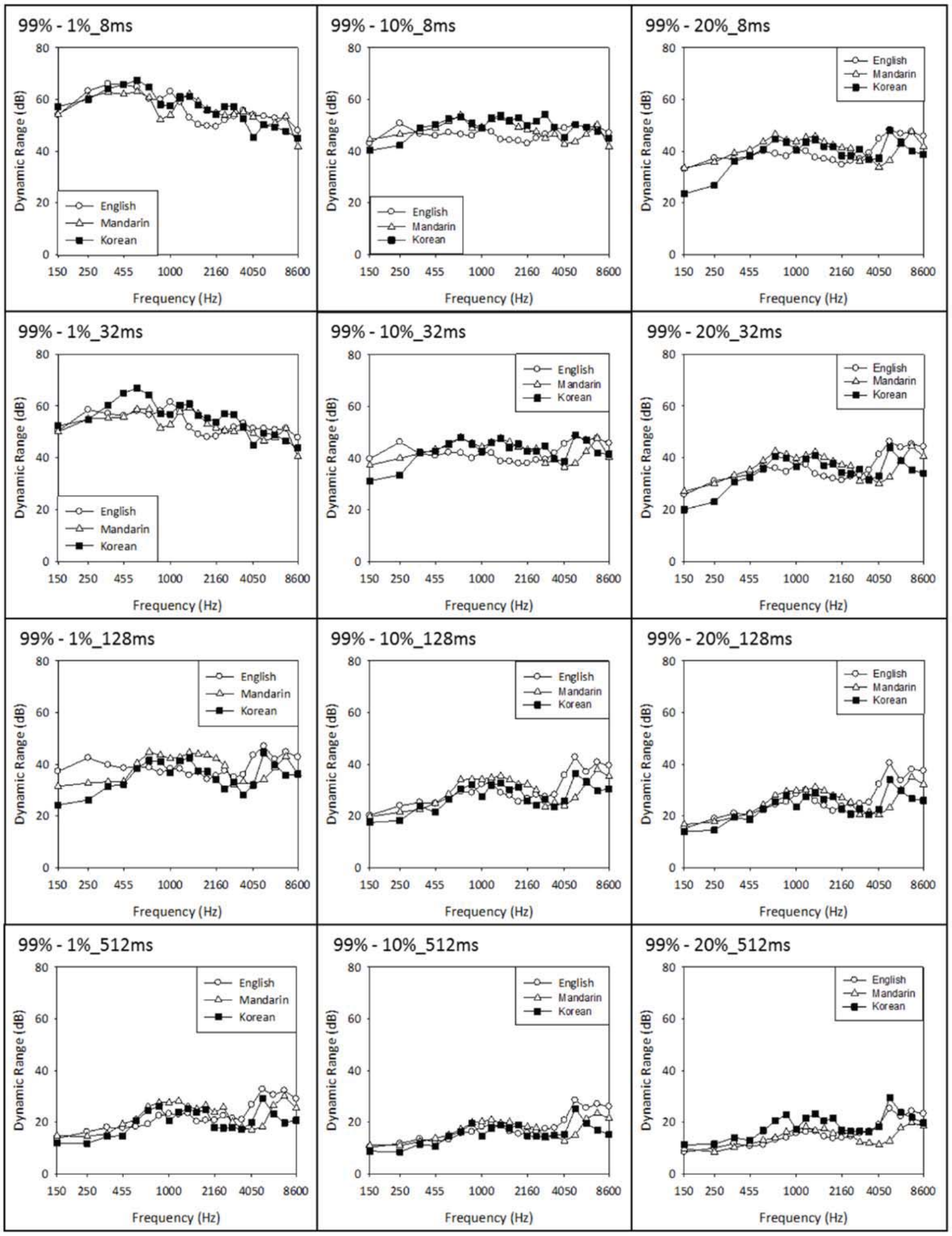


FIG. 3. The DR for three speech materials as a function of twenty-one band frequencies (150 Hz to 8600 Hz) for four integration times (8, 32, 128, and 512 ms) at the three definitions of the DR (99%-1%, 99%-10%, and 99%-20%). Three speech materials are shown: English (circle), Mandarin (triangle), and Korean (square).

Compared to Figure 2 that shows averaged level across the 21 frequency bands, Figure 3 depicts the effects of each of the 21 frequency bands (150 Hz to 8600 Hz) for the three sentence sets, for the three DR definitions, and for four specific integration times (8, 32, 128, and 512 ms). Similar to the data in Figure 2, the DR decreased as integration time increased and the DR increased as the DR definition was increased. These effects were similar across the three languages. However, the DR for different languages varied depending on frequency region. DRs for Korean were smaller than band-specific DRs for English and Mandarin in low-frequency bands (less than the CF of 455 Hz). For example, in the 99%-20% condition at 8 ms for the CF of 250 Hz, DRs for Korean, Mandarin, and English were 26.75 dB, 35.75 dB, and 37.25 dB, respectively. DRs for English were smaller than band-specific DRs for Mandarin and Korean in mid-frequency bands (between the CF of 455 Hz and 4050 Hz). For instance, in the 99%-10% condition at 32 ms for the CF of 1375 Hz, DRs for English, Mandarin, and Korean were 38.75 dB, 47.5 dB, and 47 dB, respectively. DRs for English were larger than band-specific DRs for Korean and Mandarin in high-frequency bands (above the CF of 4050 Hz). For example, in the 99%-1% condition at 128 ms in the CF of 4050 Hz, the DR for English was 43.5 dB, but DR for both Korean and Mandarin was 32 dB.

IV. DISCUSSION

In this study, the DRs for three languages (Korean, English, and Mandarin) were quantified using different definitions of DR (99%-20%, 99%-10%, and 99%-1%), for several integration times (from 1 ms to 512 ms) and in different frequency bands (from the CF of 150 Hz to the CF of 8600 Hz). The DR was affected in similar ways with changes in DR definition and integration time across the three languages. However, across-language differences in DR were evident when considering frequency-band effects. Specifically, the DR for Korean was smaller than the DRs for Mandarin and English in low-frequency bands (less than the CF of 455 Hz). The DR for English was smaller than the DR for Mandarin and Korean in mid-frequency bands (between the CF of 455 Hz and 4050 Hz). The DR for English was larger than the DRs for Korean and Mandarin in high-frequency bands (above the CF of 4050 Hz).

In the present study, the DR for all three languages varied across frequency bands (Fig. 3). For example, in the 99%-1% condition (the largest DR) at 8 ms (the shortest integration time in Fig. 3), variations of the DR for Korean were between 67.5 dB (the CF of 570 Hz) and 45 dB (the CF of 8600 Hz). The difference between the largest DR (67.5 dB) and the smallest DR (45 dB) for Korean was 22.5 dB (67.5 dB–45 dB). Also, in the 99%-20% condition (the smallest definition of the DR) at 512 ms (the longest integration time), variations of the DR for English were between 24.25 dB (the CF of 7050 Hz) and 8.5 dB (the CF of 150 Hz). The difference between the largest and smallest DR for English was 15.75 dB (24.25 dB–8.5 dB). In addition, in the 99%-20% condition at 512 ms, the differences for Mandarin and Korean were 11.22 dB and 8.25 dB respectively. This finding is similar to results from Byrne et al. (1994) (e.g., English and Mandarin) in which the DR for the 99%-1% condition at 125 ms was different from frequency bands between among 400 Hz, 1000 Hz, and 4000 Hz. For example, the differences of DR at 400 Hz and 4000 Hz were approximately 5 dB to 8 dB in English and Mandarin. The present results show that the DR for speech depends not only on the DR definition and integration time but on the frequency band.

The results of this study show that the DR varies across English, Korean, and Mandarin (Fig. 3). These results have implications for hearing-aid fittings for individuals who speak different languages. For example, because the SII depends on the DR, language-specific DRs may yield better predictions for intelligibility in the SII compared to the values used for the

ANSI standard (1997/R2012)). For instance, when the DR of speech was calculated for the 99% -10% condition at 32 ms in the CF of 250 Hz, DRs for English and Korean were 46.25 dB and 33.5 dB, respectively. In the case of the DR for English, each decibel represents 1/46.25 (2.16 %) of the audible signal range that contributes to speech intelligibility. However, in the case of Korean, each decibel represents 1/33.5 (2.98%) of the audible signal. Consider, for example, a linear amplification system in which the average signal level is amplified to be 15 dB above the impaired auditory threshold at 250 Hz. For an English speaker, approximately 82 percent of the speech would fall above threshold, while for a Korean speaker approximately 95 percent would fall above threshold because of the reduced DR in Korean. Thus, the audible signal range for the Korean speaker will be greater than for the English speaker, and the SII computed using the DR measured for the language will be higher. The desired hearing-aid gain required to achieve a specified SII value depends on both the average signal level and its DR. A lower average level may require increased gain to ensure audibility, but a reduced DR requires less amplification to place the speech minima above the impaired auditory threshold.

The speech DR used in the SII is based on a different analysis procedure (Dunn and White, 1940) than the one used in this paper. In the analysis used by Dunn and White, the signals were passed through a filter bank comprising analog infinite-impulse response (IIR) filters. The RMS level of the filter outputs was computed over a 125 ms integration time using a rectangular window, and there was no overlap between segments. To assess the effects of the differences in procedures, the measurements presented in this paper were compared to a digital simulation of the Dunn and White (1940) approach. Frequency analysis was provided by a bank of digital 21-pole IIR Butterworth filters at the same center frequencies as used for the SII critical-band procedure, and the filter order was chosen to replicate the frequency response shown in Figure 2 of Sivian et al. (1931). The filtered output was divided into 125 ms segments, and the RMS level was computed for each segment. The results of this simulation are compared to our analysis using a 256 ms raised cosine window in Figure 4. The 256-ms raised cosine window was selected for our analysis because it has approximately the same total power as the 125-ms window used by Dunn and White when integrated over the window length. The differences between the two approaches are less than 2 dB in most frequency bands.

