http://dx.doi.org/10.17959/sppm.2020.26.1.89 

# Phonotactics and speech syllabification by Korean learners of English*

Miyeon Ahn
(Hankyong National University)

**Ahn, Miyeon. 2020. Phonotactics and speech syllabification by Korean learners of English.** *Studies in Phonetics, Phonology and Morphology* 26.1. 89-101. The purpose of this study is to explore Korean-speaking listeners' use of English phonetic cues in speech syllabification. By partially replicating the study of Coetzee (2011) with Korean learners of English, we found that native language knowledge plays an important role in determining the syllable structure of a speech stream. Structures ill-formed in English were perceptually modified by Korean listeners based on their native phonotactics. The results of the study suggest that the same acoustic stimuli are variously interpreted depending on the native language, and explain how phonological grammar works in cross-language speech perception. **(Hankyong National University, Assistant Professor)**

Keywords: Speech perception, Korean listeners, phonotactics, vowel epenthesis, English

## 1. Introduction

Speech sounds in an utterance are a continuously connected stream in which no clear linguistic boundaries are found (Klatt 1980, Mattys *et al.* 2005). In spite of the continuous aspects of speech sounds, language users perceive them as if they consist of discrete units by psychologically marking various linguistic boundaries onto the stream (Cole *et al.* 1980). Placing these linguistic boundaries such as segments, syllables or words allows the speech chunks to be broken down into smaller linguistic units, which makes language users exploit perceptual processes such as segment identification and word recognition.

How listeners identify those smaller linguistic units have long been studied and it has been found that human speech involves various types of linguistic information and listeners actively utilize the information that are scattered around the target

* This work was supported by a research grant from Hankyong National University in the year of 2018.

(McQueen 2005, Newman *et al.* 2011). According to the literature of speech perception, the linguistic information may be sublexically or lexically-driven (Gow and Gordon 1995, Davis *et al.* 2002). Davis *et al.* (2002) explains that phonetic variations such as English voiceless aspiration or dark [l] at the end of words are acoustic variants that are sublexical cues as well as phonotactic regularities such as no word-final lax vowels. On the other hand, Possible Word Constraints (PWC, Norris *et al.* 1997), which determine whether a piece of stream constitutes a possible word, is an example of lexically-driven accounts. Norris *et al.* (1997) and Norris *et al.* (2001) showed that in a word-spotting task such as *vuffapple* vs. *fapple*, the English word *apple* was more easily spotted in *vuffapple* since [f] in *fapple* is not a viable word. These sublexical and lexical cues are not independent but rather converged since listeners incorporate all sorts of information that are available from multiple sources (Mattys *et al.* 2005).

Recognizing native speech sounds and identifying the linguistic boundaries to mark segments, syllables or words is not much challenging because various allophonic and boundary cues are available and native listeners make full use of the cues during their speech process (Weber and Cutler 2006). The same kind of speech recognition, however, may not be successfully accomplished for non-native listeners, which reflects language experience is preconditioned in speech recognition. In addition, due to language-specific knowledge, language users with different native language backgrounds may interpret the same speech signal differently. The role of various lexical and sublexical cues in the perception of native listeners has been widely discussed and the contribution of each individual cue during perceptual process have been well examined so far. However, the interpretation of the same linguistic cues among listeners with distinct language background has been less clear.

In this study, we explore how non-native listeners display unique perceptual patterns that are different from native listeners. By examining Korean leaners of English (i.e., Korean listeners), we discuss how ill-formed structures are processed by function of the native phonotactics. To this end, we discuss the study of Coetzee (2011) who explored English listeners' allophonic cue use and partially replicate the same study with Korean listeners.[1] Focusing on the fact that English voiceless stops are aspirated in stressed-syllable initial only, Coetzee (2011) argued that English listeners would be sensitive to syllabifically conditioned allophonic cues. He explains

[1] We replicate the Experiment 1 of Coetzee (2011) by employing the audio files used in his Experiment 2 in Coetzee (2011).

that, unlike the VsC$^{h}$ sequence wherein the sC$^{h}$ sequence is followed by a vowel, $^{*}$[#sC$^{h}$ can is perceptually illegal sequence in English. Thus, English listeners perceptually modify these ill-formed structures by epenthesizing an illusory vowel so that the sequence such as [sp$^{h}$ika] would be perceived as [səp$^{h}$ika]. The similar structure with the preceding vowel such as [la-sp$^{h}$ika], on the other hand, does not involve vowel epenthesis since the word-medial [s] is resyllabified as a coda of the preceding syllable, which becomes a well-formed structure.