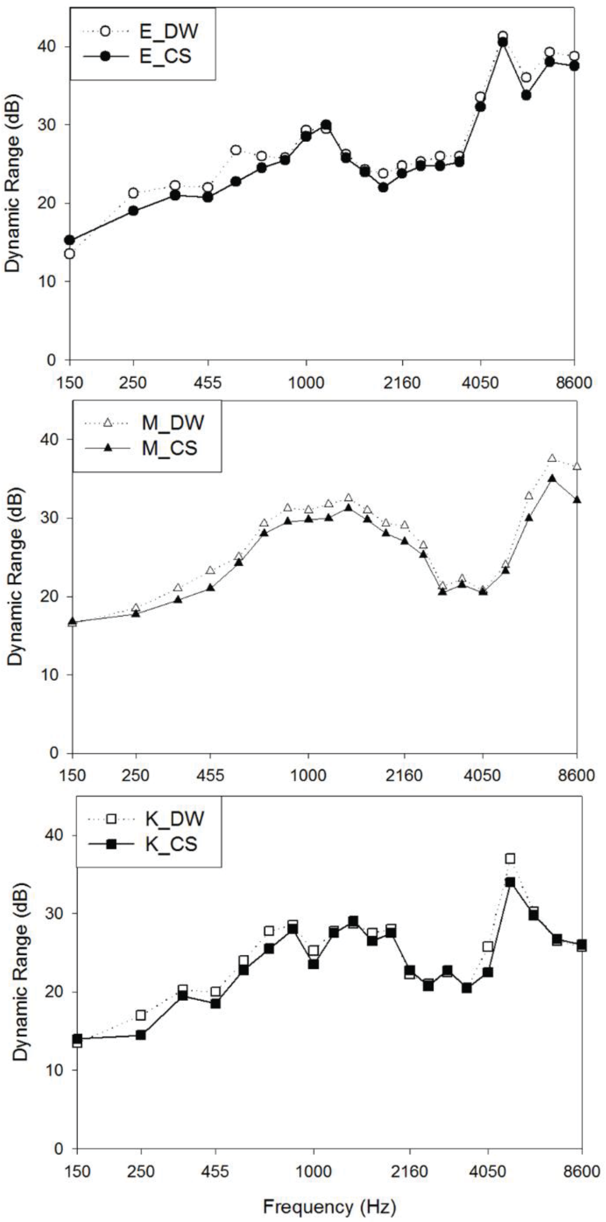


FIG. 4. The DR approached by Dunn and White (1940) (DW, unfilled shapes) and by the current study (CS, filled shapes). Three speech materials are shown: English (E, circle), Mandarin (M, triangle), and Korean (K, square).

Further study is needed before the DR measurements reported here can be integrated into the SII. In the English-language SII, the DR of speech is assumed to be approximately 30 dB, and the same DR is used for all frequency bands (Dunn and White, 1940; Beranek, 1947; Pavlovic, 1988; ANSI, 1997/R2012); Studebaker et al., 1999). While the DR currently used in the SII has been motivated by the measured DR of English, the 30-dB value is actually a "best-fit" of the SII model to the speech intelligibility data. The measurements reported in this paper show differences between the DR of Korean and Mandarin and that of English, which suggests that the accuracy of the SII for these languages may be improved by choosing a DR other than the 30 dB used for English. However, as for English, the value that provides the best SII curve fit may differ from the measured DR for the language.

The current results provide a baseline for developing the SII for different languages. First, since the DR varies across languages, the DR needs to be measured for each specific language. Second, the interactions of DR definition and integration time should be considered. The most accurate implementation of the SII for a given language may require adjusting both the integration time and DR definition, and these may differ from the values used for the ANSI standard (1997/R2012)). Third, the SII's for specific languages should consider the frequency-dependent nature of the DR. Fourth, the development of language-specific SIIs should consider a DR that has been optimized by a best-fit factor.

The observed differences in DR across languages suggest that the best-fit DR for Korean and Mandarin may be different than the best-fit for English. Thus, the findings of this study will be applied for the SIIs of different languages to predict more accurate intelligibility performance.

CHAPTER III

STUDY 2

Band-importance function for Korean sentences

ABSTRACT

The purpose of this study was to derive the band-importance function (BIF) for Korean sentences. Seventy-eight native Koreans with normal hearing were tested with twenty-one low-pass (LP) and high-pass (HP) filtering conditions at varying signal-to-noise ratios (SNRs). Stimuli included 250 standardized Korean sentences presented in speech-shaped noise. BIFs were derived using an established procedure (S&S procedure) reported by Studebaker and Sherbecoe [J. Speech Lang. Hear. Res. 34, 427-438 (1991)] and by a recently published method (Kates procedure) reported by Kates [J. Acoust. Soc. Am. 134(5), EL459-EL464 (2013)]. BIFs produced by the two different procedures were similar. The BIF for Korean sentences showed that the frequency bands contributing most to speech intelligibility were near a center frequency (CF) of 350 Hz. The importance weights of the frequency bands below a CF of 630 Hz were 37.6% for the S&S procedure and 36.1% for the Kates procedure. Compared to published English and Cantonese BIFs for sentence and discourse materials, the present results show that low-frequency regions (less than a CF of 840 Hz) were more important in the BIF for Korean sentences. This BIF will become an integral part of the speech intelligibility index (SII) for Korean.

I. INTRODUCTION

The Speech Intelligibility Index (SII) is a model describing the contribution of audible speech cues in given frequency bands to speech intelligibility (Amlani et al., 2002). The SII is a tool that allows researchers and clinicians to quantify the impact that hearing loss has on the audibility of speech in both quiet and noisy conditions. Thus, the SII is able to predict how loss of audibility due to hearing loss affects speech intelligibility (Macrae and Brigden, 1973; Pavlovic, 1984; Wilde and Humes, 1990; Beutelmann and Brand, 2006). The SII can be applied clinically in hearing-aid fittings to quantify how much audibility is restored due to amplification (Mueller and Killion, 1990; Rankovic, 1991; Amlani et al., 2002; Hornsby, 2004; Killion and Mueller, 2010). The SII can be expressed as an algebraic equation:

$$SII = \sum_{k=1}^{K} A_k I_k, \tag{1}$$

In Equation 1, K is the number of SII computational frequency bands, A_k is the band-audibility function (BAF) in frequency band k, and I_k is the band-importance function (BIF) in frequency band k (ANSI-S3.5, 1997/R2012). The BIF is a value characterizing the relative significance of different frequency bands to speech intelligibility (ANSI-S3.5, 1997/R2012). The importance of each frequency band is expressed by a weighting factor from 0.0 to 1.0 with the sum across frequency bands of the entire importance function equal to 1.0. The BAF refers to the proportion of speech energy that is above the listeners' hearing threshold in a given frequency band (ANSI-S3.5, 1997/R2012; Amlani et al., 2002).

The SII can be used to predict speech intelligibility using a transfer function (*S*), which represents the relationship between SII values and intelligibility scores. Equation 2, which produces the transfer functions with a high degree of correlation between the obtained and predicted data, was originally proposed by Fletcher and Galt (1950), and includes regression equations with fitting constants that were described by Studebaker and Sherbecoe (1991):

$$S = (1 - 10^{-\frac{PA}{Q}})^{N}, \tag{2}$$

In Equation 2, *S* is the intelligibility score, *P* is a value for the measure of the talker's and listener's proficiency, *A* is the SII value, and *Q* and *N* are fitting constants.

The BIF, which is a critical attribute in the calculation of the SII, depends on the nature of the speech materials. In English, for example, different stimuli produced different band-importance weights (Studebaker et al., 1987; Studebaker and Sherbecoe, 1991; Studebaker et al., 1993; DePaolis et al., 1996). Studebaker et al. (1987) reported that low-frequency information was more important in the BIF for continuous discourse than in the BIFs for words (Black, 1959) or nonsense syllables (French and Steinberg, 1947). In the case of continuous discourse, frequency regions of maximum importance were between 400 and 500 Hz. However, the frequency regions of maximum importance were between 2000 and 3000 Hz in the case of words (Black, 1959) or nonsense syllables (French and Steinberg, 1947).

Although the BIF varies across different types of speech stimuli, sentences can serve as a compromise stimulus (DePaolis et al., 1996). DePaolis et al. (1996) determined BIFs for English words, sentences, and continuous discourse. BIFs for the speech types were significantly different between words and continuous discourse, but the BIF for sentences was not significantly different from that of either the word or continuous discourse. This result indicates that the BIF for sentences can serve as a representative stimulus.

The BIF also depends on the language. Wong et al. (2007) determined the SII for Cantonese sentences. Intelligibility (percent keywords correct) was measured in various filtering and signal-to-noise ratio (SNR) conditions. The authors reported that low-frequency information was more important for Cantonese speech understanding compared to English (DePaolis et al., 1996). In the case of Cantonese sentences, the most important frequencies were around 1600 Hz and the band importance weight for frequencies less than 355 Hz was 0.136 in one-third octave bands. In the case of English sentences, however, the most important frequencies were around 2000 Hz, and the band importance weight for frequencies less than 400 Hz was 0.0464 for near-octave bands (DePaolis et al., 1996).

The reason for the differences in BIFs across languages can be related to differences in acoustic and linguistic characteristics. Wong et al. (2007) attributed the differences in the BIF between Cantonese and American English to be due to different phonological characteristics including the tonal nature of Cantonese. Cantonese tones carry different lexical meanings with pitch variations (Dodd and So, 1994). For example, a syllable of ‘si’ with high pitch means ‘poem,’ the same syllable with mid pitch means ‘examination,’ and the same syllable with low pitch means ‘matter’ (Browning, 1974). Because pitch variations caused by changes in fundamental frequency (F0) provide the main cues for tone perception, low-frequency information might be more important in tonal languages like Cantonese compared to non-tonal languages like English (Cheung, 1992; Wong et al., 2007).

This paper considers the BIF for Korean, which is the official language of South and North Korea (Kim & Pae, 2007). Korean is spoken by more than 75 million people in the world as a first language (Kim & Pae, 2007). The Korean language may also have a different BIF compared to other languages such as English (non-tonal language) and Cantonese (tonal language) because of differences in the characteristics of Korean (e.g., frequency distributions of phonemes and syllable structures).

The Korean language has different frequency distributions of phonemes compared to other languages (e.g., English and Cantonese). Lee et al. (2005) showed that the frequency distributions of several Korean consonants differed from the frequency distributions of English consonants. For example, the Korean /p/ showed a frequency range from 1161 Hz to 2331 Hz, but the English /p/ showed a frequency range from 550 Hz to 1500 Hz (Halle et al., 1957) (Fig. 5). The frequency distributions of Korean phonemes also differed from Cantonese. For instance, the Cantonese /u/ was approximately 355 Hz lower than the Korean /u/ in F2 (Leung et al., 1992; Lee et al., 2005) (Fig. 6). Because the frequency of occurrence of Korean phonemes differs from other languages (e.g., English and Cantonese) (Byun, 2001; Thomas, 2005), frequency importance weights for Korean may be different from other languages.

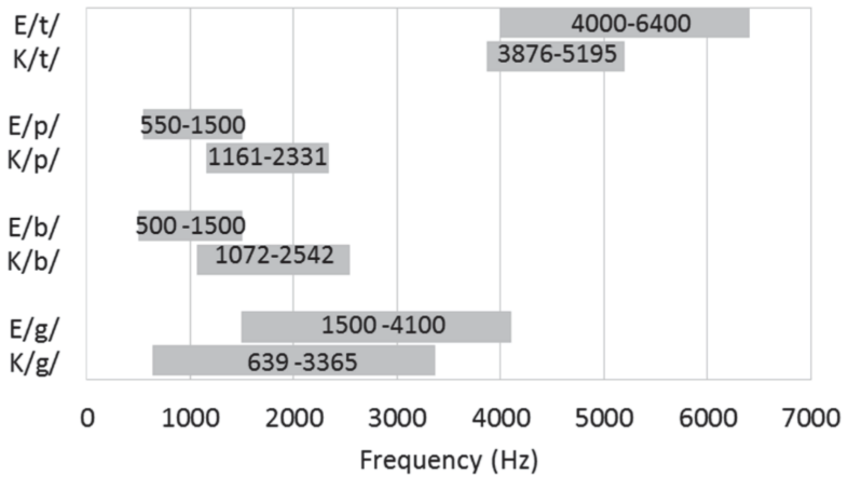


FIG. 5. Examples for frequency distributions of English (E) and Korean (K) consonants.

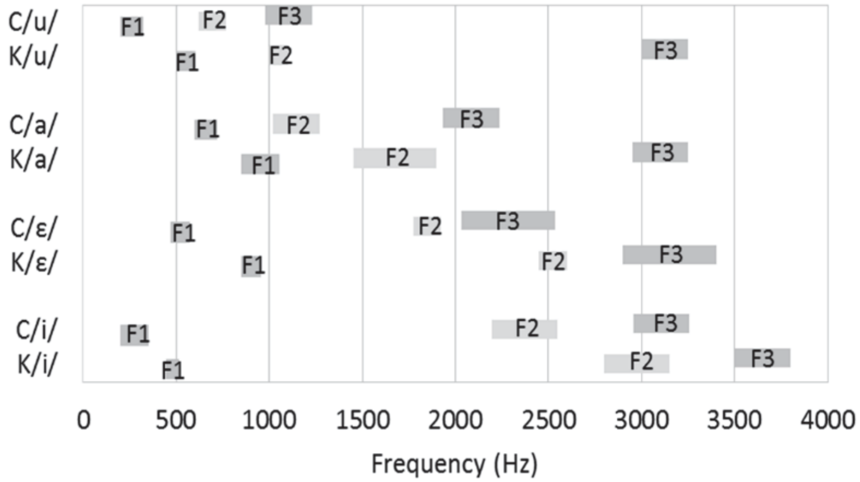


FIG. 6. Examples for frequency distributions of Cantonese (C) and Korean (K) vowels. ‘F1’ represents first formants, ‘F2’ represents second formants, and ‘F3’ represents third formants.

The Korean language also differs in syllable structures compared to English. In this paper, consonants are expressed as (C), and vowels are expressed as (V). In the case of English, the smallest syllable is (V), and the longest syllable is (CCCVCCCC) like ‘strength’ (McLeod, 2007; Smit, 2007). While the Korean syllables always contain similar numbers of consonants and vowels (e.g., (V), (CV), (VC), and (CVC)), English syllables sometimes have more consonants than vowels. These different syllable structures indicate that low-frequency information may have a relatively more important role in Korean in terms of the BIF compared to English because there are more overall consonants in English and that consonants are mostly high frequency (Mines et al., 1978).

The purpose of this study was to derive the BIF for Korean sentences and use it in the development of a Korean Speech Intelligibility Index (KSII). The BIF for Korean sentences will provide information regarding the relative significance of different frequency bands to speech intelligibility for the Korean language. Thus, the accuracy of the SII for Korean may be improved by using the Korean BIF. Furthermore, more accurate prediction for speech intelligibility will be feasible in hearing-aid fittings for Koreans, if the SII based on Korean BIF is used.

II. METHODS

A. Participants

Seventy-eight normal-hearing listeners participated in this study. There were forty-eight male and thirty female participants with a mean age of 29.3 years (age ranging from 20 to 40 years). All participants were native Korean speakers. Their auditory thresholds were 20 dB Hearing Level (HL) or better at octave frequencies from 250 Hz to 8000 Hz (ANSI, 2010). To be included as a native Korean speaker, subjects were required to meet six criteria which were based on language experience and a proficiency questionnaire (LEAP-Q) (Marian et al., 2007). First, the participants' parents should use Korean as their first language. Second, the participants should be born in Korea. Third, the participants should have learned Korean as their first language during childhood. Fourth, the participants should be graded on their ability as a native-

Korean speaker with a minimum of 6 on a scale of 1 to 7 (where 7 represents native-like ability and 1 represents no ability at all). Fifth, the participants should use Korean as their current and primarily spoken language. Finally, the participants should mainly speak Korean at home. For the intelligibility task, listeners were tested monaurally (right ear) and individually in a double-walled sound-attenuated booth. Participants were compensated $10 per hour for their participation. The study was approved by the institutional review board of the University of Colorado.

B. Stimuli

Korean Hearing-In-Noise Test (HINT) sentences were used for this study (Moon et al., 2008). The sentences were spoken by a male professional actor and were recorded at a sampling rate of 44.1 kHz. The actor was instructed to maintain a normal voice effort, pace, and clarity. Frequencies of occurrence of phonemes for the sentences from the present study were similar to frequencies of occurrence of Korean phonemes from Korean drama and news scripts (Byun, 2001) (see Appendix [A1]). A total of 250 sentences were used, with 25 sets of 10 sentences per set (1 practice set; 24 test sets). Each test set comprised 33 keywords, and all test sets had similar long-term average speech spectra. Sentences were mixed with speech-shaped noise which was matched to the spectrum of the sentences to create signals that varied in SNR. Five SNR conditions were used (-8, -4, 0, 4, and 8 dB SNR). The noisy sentences were adjusted to have a level of 65 dB SPL. The noisy speech sentences were filtered through 21 low-pass (LP) filters (200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400, 7700, and 9500 Hz) and 21 high-pass (HP) filters (100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400, and 7700 Hz). These 42 cutoff frequencies were chosen to be consistent with the SII critical-band calculation procedure (ANSI, 1997/R2012). Linear-phase finite impulse response (FIR) filters were used having a rejection slope of 96 dB/octave at the desired cutoff frequencies.

C. Equipment

The stimuli were routed to an audiometer (GSI61, Grason-Stadler, Eden Prairie, MN) and presented to the participant's right ear through a headphone (TDH-50P, Telephonics Corporation, Farmingdale, NY). The output level of the headphone was calibrated using a 1 kHz pure tone which has the same average root-mean-square (RMS) level of the speech stimuli. To estimate the speech level, all of the sentences were concatenated without any separating pauses and then were digitized at a 44.1 kHz sampling rate in a low-pass filtered condition at 9500 Hz.

D. Test procedure

The experimental protocol was similar to the procedures used by Studebaker and Sherbecoe (1991), Eisenberg et al. (1998), and Wong et al. (2007). Participants listened to the sentences over headphones. The task was to listen to a sentence and then repeat back as much of the sentence as possible. Intelligibility (percent keywords correct) was then measured for sentences processed in multiple filtering conditions (21 LP and 21 HP filters) and in several SNR conditions (5 SNRs). The total number of conditions was 210 (42 filters × 5 SNRs). However, 208 conditions were used because of the research design (balanced incomplete block design, BIBD) of this study. Thus, two conditions (LP 200 Hz with -8 dB SNR and HP 7500 Hz with -8 dB SNR), which were expected to have a 0% score, were excluded. A detailed description of the BIBD is provided in the research design section below. Because the stimulus set comprised 24 (240 sentences), each listener randomly participated in a subset (24 conditions) of the total number of conditions (208 conditions). Thus, each participant did not hear the same sentence more than once. At the outset of the test block, each listener was presented with a practice block that included 10 test trials in randomized order. Feedback was not provided for the practice or the test trials.

E. Research design and statistical analysis

There are two primary research designs used in the present analysis: between-groups and within-subjects designs. The premise of a between-groups design is that each group is assigned to a different condition in the experiment. Thus, each group can only participate in one condition among the total number of conditions (210 filtering/SNR conditions). However, the between-groups design requires too many participants for the present study. For example, if 10 participants are considered for each group, 2100 native Koreans will be required for 210 conditions (Table I). The premise of a within-subjects design is that the same participants participate in all of the experimental conditions. Thus, each participant should participate in all of the filtering/SNR conditions (210 conditions). However, the within-subjects design requires too many sentences for the present study. For example, if 10 sentences are considered for each condition, 2100 sentences will be required for 210 conditions (Table I). There is no sentence material that includes 2100 Korean sentences and the HINT sentences (24 sets) were the only option for the present study.

			Between-groups design			Total
Filtering/SNR Condition (C)	C1	C2	…	C209	C210	210
Subjects	10	10	…	10	10	2100
			Within-subjects design			Total
Filtering/SNR Condition (C)	C1	C2	…	C209	C210	210
Sentences	10	10	…	10	10	2100

TABLE I. Examples of between-group and within-subjects designs for 210 filtering/SNR conditions.

Due to the limitations of the number of participants and the number of sentences relative to the number of filtering/SNR conditions (210 conditions), we adopted a mixed design that included between-groups and within-subjects designs. The mixed design satisfied two important constraints. First, each participant did not hear the same sentence more than once to control the possible learning effects. Second, the experimental goal was to measure nine listeners' intelligibility performance per condition. Therefore, each listener participated in a subset (24) of the total number of conditions, and total of 78 listeners were recruited.

To increase the statistical power, we controlled several factors in the mixed design. First, filtering/SNR conditions were randomly assigned to each participant. Second, to decrease variance in experimental error and to provide more precise comparisons among filtering/SNR conditions than is possible with a complete block design, a balanced incomplete block design (BIBD) was used in this study.

While complete block designs mean that all treatments (testing conditions) are represented in each block of the experiment, a BIBD is used when the number of testing conditions (e.g., 210 SNR/filtering conditions) exceeds the number of units per block (e.g., 24 stimuli sets per participant) (Hinkelmann and Campthorne 2005). The BIBD was introduced by Yates (1936) and has been used in many fields of study such as by Jung (1961), Fleiss (1981), and Yang et al. (2011). The mixed-block design has to satisfy three requirements: 1) The product of the number of subjects (78) and the number of stimuli (24) should be equal to the product of the number of repetitions (9) and the number of filtering/SNR conditions (208) (78 × 24 = 9 × 208 = 1872). 2) When the value for filtering/SNR conditions (208) minus 1 is divided by the value, which is the multiplication of the number of repetition (9) and the number of stimuli (24) minus 1 (24-1), the calculated value should be equal to an integer ((208 – 1) / 9 × (24 - 1) = 1). The meaning of 1 is related to the total number of experimental units that appear with a specific condition in this study. To satisfy these relations, two filtering/SNR conditions (LP 200 Hz at -8 dB SNR and HP 7500 Hz at -8 dB SNR), which were expected to have a 0% score, were excluded and 208 conditions were selected for the mixed-block design of the present study. 3) The BIBD is based on the assumption that block effects are not significant. To test this assumption, a univariate general linear model (GLM) was used to identify whether block (within-subject) effects were significant (PASW Statistics 18, IBM Corporation, Armonk, NY).