Two forced-choice experiments were designed to examine Korean listeners' perceptual patterns. These experiments test whether Korean listeners are sensitive to syllabifically conditioned allophonic cues as English listeners are and they investigate the role of Korean phonological grammar. Unlike English, aspiration results in phonemic contrast in Korean. Concerning that the difference of grammar would cause unique perceptual patterns, the comparison in cue use would allow us to investigate how Korean listeners deal with the same acoustic signal similar to or different from English listeners.

## 2. Experiment 1

### 2.1 Methods

#### 2.1.1 Participants

Forty paid participants were recruited around XX campus for the experiments. All of them were native Korean-speaking listeners and learned English as a foreign language. They reported that their stay of English-speaking countries such as USA or England was not longer than 30 months. They were in their 20s and the numbers of each gender were equally distributed.

#### 2.1.2 Stimuli and Predictions of Responses

The sound files adopted in this experiment were the same as used in Coetzee (2011). The twelve English non-words were recorded by a native speaker of English and they are provided in the second column of Table 1.

**Table 1. The stimuli adapted from Coetzee (2011)**

| Wd no. | Token 1 $CV_1CV_2C^hV_3C$ | Token 2 $CV_1CC^hV_3C$ | Token 3 $CV_1CCV_3C$ | Token 4 $CV_2C^hV_3C$ | Token 5 $CC^hV_3C$ | Token 6 $CCV_3C$ |
|---|---|---|---|---|---|---|
| 1 | [basək$^h$íp] | [bask$^h$íp] | [baskíp] | [sək$^h$íp] | [sk$^h$íp] | [skíp] |
| 2 | [fasət$^h$ík] | [fast$^h$ík] | [fastík] | [sət$^h$ík] | [st$^h$ík] | [stík] |
| 3 | [fisək$^h$án] | [fisk$^h$án] | [fiskán] | [sək$^h$án] | [sk$^h$án] | [skán] |
| 4 | [kusət$^h$íf] | [kust$^h$íf] | [kustíf] | [sət$^h$íf] | [st$^h$íf] | [stíf] |
| 5 | [lasəp$^h$íf] | [lasp$^h$íf] | [laspíf] | [səp$^h$íf] | [sp$^h$íf] | [spíf] |
| 6 | [lusət$^h$ám] | [lust$^h$ám] | [lustám] | [sət$^h$ám] | [st$^h$ám] | [stám] |
| 7 | [masək$^h$íf] | [mask$^h$íf] | [maskíf] | [sək$^h$íf] | [sk$^h$íf] | [skíf] |
| 8 | [masəp$^h$ál] | [masp$^h$ál] | [maspál] | [səp$^h$ál] | [sp$^h$ál] | [spál] |
| 9 | [misət$^h$ál] | [mist$^h$ál] | [mistál] | [sət$^h$ál] | [st$^h$ál] | [stál] |
| 10 | [pisək$^h$áf] | [pisk$^h$áf] | [piskáf] | [sək$^h$áf] | [sk$^h$áf] | [skáf] |
| 11 | [vesəp$^h$ím] | [vesp$^h$ím] | [vespím] | [səp$^h$ím] | [sp$^h$ím] | [spím] |
| 12 | [visəp$^h$áf] | [visp$^h$áf] | [vispáf] | [səp$^h$áf] | [sp$^h$áf] | [spáf] |

The original words (Token 1) were three-syllabled with $CV_1CV_2CV_3C$ structure of which the ultimate syllable (i.e., $V_3$) is stressed and the onset of the stressed syllable is aspirated as in [fisək$^h$án]. Token 2 stimuli were created by excising $V_2$ from Token 1 while Token 3 stimuli were done by excising both $V_2$ and the aspiration of the following consonant. Token 4 was created by removing the antepenultimate syllable (i.e., $CV_1$) from Token 1. The second vowel (i.e., $V_2$) was removed from Token 4 to create Token 5, and both $V_2$ and the aspiration were excised for Token 6. The whole process generated six token types for each word and 72 stimuli were created in total. See Experiment 2 of Coetzee (2011) for the detailed explanation of stimuli creation.