F. Calculation approach for deriving the band-importance function

The present study used two different procedures for deriving the BIF: The first procedure was based on an established protocol (S&S procedure) (e.g., French and Steinberg, 1947; Studebaker and Sherbecoe, 1991; Wong et al., 2007), and the second procedure was based on a recently published method (Kates procedure) (Kates, 2013).

The S&S procedure contains several steps. The procedure was proposed by French and Steinberg (1947) and has been used previously by several investigators (Studebaker and

Sherbecoe, 1991; Studebaker et al., 1993; DePaolis et al., 1996; Wong et al., 2007). The first step was to determine the relationship between the SII values and the percent correct scores using smoothed curve fitting. To generate the smoothed curves, the original percent correct scores were adjusted by four rules (e.g., Studebaker and Sherbecoe, 1991): (1) A curve at one SNR did not intersect a curve with another SNR; (2) Both LP and HP curves for any one SNR had to end at the same scores; (3) The intelligibility scores had to increase as the SNR is increased; (4) LP and HP curves at each SNR had only one crossover frequency. Then, the final smoothed lines were generated using a cubic spline-fitting procedure, which was similar to the procedure used by DePaolis et al. (1996). We adopted the cubic spline-fitting procedure because it minimized the variance between unsmoothed and smoothed data. The starting point for the relationship was determined by plotting the percent correct scores as a function of band frequency at the highest SNR (+8 dB SNR) used in the study: One curve represents the data when the low-frequencies were removed using an HP filter, while the other curve represents the data when the high-frequencies were removed using an LP filter. The intersection of these two curves determined the first point in a series of graphical processes and curve interpolations for data at different SNRs, yielding the fitting constant values for Q and N in Equation 2. The proficiency value P was assumed to be 1.

Next, the BIF was derived using the computed values of Q, N, and P through the inverse of Equation 2. In the inverse of Equation 2, the percent correct values for all filtering/SNR conditions were transformed into their associated SII values (cumulative values). Averaged cumulative SII values were then converted to separate SII values for each band. For the LP-filtered values, separate SII values were calculated by subtracting the SII value for the lower cutoff frequency from the SII value for the higher cutoff frequency. Separate SII values for the

HP-filtered conditions were calculated by the reverse procedure. The separate SII values for LP and HP filter conditions were then combined. If the value was zero, the value was not used in computing combined values. If two values at the same LP and HP cutoff frequency with the same SNR were greater than zero, an averaged value was used. Values for the same cutoff frequency band at the different SNR conditions were then averaged. The averaged values were then expanded to a 0 to 1 scale by dividing every SII value by the largest SII value which corresponded to the SII value for the highest frequency band. The final importance weight for each frequency band was then obtained by subtracting the SII value for the low-frequency edge of the band from the high-frequency edge.

The BIF was also derived by the procedure described in Kates (2013). This procedure does not depend on arbitrary data smoothing and graphical constructions. Instead, it minimizes the RMS error between the proportion correct observed in the intelligibility performance and the proportion correct predicted by the SII, in conjunction with the transformation of Equation 2. The new procedure used the functional minimization routines provided in MATLAB (version R2013a; MathWorks Inc., Natick, MA, USA). Starting with initial values of the frequency importance function, Q, N, and dynamic-range offset K, the procedure iteratively adjusted these values to produce the best match of the model output to the subject intelligibility scores. The RMS error was computed by taking the RMS difference of the predicted and measured proportions correct. More details were described in Kates (2013).

III. RESULTS

A. Statistical analysis

When the data for intelligibility scores were averaged, an arcsine transform was used to stabilize the variance (Studebaker, 1985). First, each listener's intelligibility score was transformed to an arcsine unit. Then, arcsine values from different listeners were averaged for each condition and the averaged arcsine values were used for the analysis of the data. A univariate general linear model (GLM) was used to determine whether block (within-subject) effects were significant. The GLM analysis showed that there was no significance variation of intelligibility scores among listeners in each condition ($F=1.154, p > 0.05$). This result supports the conclusion that block effects are not significant which satisfied the constraint for the BIBD.

B. Mean intelligibility scores

To derive the BIF, the averaged arcsine values were converted back to percent values. Mean intelligibility scores (percent keywords correct) in 42 filtering (21 LP and 21 HP filters) at 5 SNR conditions are reported in Table II. Overall, average intelligibility scores increased as audible frequency band and SNR increased. In the LP-filtered conditions, mean intelligibility scores increased as cutoff frequencies increased from 200 Hz to 9500 Hz. For example, at the highest SNR (+8 dB) with LP filtering, the mean intelligibility scores increased from 0.3% to 100% as cutoff frequencies increased from 200 Hz to 9500 Hz. In contrast, the mean intelligibility scores for HP filtering conditions increased as cutoff frequencies decreased from 7700 Hz to 100 Hz. For instance, at the highest SNR (+8 dB) with HP filtering, the mean intelligibility scores increased from 0% to 99% as cutoff frequencies decreased from 7700 Hz to 100 Hz.

Cutoff frequency (Hz)	SNR (dB) for low-pass filter (LP)					SNR (dB) for high-pass filter (HP)				
	-8	-4	0	4	8	-8	-4	0	4	8
100						20.5	52.5	88.2	98	99
200		0	0	0.3	0.3	14.5	52.9	84.5	98	99
300	0	0	0	0	0.7	11.4	45.8	82.5	95	99
400	0	0	0.3	1.3	3.7	6.7	30	69.4	91.6	97.6
510	0	0.7	6.4	11.1	25.3	4.4	17.8	56.9	88.2	92
630	0.7	4	13.8	32.7	48.8	3.4	17.2	38.7	84.2	89.2
770	0.3	7.1	18.2	41.8	53.1	2.4	12.8	35.4	70.7	90.2
920	1.7	5.7	18.9	53.5	54.9	2	8.4	32.3	68	85.2
1080	1.3	8.8	19.2	52.8	53.9	2.4	6.7	28.6	65.7	86.2
1270	1	7.4	30.3	51.9	66.3	1	6.1	17.5	48.8	64
1480	2.4	11.1	40.7	72.4	77.4	0	1.7	12.5	20.9	55.9
1720	4.4	11.4	45.1	76.4	90.6	1.3	0.7	9.8	17.5	43.8
2000	4.4	18.5	58.6	77.8	89.6	0	0.7	7.7	17.8	40.4
2320	3.7	21.9	63	88.6	93.6	0	1	7.7	12.8	22.9
2700	3	30.3	72.1	87.9	93	0	0.3	3.7	7.7	14.1
3150	9.1	43.8	72.4	91.2	94	0	0.7	4.4	4	6.7
3700	7.4	39.7	81.5	95	97	0	0	0	1	3
4400	11.8	39.4	78.8	94.3	99	0	0	0	0	0
5300	10.8	49.2	86.9	94.2	97.6	0	0	0	0	0
6400	18.2	62.6	90.2	98	99	0	0	0	0	0
7700	20.9	65.3	91.6	96.6	100		0	0	0	0
9500	19.2	66	92.2	97	100					

TABLE II. Mean percent correct scores for 21 high- and low-pass filtered conditions.

C. Crossover frequencies

The crossover frequency refers to the frequency which divides the frequency range into two parts, each of which contributes 50% to intelligibility. Each crossover frequency was obtained from the mean intelligibility scores of LP and HP at the same SNR. Crossover frequencies were 1208 Hz at +8 dB SNR, 1163 Hz at +4 dB SNR, 1070 Hz at 0 dB SNR, 913 Hz at -4 dB SNR, and 1200 Hz at -8 dB SNR.

D. Derivation of the relative transfer function

The BIF was derived through an iteration technique (French and Steinberg, 1947; Studebaker et al., 1987; Studebaker and Sherbecoe, 1991). Briefly, numerical constant values (Q and N) were determined from the relationship between relative SII values and mean percent scores through the Equation 2. Then, the BIF was derived using the constant values and the mean scores (proportional values) through the inverse of Equation 2. The first step was to determine a relative transfer function (RTF) that represents the relationship between mean percent scores (smoothed) and relative SII values. When deriving the RTF, the wide band condition with the highest score was assumed to have an SII value of 1.0 and other conditions were adjusted to give SII values relative to that value. Then, the mean percent scores for each condition were converted to associated SII values through the inverse of Equation 2. The difference of SII values at two cutoff frequencies was determined as the band-importance weight of that frequency band.

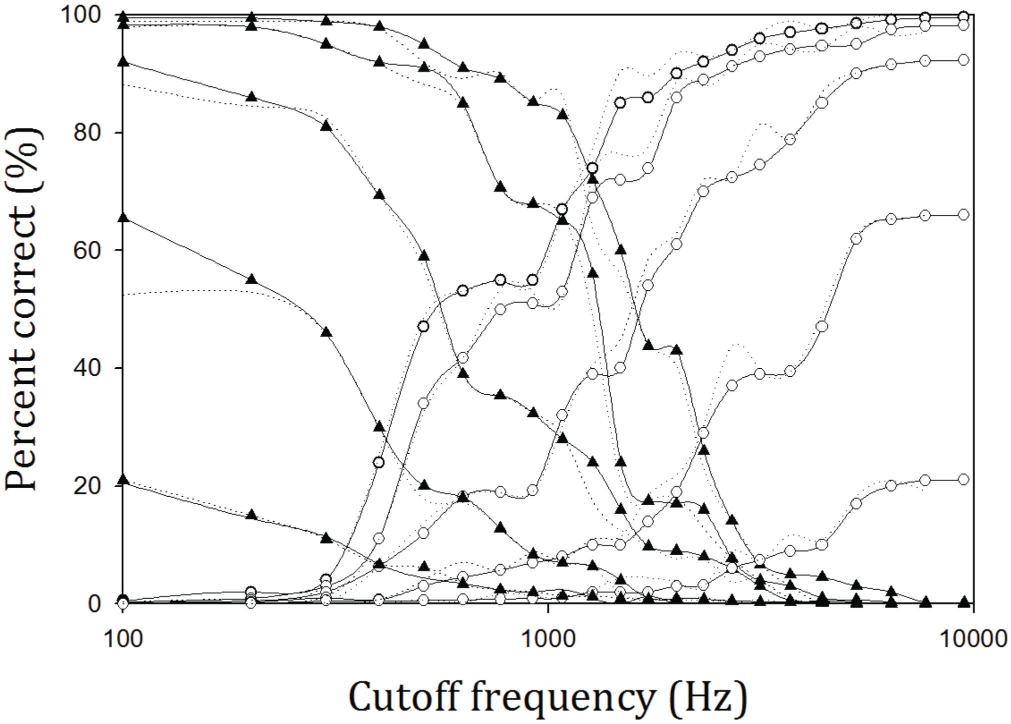


FIG. 7. Unsmoothed and smoothed mean percent correct scores. The solid lines indicate the smoothed values. Triangles indicate smoothed values for HP-filtered conditions and circles indicate smoothed values for LP-filtered conditions. This figure displays +8, +4, 0, -4, and -8 dB SNRs for both HP and LP conditions. The dotted lines indicate unsmoothed data.

To derive the RTF, the intelligibility scores for filtering/SNR conditions were fitted smoothly using the cubic-spline interpolation. Unsmoothed and smoothed data are shown in Figure 7. The smoothed data was based on the four rules that were used by Studebaker and Sherbecoe (1991) before applying the cubic-spline interpolation. Then, the curve-bisection method was used as described in the methods section to derive the RTF. As shown in Figure 8, a total of 10 SII values represented by the circles in the figure were determined and the best-fit RTF curve was derived using the MATLAB curve fitting toolbox 2.2 (Natick, MA) by Equation 2. The value *P* was assumed to be a value of 1.0 and fitting constants were determined by the 10

SII values. The values for *Q* and *N* values were 0.2635 and 15.52, respectively. The value of R-squared was 0.9856 indicating a reliable relationship between the predicted and obtained values. Overall, the intelligibility score increased as the SII value increased. The value of 0.28 SII corresponded to 20% of the intelligibility score, the value of 0.52 SII corresponded to 80% of the intelligibility score, and the value of 0.5 SII corresponded to 78.7% of the intelligibility score.

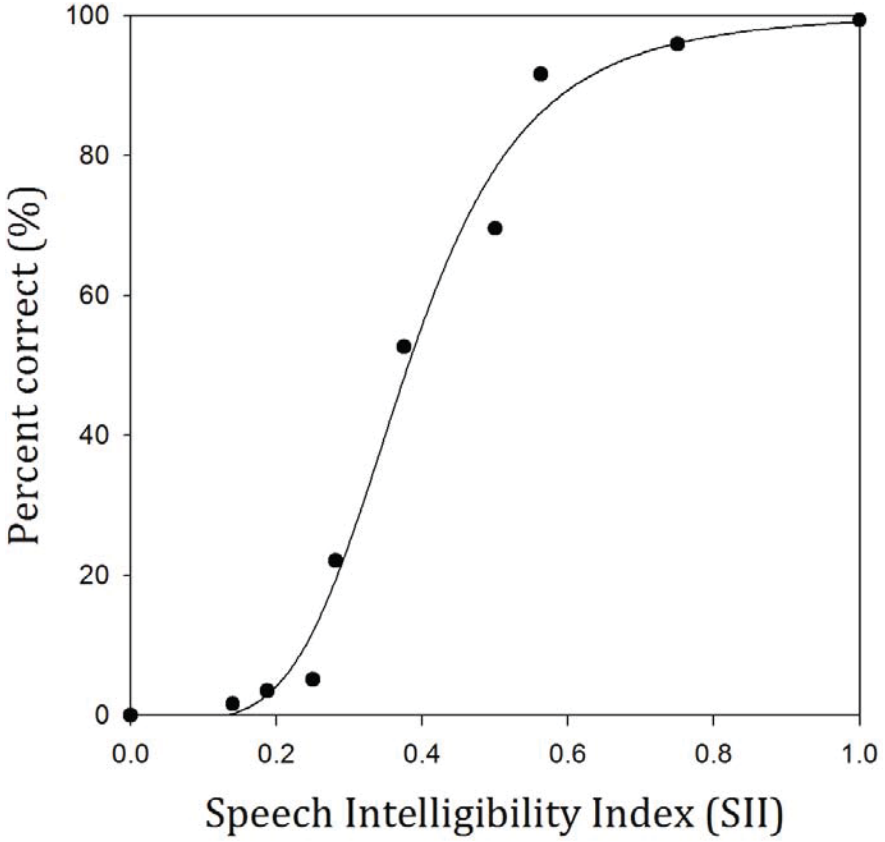


FIG. 8. Best-fit RTF. Circles represent actual obtained values. The RTF was derived by Equation 2 and the values for Q and N were 0.2635 and 15.52.

E. Derivation of the band-importance function

Two BIFs were derived by the S&S procedure and the Kates procedure. Table III and Figure 9 report band-importance weights for each frequency band. In the case of the BIF produced by the S&S procedure, the center frequency (CF) of 350 Hz accounted for the highest importance weight (8.7%) and frequencies below 630 Hz accounted for 37.6% of importance weight. The BIF produced by the Kates procedure had weights similar compared to the BIF produced by the S&S procedure. The CF of 350 Hz and 450 Hz accounted for highest weights (8.3%) and frequencies below 630 Hz accounted for 36.1% of importance weight. The highest gap between two procedures was 1% at the CF of 250 Hz and the gap was less than 1% in all other frequency bands.

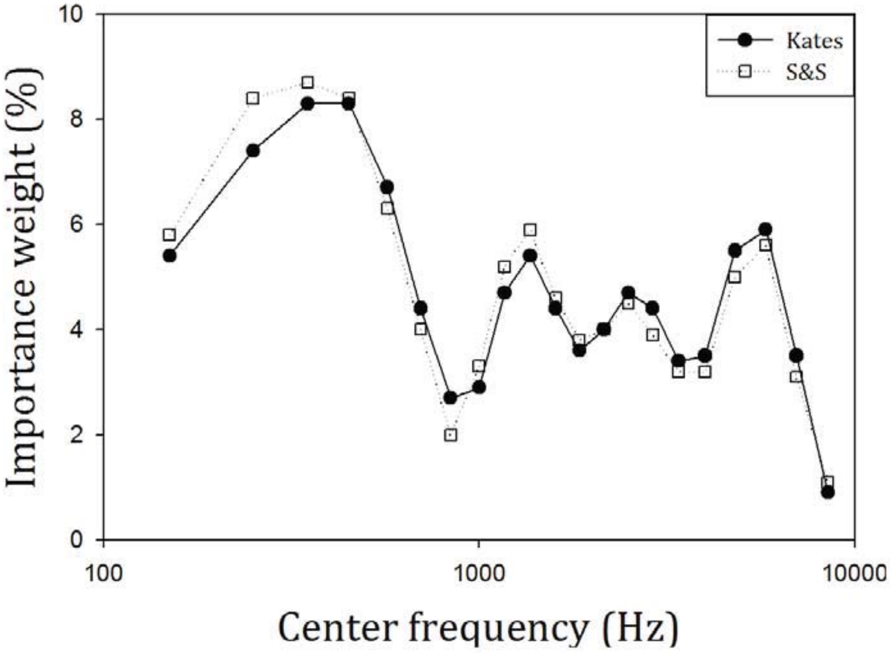


FIG. 9. BIFs for 21 frequency bands derived by the S&S procedure and the Kates procedure. Circles indicate the BIF produced by the Kates procedure and squares indicate the BIF produced by the S&S procedure.

Band limit (Hz)	Center frequency (Hz)	BIF by S&S (%)	BIF by Kates (%)	Band limit (Hz)	Center frequency (Hz)	BIF by S&S (%)	BIF by Kates (%)
100-200	150	5.8	5.4	1720-2000	1850	3.8	3.6
200-300	250	8.4	7.4	2000-2320	2150	4.0	4.0
300-400	350	8.7	8.3	2320-2700	2500	4.5	4.7
400-510	450	8.4	8.3	2700-3150	2900	3.9	4.4
510-630	570	6.3	6.7	3150-3700	3400	3.2	3.4
630-770	700	4.0	4.4	3700-4400	4000	3.2	3.5
770-920	840	2.0	2.7	4400-5300	4800	5.0	5.5
920-1080	1000	3.3	2.9	5300-6400	5800	5.6	5.9
1080-1270	1170	5.2	4.7	6400-7700	7000	3.1	3.5
1270-1480	1370	5.9	5.4	7700-9500	8500	1.1	0.9
1480-1720	1600	4.6	4.4				

TABLE III. BIFs for 21 frequency bands derived by the S&S procedure and the Kates procedure.

IV. DISCUSSION

After determining the BIF, the RTF can be transformed to the absolute transfer function (ATF) that represents the true relationship between SII values and intelligibility scores (Studebaker and Sherbecoe, 1991). For the ATF by the BIF from the S&S procedure, two steps are followed (Studebaker and Sherbecoe, 1991). First, SII values for each condition are calculated using the product of the BIF and the BAF for each frequency band. The BAF can be expressed as an algebraic equation:

$$A_i = (SNR_i + K) / D, \tag{3}$$

In Equation 3, SNR_i is the SNR in frequency band i expressed in dB, K is an offset related to the ratio of the speech peaks to the speech RMS level (French and Steinberg, 1947), and D is the dynamic range of the stimuli. The value of D is assumed to be a value of 30 dB (ANSI-S3.5, 1997/R2012). Next, unsmoothed proportion scores for all conditions between 5% and 95% are plotted as a function of SII values from the first step. Then, Q, N, and K values are determined using Equation 1. The values for Q and N are 0.2178 and 15.52, and the value of K that yields the least error is 14.15 dB. The value of R-squared is 0.9774. For the ATF by the BIF from the Kates procedure, the same procedure is followed. *Q*, *N*, and *K* values are 0.219, 15.58, and 14.33, respectively. The ATFs by BIFs from two different procedures (the S&S procedure and the Kates procedure) are shown in Figure 10. Both ATFs have almost the same values. Approximately, the SII value of 0.22 corresponds to the 20% of percent correct intelligibility score, the SII value of 0.3 corresponds to the 50%, and the SII value of 0.41 corresponds to the 80.1% in both ATFs.