Token 1 of each word was then paired with Token 3 or Token 2 as in Experiment 1 in Coetzee (2011), which were referred to as Condition 1 and 3, respectively, in this study. Similarly, Token 4 was paired with Token 6 or Token 5 to be Condition 2 and 4 as in Table 2.

**Table 2. Stimuli pairing for same vs. different task and expected responses to target pairs[2]**

| Condition | Stimuli paired | | Eng listeners | Kor listeners |
|---|---|---|---|---|
| **Condition 1** | [Token 1 ~ 3] | [fisək$^h$án ~ fiskán] | Different | Different |
| **Condition 2** | [Token 4 ~ 6] | [sək$^h$án ~ skán] | Different | Different |
| **Condition 3** | [Token 1 ~ 2] | [fisək$^h$án ~ fisk$^h$án] | Different | Same |
| **Condition 4** | [Token 4 ~ 5] | [sək$^h$án ~ sk$^h$án] | Same | Same |

All the conditions consist of two tokens that are actually different. Since Tokens in Conditions 1, 2 and 3 are well-formed structures in English and the two tokens in each condition differs by one vowel, referring to the results of Coetzee (2011), English listeners would be able to relatively well distinguish the differences between given stimuli pairs. The listeners, however, would have difficulty in distinguishing the two tokens in Condition 4 since Token 5 is ill-formed so [sk$^h$án] would be perceptually modified to [sək$^h$án] by applying the phonological rule of $^*$#sk$^h$, which leads English listeners to perceive the two tokens to be different.

On the other hand, Korean listeners are expected to evaluate the conditions with different reasons. In Condition 1, Token 3 [fiskán] is considered to be ill-formed since neither a consonant cluster is allowed nor /s/ is an eligible coda in Korean. It thus would be perceptually modified to [fisəkán] which is still different from Token 1 [fisək$^h$án] due to the absence of aspiration. Therefore, while English listeners would determine Token 1~3 to be different due to the presence of $V_2$ vowel, Korean listeners' perception would be based on the presence of aspiration. The same rule applies to Condition 2. Condition 3 is where the difference between English vs. Korean listeners can be found. Token 2 [fisk$^h$án] would be perceived as [fisək$^h$án] after perceptual vowel epenthesis unlike English listeners who process the token as a legal percept [fis.k$^h$án]. Similarly, Token 5 of Condition 4 involves vowel epenthesis, which becomes the same as Token 4 [sək$^h$án].

### 2.1.3 Procedure

The participants were explained that they would listen to English non-words and that their task was to determine whether the two words were the same or different. They showed

[2] The expected responses of English listeners are based on Coetzee (2011).

their response by clicking their choices on the screen of a laptop with a mouse. The four conditions were presented to participants one at a time. A pair of each condition was randomly played and the next sound was not played until a participant made a choice. Each sound file was repeated three times and all conditions were counterbalanced. The whole experiment was conducted using a headset in a quiet room.

In order to measure listeners' sensitivity to the stimuli, *d*-prime value of *Signal Detection Theory* was calculated (Macmillan and Creelman 1991). The higher *d*-prime value corresponds to higher listeners' sensitivity. Higher sensitivity means listeners are sensitive to the difference of two tokens and they correctly make a decision in distinguishing them. Thus, the higher *d*-prime value can be interpreted as less epenthesis being introduced. When *Hit rate* and *False alarm* corresponds to 1 or 0, the loglinear approach was applied (Hautus 1995, Stanislaw and Todorov 1999). Two word groups (number 1 and 9) found to be incorrectly encoded during program scripting were excluded from the analysis.

### 2.2 Results and Discussion

Figure 1 shows the Korean listeners' mean *d'* values based on their responses to the four conditions. As in the figure, the *d'* values in Condition 1 and 2 were higher than the values in Condition 3 and 4, respectively. The *d'* value in Condition 2 was higher than those in Condition 3 and 4. The value was lowest in Condition 4.

The Korean listeners' responses were statistically analyzed both by subject and by item. Two one-way analysis of variance (ANOVA) was conducted and it was found that the differences in *d'* were significant in by subject analysis ($F(3, 156) = 10.465$, $p < 0.001$) but not in by item ($F(3, 36) = 2.817$, $p = .053$). Tukey HSD further showed that Condition 1 and 3 ($p < .05$) and Condition 2 and 4 ($p < .05$) were significantly different but those significance was not found between other conditions.