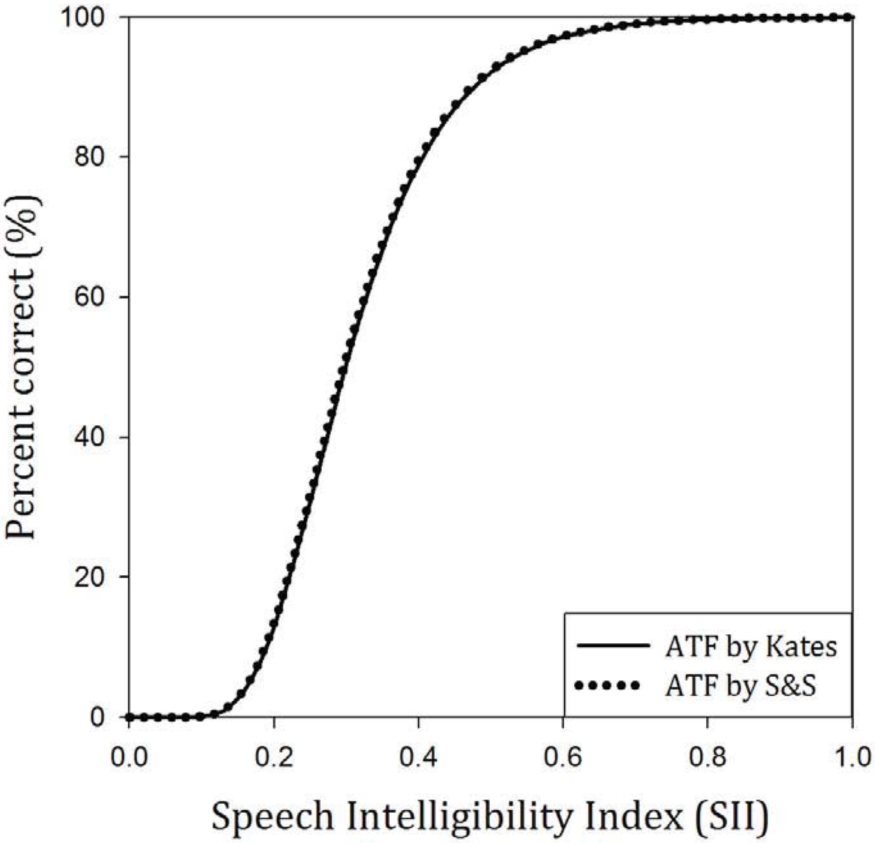


FIG. 10. The ATF by band-importance weights from the S&S procedure and the Kates procedure. The DR is assumed to be a 30 dB.

In the SII for English, the DR of speech is assumed to be 30 dB across all frequency bands (Dunn and White, 1940; Pavlovic, 1988; ANSI S3.5-1997/2012), which is based on a "best-fit" of the SII model to the speech intelligibility data. The ATF for Korean in the present study is also derived based on the DR of 30 dB. However, the DR can be dependent on the language (Jin et al., 2014). Jin et al. (2014) reported that the Korean speech DR is different from English and Mandarin speech DRs, and this suggests that more accurate ATF for Korean may be derived by choosing a DR other than the 30 dB. Thus, further studies are required to derive the "best-fit" ATF for Korean.

The band importance weights for the S&S procedure and the Kates procedure are similar (Table III and Fig. 9). The proportion correct scores computed from band-importance weights by the S&S procedure and the proportion correct computed from band-importance weights by the Kates procedure are compared in Figure 11. The RMS errors for the S&S procedure and for the Kates procedure are 0.038 and 0.042, respectively. The Pearson correlation coefficients are 0.994 for the S&S procedure and 0.993 for the Kates procedure ($p < 0.001$). These results indicate that both procedures accurately predict intelligibility scores.

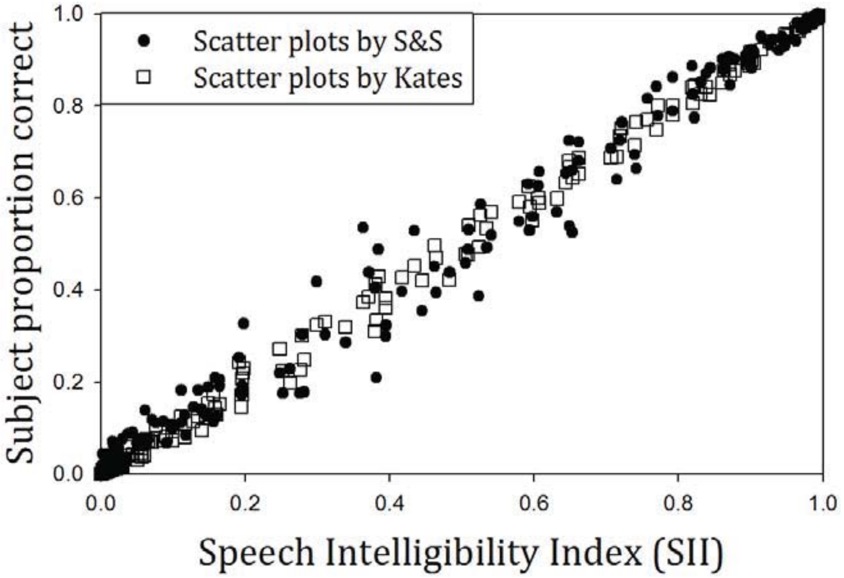


FIG. 11. Scatter plots showing the proportion correct scores as a function of the predicted SII values. Circles are generated by the S&S procedure and squares are generated by the Kates procedure.

In the present study, the data smoothing is based on the four rules by Studebaker and Sherbecoe (1991) and the cubic spline-fitting curve that used by DePaolis et al., 1996. There are several reasons to choose these methods for the data smoothing. First, intelligibility in this study is measured in 42 filtering bands, and there are several conditions where the intelligibility remains constant or even decreases as the bandwidth is increased. However, the calculation for the BIF from the S&S procedure does not allow any decrement of the score as the bandwidth increases because the S&S procedure is based on the assumption that the intelligibility scores increase monotonically as the bandwidth or SNR increases. Thus, data adjustments by the four rules are indispensable. Second, we select a curve fitting procedure which yields the least variation between unsmoothed and smoothed data among various curve fitting methods.

Although curve fittings such as logistic or sigmoid curves yield smoother curves compared to the cubic spline fit, variations between unsmoothed and smoothed data by other curve fitting techniques are bigger than the variation in the cubic spline fit. Because the cubic spline-fitting procedure provides precise calculation values for polynomial coefficients so it could minimize the assumption for values between scores for each cutoff frequency, the cubic spline-fitting is adopted.

The comparison of cumulative BIFs for Korean sentences by the S&S procedure (the present study), Cantonese sentences (Wong et al., 2007), English sentences (Healy et al., 2013), English average speech (Pavlovic, 1987), and English short passages (Cox and McDaniel, 1984) is shown in Figure 12. Compared to BIFs from other languages (Cantonese and English), low-frequency areas less than 840 Hz (CF) are more important for the Korean BIF. For example, the cumulative band-importance weight (the CF of 840 Hz) for Korean sentences is 43.6% but other languages' band-importance weights are 43% for English short passages (Cox and McDaniel, 1984), 35.8% for Cantonese sentences (Wong et al., 2007), 32% for English sentences (Healy et al., 2013), and 30.6% for English average speech (Pavlovic, 1987). The biggest difference of the cumulative band-importance weights between Korean sentences and other languages is observed in the CF of 350 Hz. The cumulative band-importance weight for Korean sentences is 22.9% but other languages' band-importance weights are 14.3% for English short passages (Cox and McDaniel, 1984), 11.8% for Cantonese sentences (Wong et al., 2007), 11.4% for English sentences (Healy et al., 2013), and 7.8% for English average speech (Pavlovic, 1987). Based on the comparison of cumulative BIFs for Korean, Cantonese, and English, we may conclude that low-frequency areas (less than the CF of 840 Hz) are more important in the Korean BIF than for other languages.

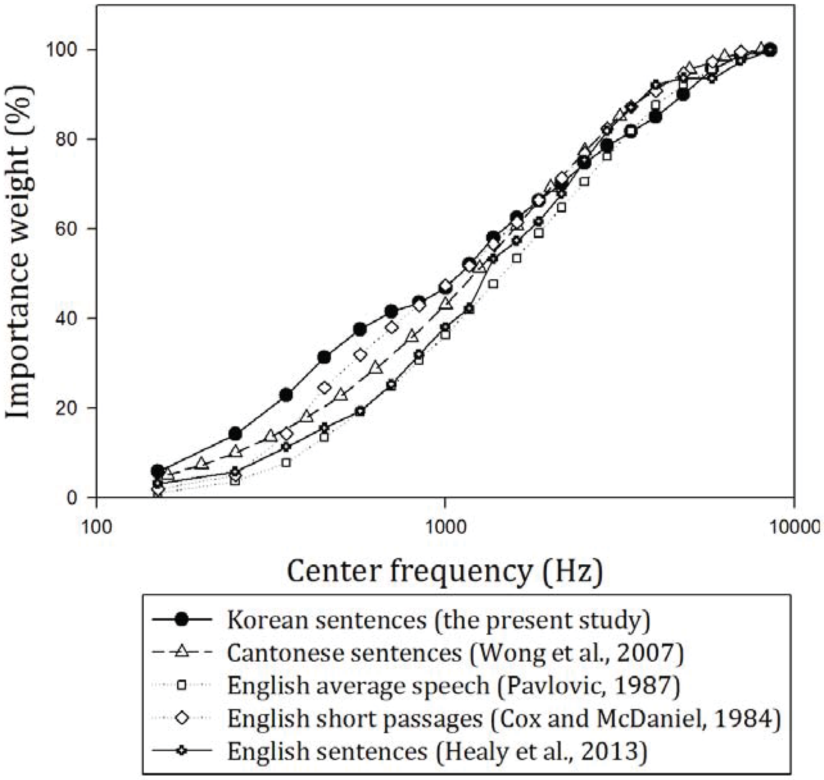


FIG. 12. Comparison of cumulative BIFs for Korean sentences and other languages (sentences, average speech, and short passages).

The BIF for Korean sentences has implications for using the SII for Korean in clinical contexts. First, the SII for Korean may provide a baseline to develop HA fitting prescriptions for Korean. For example, NAL-NL1 (National Acoustics Labs, Non-Linear, version 1) is a prescriptive procedure that uses the English SII model to derive optimal frequency-gain characteristics (Dillon, 1999). If a HA fitting prescription is based on the SII for Korean, the prescription will provide appropriate gains for Korean. Second, the SII for Korean may be applied to the prediction of HA outcomes for Korean speakers, and employed in the prediction of aided gain. (Mueller and Killion, 1990; Rankovic, 1991; Amlani et al., 2002; Hornsby, 2004; Killion and Mueller, 2010). Thus, more precise prediction for the improvement of speech intelligibility due to amplification will be feasible in HA fittings for Korean speakers using the SII for Korean.

In the present study, BIFs for Korean sentences were based on the intelligibility scores measured in 42 filtering (21 LP and 21 HP filters) at 5 SNR conditions and were derived using two different procedures (the S&S procedure and the Kates procedure). The BIF weights produced by the two different procedures were similar, and both procedures accurately predicted intelligibility scores. Thus, the Kates procedure could be used to cross-check the BIF by comparing with the S&S procedure. In addition, it may be regarded as another method to derive the BIF. Because the BIF for Korean sentences were different than the BIFs from other languages (Cantonese and English), the SII based on the BIF for Korean sentences would be expected to provide more accurate predictions of intelligibility of Korean sentences. In addition, if the best-fit BAF for Korean is integrated with the BIF for Korean, it will yield the best-fit SII for Korean.

CHAPTER IV

CONCLUSIONS

The main purpose of this dissertation was to provide a baseline for developing the SII for Korean. Thus, two studies were conducted. In the first study, the DRs for three languages (Korean, English, and Mandarin) were quantified using different definitions of DR (99%-20%, 99%-10%, and 99%-1%), for several integration times (from 1 ms to 512 ms), and in different frequency bands (from a CF of 150 Hz to a CF of 8600 Hz). The DR was affected in similar ways with changes in DR definition and integration time across the three languages. However, across-language differences in DR were evident when considering frequency-band effects. Specifically, the DR for Korean was smaller than the DRs for Mandarin and English in low-frequency bands (less than a CF of 455 Hz). The DR for English was smaller than the DR for Mandarin and Korean in mid-frequency bands (between a CF of 455 Hz and a CF of 4050 Hz). The DRs Korean and Mandarin were smaller than the DR for English in high-frequency bands (above a CF of 4050 Hz).

In the second study, BIFs for Korean sentences were derived. The BIFs for Korean sentences were based on the intelligibility scores measured for a large number of filtering and 5 SNR conditions and were derived using two different procedures (the S&S procedure and the Kates procedure). Both procedures yielded similar BIF weights and accurately predicted intelligibility scores. The BIF for Korean sentences showed that the frequency bands contributing most to speech intelligibility were near a CF of 350 Hz. The importance weights of the frequency bands below a CF of 630 Hz was 37.6% for the S&S procedure and 36.1% for the Kates procedure. Compared to published English and Cantonese BIFs, the present results showed that low-frequency regions (less than a CF of 840 Hz) were more important in the BIF for Korean sentences.

Results of this dissertation provide a baseline for developing the SII for the Korean language. First, since the DR for Korean speech is different from other languages, the DR needs to be measured for the Korean language when deriving the BAF. Second, the interactions of DR definition and integration time should be considered. The most accurate implementation of the SII for the Korean language may require adjusting both the integration time and the DR definition, and these may differ from the values used for the ANSI standard (1997/R2012)). Third, the SII for the Korean languages should consider the frequency-dependent nature of the DR. Fourth, the development of the SII for Korean should consider a DR that has been optimized by a best-fit factor. When deriving the ATF for Korean in the second study, the BIF for Korean sentences was considered but the BAF was based on the DR of 30 dB that is a "best-fit" of the English SII model to the speech intelligibility data. The reason the 30 dB was selected for the DR in the BAF is that the ATF accurately predicted intelligibility scores ($R^{\wedge 2} = 0.9774$). However, more accurate BAF for Korean might be derived by choosing a DR other than the 30 dB because DRs depend on the language. Thus, further studies are required to derive the "best-fit" BAF for Korean and to determine the sensitivity of the prediction accuracy to the DR used in the equations. Fifth, because the BIF for Korean sentences is different than the BIFs from other languages (Cantonese and English), the SII for Korean sentences should base on the BIF for Korean sentences. The SII based on the BIF for Korean sentences will provide more accurate predictions of intelligibility of Korean sentences. Lastly, if the best-fit BAF for Korean is integrated with the BIF for Korean, it will yield the best-fit SII for Korean.

The findings of this dissertation will be applied for the SII for Korean to predict more accurate intelligibility performance. In addition, the SII for Korean will be used as a basis to develop HA fitting prescriptions for Korean. For example, the DR based on the Korean language may yield better predictions of intelligibility for Korean in the SII compared to the values used for the ANSI standard (1997/R2012). When the DR of speech was calculated for the 99%-10% condition at 32 ms in the CF of 250 Hz, DRs for English and Korean were 46.25 dB and 33.5 dB, respectively. In the case of the DR for English, each decibel represents 1/46.25 (2.16 %) of the audible signal range that contributes to speech intelligibility. However, in the case of Korean, each decibel represents 1/33.5 (2.98%) of the audible signal. Consider, for example, a linear amplification system in which the average signal level is amplified to be 15 dB above the impaired auditory threshold at 250 Hz. For an English speaker, approximately 82 percent of the speech would fall above threshold, while for a Korean speaker approximately 95 percent would fall above threshold because of the reduced DR in Korean. Thus, the audible signal range for the Korean speaker will be greater than for the English speaker, and the SII computed using the DR measured for the language will be higher. The desired hearing-aid gain required to achieve a specified SII value depends on both the average signal level and its DR. A lower average level may require increased gain to ensure audibility, but a reduced DR requires less amplification to place the speech minima above the impaired auditory threshold.

The BIF for Korean sentences also has implications for using the SII for Korean in clinical contexts. For instance, a band-importance weight for bandwidth between 100 Hz to 400 Hz in Korean sentences (S&S procedure) was 22.9% and a band-importance weight for the same bandwidth in English sentences was 4.63% (DePaolis et al., 1996). If a Korean speaker and an English speaker has same degree of severe hearing loss in the low-frequency areas (bandwidth between 100 Hz to 400 Hz), the Korean speaker shows lower intelligibility performances compared to the English speaker due to the difference of the band-importance weight between two languages. If the entire DR for the low-frequency area can be restored by a HA gain, the improvement of the SII will be approximately 0.23 for the Korean speaker and 0.05 for the English speaker. Because BIFs are different between two languages, a HA outcome for the Korean speaker is different compared to a HA outcome for the English speaker. This example indicates that more precise prediction for the improvement of speech intelligibility due to amplification will be feasible in HA fittings for Korean speakers when considering the SII based on the Korean BIF.

In addition, the SII for Korean may provide a baseline to develop HA fitting prescriptions for Korean. For example, NAL-NL1 (National Acoustics Labs, Non-Linear, version 1) is a prescriptive procedure that uses the English SII model to derive optimal frequency-gain characteristics (Dillon, 1999). If a HA fitting prescription is based on the SII for Korean, the prescription will provide more accurate gains for Korean compared to prescriptions based on other languages' SII.

REFERENCES

American National Standards Institute. (1997(R2012)). *Methods for calculation of the Speech Intelligibility Index, ANSI S3.5-1997.* New York: Acoustical Society of America.

American National Standards Institute. (2010). *Specifications for Audiometers, ANSI S3.6-2010.* New York: ANSI.

Amlani, A. M., Punch, J. L., & Ching, T. Y. C. (2002). Methods and applications of the audibility index in hearing aid selection and fitting. *Trends in Amplification*, 6, 81-129.

Beranek, L. L. (1947). The design of speech communication systems. *Proceedings of the IRE*, *35*(9), 880-890.

Beutelmann, R., & Brand, T. (2006). Prediction of speech intelligibility in spatial noise and reverberation for normal hearing and hearing impaired listeners. *Journal of the Acoustical Society of America*, 120(1), 331-342.

Black, J. W. (1959). Equally contributing frequency bands in intelligibility testing. *Journal of the Speech and Hearing Research*, 2, 81-83.

Boothroyd, A. (2008). The performance/intensity function: An underused resource. *Ear and Hearing*, 29, 479-491.

Boothroyd, A., Erickson, F. N., & Medwetsky, L. (1994). The hearing aid input: A phonemic approach to assessing the spectral distribution of speech. *Ear and Hearing*, 15(6), 432-442.

Browning, L. K. (1974). The Cantonese dialect with special reference to contrasts with Mandarin as an approach to determining dialect relatedness, Ph.D. dissertation, Georgetown University.

Byrne, D., Dillon, H., Tran, K., et al. (1994). An international comparison of long-term average speech spectra. *Journal of the Acoustical Society of America*. 96, 2108–2120.

Byun, S. W. (2001). Frequencies of Korean phonemes and reliability of Korean phonetically balanced word lists. *Korean Journal of Otolaryngology-Head and Neck Surgery*, 44(5), 485-489.

Chasin, M. (2011). Setting hearing aids differently for different languages. *Seminars in Hearing*, 32(2), 182-188.

Cheung, P. P. (1992). *Tonal confusions in Cantonese at different signal-to-noise ratios*. B. Sc. Dissertation, University of Hong Kong.

Ching, T. Y., Scollie, S. D., Dillon, H., et al. (2010). A cross-over, double-blind comparison of the NAL-NL1 and the DSL v4. 1 prescriptions for children with mild to moderately severe hearing loss. *International Journal of Audiology*, 49(S1), S4-S15.

Cox, R. M., & McDaniel, D. M. (1984). Intelligibility ratings of continuous discourse: Application to hearing aid selection. *Journal of the Acoustical Society of America*, 76(3), 758-766.

DePaolis, R. A., Janota, C. P., & Frank, T. (1996). Frequency importance functions for words, sentences, and continuous discourse. *Journal of Speech and Hearing Research*, 39, 714-723.

Dillon, H. (1993). Hearing aid evaluation: Predicting speech gain from insertion gain. *Journal of Speech and Hearing Research*, 36, 621-633.

Dillon, H. (1999). NAL-NL1: A new procedure for fitting non-linear hearing aids. *The Hearing Journal*, 52(4), 10-16.

Dodd, B. J., & So, L. K. H. (1994). The phonological abilities of Cantonese-speaking children with hearing loss. *Journal of Speech, Language, and Hearing Research*, 37, 671-779.

Dunn, H. K., & White, S. D. (1940). Statistical measurements on conversational speech. *Journal of the Acoustical Society of America*, *11*, 278-288.

Eisenberg, L. S., Dirks, D. D., Takayanagi, S., & Martinez, A. S. (1998). Subjective judgments of clarity and intelligibility for filtered stimuli with equivalent speech intelligibility index predictions. *Journal of Speech, Language, and Hearing Research*, 41, 327-339.

Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. Applied Psychological Measurement, 5(1), 105-112.

Fletcher, H., & Galt, R. H. (1950). The perception of speech and its relation to telephony. *Journal of the Acoustical Society of America*, 22, 89-151.

French, N. R., & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical society of America*, 19, 90-119.

Halle, M. G. W., Hughes J. P., & Radley, A. (1957). Acoustic Properties of Stop Consonants. *Journal of the Acoustical Society of America*, 29(1), 107-116.

Healy, E. W., Yoho, S. E., & Apoux, F. (2013). Band importance for sentences and words reexamined. *Journal of the Acoustical Society of America*, 133(1), 463-473.

Hinkelmann, K., & Kempthorne, O. (2005). Design and Analysis of Experiments, Volume 2, Advanced Experimental Design, Wiley.