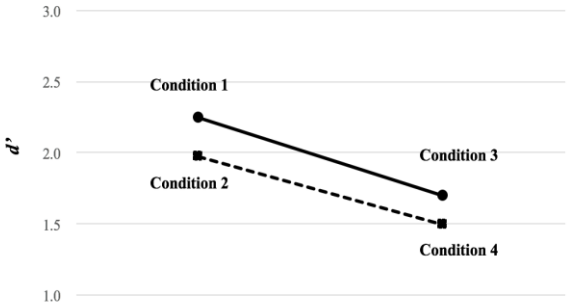


**Figure 1. The analysis of Korean listeners' *d'* value of Experiment 1**

The highest *d'* value in Condition 1 indicates that Korean listeners are sensitive to the difference of Token 1 and 3 which involves both vowel and aspiration differences. It is rather straightforward that the more cue differences lead to easier distinction since listeners would best use the cues that are available. The relatively low *d'* value in Condition 3, on the other hand, shows that the listeners had difficulties in distinguishing [fisək$^{h}$án ~ fisk$^{h}$án] that are differed by only the presence of the vowel. This result allows us to infer that Korean listeners' decision in Condition 1 was based on the aspiration rather than a vowel and it suggests that the listeners perceptually modified one of the tokens of Condition 3 to perceive them as the same token. In other words, the listeners either perceptually inserted a vowel in Token 2 [fisk$^{h}$án] to be same as [fisək$^{h}$án] or perceptually deleted a vowel in Token 1 [fisək$^{h}$án] so that it becomes same as Token 2 [fisk$^{h}$án]. The same discussion applies to Condition 2 and 4. The listeners' decision is likely to be due to the aspiration in Condition 2 and one of the tokens in Condition 4 was modified to be the same as the other.

## 3. Experiment 2

In order to examine whether perceptual modification is processed to introduce vowel epenthesis or deletion of illegal consonants, we designed Experiment 2 in which listeners were asked to identify the number of syllables of each token.

## 3.1 Methods

### 3.1.1 Participants

The subjects of Experiment 1 also participated in Experiment 2.

### 3.1.2 Stimuli and Predictions of Responses

The stimuli used in Experiment 2 were the same as those in Experiment 1. The 72 tokens in Table 1 were provided one by one instead of stimuli pairs. While the stimuli in Token 1 [fisək$^{h}$án] and 4 [sək$^{h}$án] were well-formed, those in the rest (i.e., Token 2, 3, 5 and 6) were ill-formed structures in Korean. As illustrated in Table 3, Korean listeners, thus, were expected to perceptually modify those in Token 2, 3, 5 and 6. If perceptual modification introduces vowel epenthesis so that listeners perceive an illusory vowel, the extended syllable structure leads to the increase of the number of syllables. The incorrect perception of the extended number of syllables is expected to be lower in well-formed structures than ill-formed.

**Table 3. Predictions for Korean listeners' syllable count task**

| Token | Example Stimuli | Number of syllables | Eng listeners | Kor listeners |
|---|---|---|---|---|
| **Token 1** | [fisək$^{h}$án] | 3 | 3 | 3 |
| **Token 2** | [fisk$^{h}$án] | 2 | 2 | 3 |
| **Token 3** | [fiskán] | 2 | 2 | 3 |
| **Token 4** | [sək$^{h}$án] | 2 | 2 | 2 |
| **Token 5** | [sk$^{h}$án] | 1 | 2 | 2 |
| **Token 6** | [skán] | 1 | 1 | 2 |

### 3.1.3 Procedure

Each token was randomly played and the laptop monitor showed 5 syllable count choices from 1 to 5. The participants were instructed to identify the number of syllables, after listening to each word, by clicking one of the numbers on the screen. The whole sound file was repeated twice. Words numbered 4, 5, 7, 10 and 12 that involve word-final [f] were excluded from data analysis as the sound [f] is not an eligible coda in Korean and words with this sound may trigger an additional vowel at the end of the words.