Hornsby, B. W. (2004). The Speech Intelligibility Index: What is it and what's it good for?. *The Hearing Journal*, 57(10), 10-17.

Humes, L. E. (2002). Factors underlying the speech-recognition performance of elderly hearing-aid wearers. *Journal of the Acoustical Society of America*, *112*, 1112-1132.

Jin, I. K., Kates, J. M., & Arehart, K. H. (2014). The dynamic range for speech materials in Korean, English, and Mandarin: A cross-language comparison. *Journal of Speech, Language, and Hearing Research*, 57, 2024-2030.

Jung, A., F. (1961). Interviewer Differences among Automile Purchasers, *Journal of the Royal Statistical Society*. Series C (Applied Statistics), 10(2), 93-97.

Kates, J. M. (2013). Improved estimation of frequency importance functions. *Journal of the Acoustical Society of America*, 134(5), EL459-EL464.

Keidser, G., Dillon, H., Carter, L., & O'Brien, A. (2012). NAL-NL2 empirical adjustments. *Trends in amplification*, *16*(4), 211-223.

Killion, M. C., & Mueller, H. G. (2010). Twenty years later: A new Count-The-Dots method. *The Hearing Journal*, 63, 10-17.

Kim, M., & Pae, S. (2007). *Korean speech acquisition (In S. McLeod (Ed.)).* The international guide to speech acquisition, Clifton Park, NY: Thomson Delmar Learning, 472-482.

Lee, J. H., Jang, H. S., & Chung, H. J. (2005). A Study on Frequency Characteristics of Korean Phonemes. *Audiology*, 1, 59-66.

Leung, S. H., Luk, A., Liu, G. K. F., & Lun, C. S. (1992). An ARMA model for extracting Cantonese phoneme characteristics. In Proceedings of the Fourth Australian International Conference on Speech Science and Technology, St. John College, The University of Queensland, Brisbane, Australia, 692-697.

Li, F., Edwards, J., & Beckman, M. (2007, August). Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. *The XVIth International Congress of Phonetic Sciences,* 4, 917-920.

Lobdell, B. E., & Allen, J. B. (2007). A model of the VU (volume-unit) meter, with speech applications. *Journal of the Acoustical Society of America*, *121*, 279-285.

Macrae, J. H., & Brigden, D. N. (1973). Auditory threshold impairment and everyday speech reception. *Audiology*, 12, 272-290.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50, 940-967.

McLeod, S. (2007). Australian English speech acquisition (In S. McLeod (Ed.)). The international guide to speech acquisition, Clifton Park, NY: Thomson Delmar Learning, 241-256.

Mines, M. A., Hanson, B. F., and Shoup, J. E. (1978). Frequency of occurrence of phonemes in conversational English. *Language and speech*, 21(3), 221-241.

Moon, S. K., Kim, S. H., Mun, H. A., Jung, H. K., Lee, J. H., Choung, Y. H., & Park, K. H. (2008). The Korean Hearing In Noise Test. *International Journal of Audiology*, 47, 375-376.

Mueller, H. G., & Killion, M. C. (1990). An easy method for calculating the articulation index. *The Hearing Journal*, 43(9), 14-17.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95, 1085-1099.

Pavlovic, C. V. (1984). Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment. *Journal of the Acoustical Society of America*, 75, 1085-1099.

Pavlovic, C. V. (1987). Derivation of primary parameters and procedures for use in speech intelligibility predictions. *Journal of the Acoustical Society of America*, 82(2), 413-422.

Pavlovic, C. V. (1988). Articulation index predictions of speech intelligibility in hearing aid selection. *Asha*, *30*(6-7), 63-65.

Rajkumar, S., Muttan, S., Jaya, V., et al. (2013). Comparative Analysis of Different Prescriptive Formulae Used in the Evaluation of Real Ear Insertion Gain for Digital Hearing Aids. *Universal Journal of Biomedical Engineering*, 1(2), 32-41.

Rankovic, C. M. (1991). An application of the articulation index to hearing aid fitting. *Journal of speech, Language, and Hearing Research*, 34, 391-402.

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2009). The dynamic range of speech, compression, and its effect on the speech reception threshold in stationary and interrupted noise. *Journal of the Acoustical Society of America*, *126*, 3236-3245.

Sivian, L. J., Dunn, H. K., & White, S. D. (1931). Absolute amplitudes and spectra of certain musical instruments and orchestras. *Journal of the Acoustical Society of America*, *2*(3), 330-371.

Smit, A. B. (2007). General American English speech acquisition (In S. McLeod (Ed.)). The international guide to speech acquisition, Clifton Park, NY: Thomson Delmar Learning, 128-147.

Stiles, D. J., Bentler, R. A., & McGregor, K. K. (2012). The speech intelligibility index and the pure-tone average as predictors of lexical ability in children fit with hearing aids. *Journal of Speech, Language and Hearing Research*, *55*(3), 764-778.

Studebaker, G. A. (1985). A rationalized arcsine transform. *Journal of Speech, Language, and Hearing Research*, 28(3), 455-462.

Studebaker, G. A., & Sherbecoe, R. L. (1991). Frequency importance and transfer functions for recorded CID W-22 word lists. *Journal of Speech, Language, and Hearing Research*, 34, 427-438.

Studebaker, G. A., & Sherbecoe, R. L. (1993). Frequency-importance functions for speech recognition. *Acoustical factors affecting hearing aid performance*, 185-204.

Studebaker, G. A., Pavlovic, C. V., & Sherbecoe, R. L. (1987). A frequency importance function for continuous discourse. *Journal of the Acoustical Society of America*, 81, 1130-1138.

Studebaker, G. A., Sherbecoe, R. L., & Gilmore, R. L. (1993). Frequency importance and transfer functions for the Auditec of St. Louis recordings of the NU-6 word list. *Journal of Speech, Language, and Hearing Research*, 36, 799-807.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., et al. (1999). Monosyllabic word recognition at higher-than-normal speech and noise levels. *Journal of the Acoustical Society of America*, *105*, 2431-2444.

Thomas, T. W. C. (2005). The Effects of Occurrence Frequency of Phonemes on Second Language Acquisition: A Quantitative Comparison of Cantonese, Mandarin, Italian, German and American English, LIN1002 Invitation to Linguistics, Chinese University of HongKong, 1-20.

Tu, C. (1991). Speech spectrum analysis of English and Mandarin: Implications for hearing aid evaluation. Master's thesis, 1-57.

Wilde, G., & Humes, L. E. (1990). Applications of the articulation index to the speech recognition of normal and impaired listeners wearing hearing protection. *Journal of the Acoustical Society of America*, 87, 1192-1199.

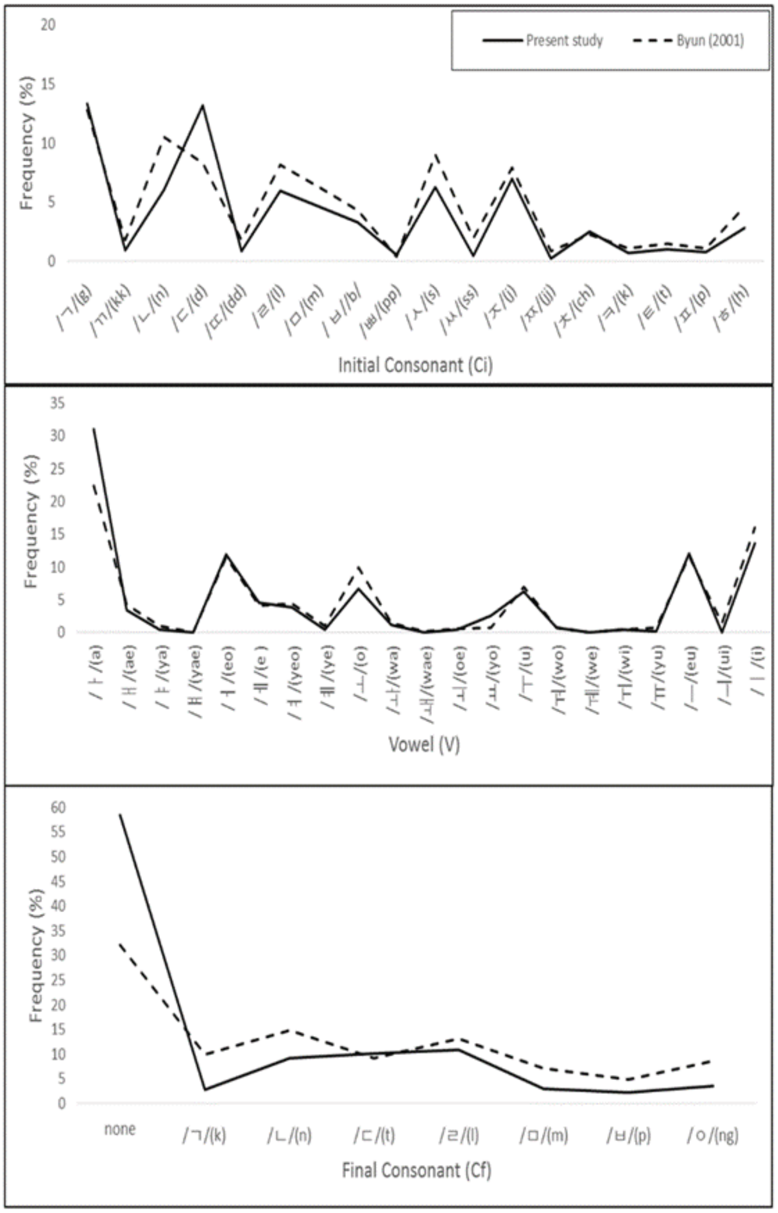
Wong, L. L. N., Ho, A. H. S., Chua, E. W. W., & Soli, S. D. (2007). Development of the Cantonese speech intelligibility index. *Journal of the Acoustical Society of America,* 121, 2350-2361.

Wong, L. L., Soli, S. D., Liu, S., et al. (2007). Development of the Mandarin hearing in noise test (MHINT). *Ear and hearing*, 28(2), 70-74.

Yang, Z. F., Lee, P. C., Chen, W. H., & Leu, J. G. (2011). Extension of structural watermarks based on balanced incomplete block designs. *Journal of Information Hiding and Multimedia Signal Processing*, 2(4), 354-365.

Yates, F. (1936). Incomplete randomized blocks. *Annals of Eugenics*, 7(2), 121-140.

APPENDIX A.



A. 1. Frequency of occurrence of phonemes which used in the present study (solid line) and studied by Byun (2001) (dotted line). Upper graph represents initial consonants (Ci), middle graph represents vowels (V), and lower graph represents final consonants (Cf). Byun (2001) investigated frequencies of occurrence of Korean phonemes from drama and news scripts (total 171,824 syllables).