### 3.2 Results and Discussion

Figure 2 describes the percentage of correct vs. error in syllable count task of Korean learners of English. It is represented that the % of correct was higher than % of error in Token 1 and 4 but it was consistently lower in the rest, i.e., Token 2, 3, 5 and 6. It should be noted that tokens with high % of correct were well-formed structures while ones with low % of correct were ill-formed.

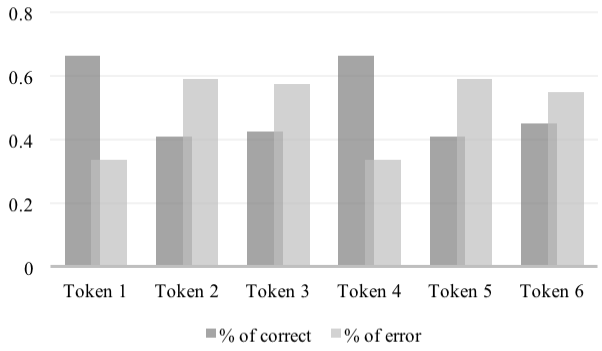


**Figure 2. The percentage of correct and error in syllable count**

Figure 3 presents the error patterns in detail. Both Token 2 [fisk$^{h}$án] and Token 3 [fiskán] are two-syllabled words, but 44% of the total responses identified them as three-syllabled, which means that these two tokens incurred vowel epenthesis. Similarly, Token 5 [sk$^{h}$án] and Token 6 [skán] were considered to be extended to two-syllabled words. In all Token conditions, the major source of errors was attributed to the fact that listeners identified those tokens with an additional syllable. It can be inferred then that Korean listeners perceptually modified the ill-formed by adopting vowel epenthesis rather than vowel deletion.

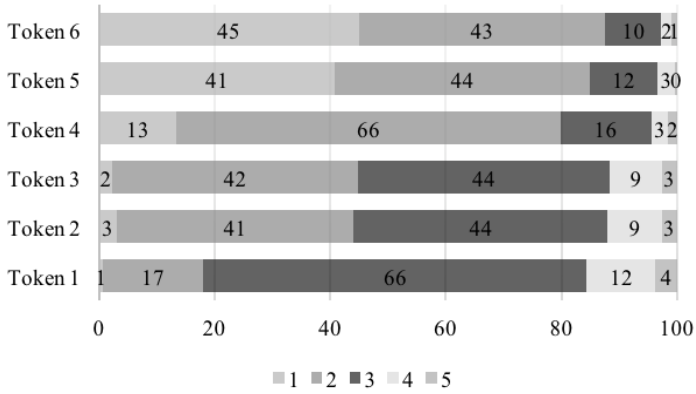


**Figure 3. The percentages of indicated number of syllables**

It also should be noted that although a number of responses in both Token 5 and 6 is related to vowel epenthesis, the foremost response was 1 syllable in Token 6 while it is 2 in Token 5. Token 5 [sk$^{h}$án] is against English phonotactics and it is the token that incurs an epenthetic vowel to English listeners. Although being nonnative listeners, the participants presumably adopted linguistic knowledge of English $^{*}$[#sk$^{h}$ to some extent.

## 4. General Discussion

The results of the current study coincide with the discussions of the literature in cross-language speech perception in that listeners modulate their perception based on their native language knowledge (Best 1995) and that first-language phonotactics influences nonnative listening (Weber and Cutler 2006). It thus can be inferred that listeners with a different language background interpret the same acoustic details differently.

In the present study, Korean listeners' cue use was compared with English listeners. English aspiration is applied only in syllable-initial, which is crucially adopted by English listeners. When an ill-formed structure such as [sk$^{h}$án] is heard, it is perceptually modified to [sək$^{h}$án] so that the aspirated stop appears syllable-initially. Although [fisk$^{h}$án] is acoustically similar to [sk$^{h}$án] except the initial syllable, perceptual modification to [fisək$^{h}$án] is not applied since the stop can still be syllable-initial.

On the other hand, Korean listeners exploit perceptual epenthesis with different reasons. Vowel epenthesis in Condition 3 or 4 of Experiment 1 or Token 5 [sk$^{h}$án] or 6 [skán] of Experiment 2 is incurred not because of aspiration rule of English but because of restrictions regarding consonant clusters such as $^{*}$CC that has to do with syllable structures. In addition, Korean listeners also epenthesize a vowel in Condition 1 or 2 of Experiment 1 or Token 2 or 3 of Experiment 2 where English listeners do not. Vowel epenthesis in these contexts such as [fisk$^{h}$án] has to do with a different sort of phonological rules regarding consonantal contact that does not allow the co-occurrence of certain sounds. According to Kabak and Idsardi (2007), it is suggested that, among various types of phonotactic knowledge, syllable structure restrictions result in more vowel epenthesis than restrictions on consonantal contact. Relatively low *d'* values in Condition 3 and 4 of Experiment 1 accord with their discussion in that Korean listeners are more sensitive to consonantal contact rather than syllable structure.

Based on the results of the two experiments, we showed that language users' cue use varies depending on their native language. In contrast to English listeners who adopt the presence of aspiration as English allophonic cues on their speech perception, the same information was used as phonemic cues to Korean listeners. It should be further discussed that whether Korean listeners would switch their phonemic cue use into allophonic as their fluency of English increases.

## REFERENCES

BEST, CATHERINE T. 1995. A direct realist view of cross-language speech perception. In W. Strange (ed.). *Speech perception and Linguistic Experience*. 171-204. York Press.

COETZEE, ANDRIES W. 2011. Syllables in speech processing: Evidence from perceptual epenthesis. In Charles E. Cairns and Eric Raimy (eds.). *Handbook of the Syllable* (*Brill's Handbooks in Linguistics* vol. 1). 295-328. BRILL.

COLE, RONALD A., JOLA JAKIMIK and WILLIAM E. COOPER. 1980. Segmenting speech into words. *Journal of the Acoustical Society of America* 64, 1323-1332.

DAVIS, MATTHEW H., WILLIAM D. MARSLEN-WILSON and M. GARETH GASKELL. 2002. Leading up the lexical garden path: segmentation and ambiguity in

spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 28.1, 218-244.

GOW, DAVID W. and PETER C. GORDON. 1995. Lexical and prelexical influences on word segmentation. *Journal of Experimental Psychology: Human Perception and Performance* 21, 344-359.

HAUTUS, MICHAEL J. 1995. Corrections for extreme proportions and their biasing effects on estimated values of *d'*. *Behavior Research Methods, Instruments, & Computers* 27, 46-51.

KABAK, BARIS and WILLIAM IDSARDI. 2007. Perceptual distortions in the adaptation of English consonant clusters: syllable structure or consonantal contact constraints? *Language and Speech.* 50, 23-52.

KLATT, DENNIS H. 1980. Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (ed.). *Perception and production of fluent speech.* 243-288. Hillsdale, NJ: Erlbaum.

MCQUEEN, JAMES M. 2005. Speech perception. In K. Lamberts and R. Goldstone (eds.), *The Handbook of Cognition.* 255-275. London: Sage Publications.

MATTYS, SVEN L., LAURENCE WHITE and JAMES F. MELHORN. 2005. Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General* 134.4, 477-500.

MACMILLAN, NEIL A. and DOUGLAS C. CREELMAN. 1991. *Detection Theory: A User's Guide*. Psychology Press.

NEWMAN, ROCHELLE S., JAMES R. SAWUSCH and TYLER WUNNENBERG. 2011. Cue and cue interactions in segmenting words in fluent speech. *Journal of Memory and Language* 64, 460-476.

NORRIS, DENNIS, JAMES M. MCQUEEN, ANNE CUTLER and SALLY BUTTERFIELD. 1997. The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology* 34, 191-243.

NORRIS, DENNIS, JAMES M. MCQUEEN, ANNE CUTLER, SALLY BUTTERFIELD and RUTH KEARNS. 2001. Language-universal constraints on speech segmentation. *Language and Cognitive Processes* 16, 637-660.

STANISLAW, HAROLD and NATASHA TODOROV. 1999. Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers* 31, 137-149.

WEBER, ANDREA and ANNE CUTLER. 2006. First-language phonotactics in second-language listening. *Journal of Acoustical Society of America* 119.1, 597-607.

Miyeon Ahn (Assistant Professor)
Department of English
Hankyong National University
327 Jungang-ro, Ansung-si
Gyeonggi 17579, Republic of Korea
e-mail: mahn@hknu.ac.kr

Received: March 20, 2020
Revised: April 25, 2020
Accepted: April 27, 2020