



# The acquisition of boundary tones in spontaneous speech by Korean learners of English\*

Wook Kyung Choe\*\*

*College of General Education, Dong-A University, Busan, Korea*

## Abstract

The current study was designed to investigate which type of phrase boundary tones high-intermediate Korean learners of English used in their spontaneous speech. These boundary tones were compared to those used in native speakers' spontaneous speech to examine whether the learners successfully acquired the use of boundary tones. To achieve this purpose, 10 Korean learners of English and four native speakers of English participated in the current study. The participants were asked to summarize the stories of short videos, and the tonal and the phrasing patterns of the obtained spontaneous speech were analyzed using Tone and Break Indices (ToBI) transcription conventions. The results indicated that both the native speakers and the Korean learners frequently marked their intonational phrase boundaries with high boundary tones. However, regarding the prosodic phrase positions within a sentence, Korean learners frequently used steep rising tones (i.e., H-H%) while native speakers used gradual rising tones (i.e., L-H%) for sentence-final intonational phrases. Overall, the findings suggested that high-intermediate Korean learners understood the forward-looking function of the high boundary tones and that they were able to make use of these tones to mark intonational phrases in their spontaneous speech.

**Keywords:** prosodic acquisition, boundary tones, phrase accents, spontaneous speech, Korean learners of English

## 1. Introduction

Prosody, one of the critical aspects of spoken language, contains and is used to express various linguistic and paralinguistic information. Due to its abundant linguistic information, studies on second language acquisition suggested that non-native-like realization of prosody impacted on the degree of foreign-accentedness of learners' speech (e.g., Holden & Hogan, 1993;

Munro, 1995; Munro & Derwing, 1999). Some studies even argued that the inappropriate prosody by non-native speakers more negatively affected the extent of intelligibility and foreign-accentedness than the inappropriate segmental production (e.g., Anderson-Hsieh et al., 1992; Herwings, 1995; Trofimovich & Baker, 2006). Among different aspects of prosody, many of the studies on prosodic acquisition have examined the realization of word-level prosody by language learners (e.g., Anderson-Hsieh, et

\* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A5A8028548).

\*\* regina1004@dau.ac.kr, Corresponding author

Received 14 November 2020; Revised 12 December 2020; Accepted 14 December 2020

© Copyright 2020 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

al., 1992; Trofimovich & Baker, 2006). Therefore, the current study was designed to investigate the acquisition of phrase-level prosody.

Prior to investigating the acquisition of phrase-level prosody by second language learners, a brief introduction of the structure and characteristics of prosodic phrases should be provided. Based on how to define prosodic phrases, different theories suggest the different numbers and structures of prosodic phrases. Since the current study used Tone and Break Indices (ToBI) transcription conventions (Beckman & Ayers, 1997; Beckman & Hirschberg, 1994; Beckman et al., 2005) to analyze the prosodic structure of the obtained data, it followed prosodic phrase structures focusing on suprasegmental features (Beckman & Pierrehumbert, 1986). In this theory, the highest prosodic phrase is an Intonational Phrase (hereafter IP), whose existence and nature are agreed upon among most of the theories. An IP is defined by a following pause, pre-boundary lengthening, and a complete intonational contour (e.g., Selkirk, 1984). Despite the consent on the definition and the existence of IPs among theories, the existence, name, and nature of mid-level prosodic phrases varied with different theories. The mid-level prosodic phrase assumed in the current study is an intermediate phrase (hereafter ip), which is defined by one or more pitch accents and a phrase accent.

Among several components of phrase-level prosody, intonation—including pitch accents, intonational contours, phrase accents, and boundary tones—is known to indicate discourse information of utterances such as a sentence type, a focus, and an information structure (e.g. Couper-Kuhlen, 2001). Especially, as for the intonational contours, falling tones are usually known to mark declarative sentences and/or the completion of an utterance, while rising tones are known to mark interrogative sentences and/or the continuity to the upcoming phrases (e.g., Ladd, 1996; Ohala, 1983). Specifically, the use of High phrase accents (H-) for ips and High boundary tones (H%) for IPs in declarative utterances is considered to show either the speakers' uncertainty of the contents (Smith & Clark, 1993; Shokier, 2008; Tomlinson Jr. & Fox Tree, 2011) or the forward-looking function (Du Bois et al., 1993; Fletcher et al., 2002; Pierrehumbert & Hirschberg, 1990). For example, Smith & Clark (1993) found that when speakers were not sure about the contents of their speech, they tended to use rising tones. In addition, Shokier (2008) suggested that listeners interpreted an IP with a H-H% tone as a statement with less certainty than an IP with a L-H% tone. With respect to the forward-looking function, Pierrehumbert & Hirschberg (1990) argued that speakers marked their ips with H- and their IPs with H% when the currently-speaking prosodic phrase was related with and continued to the upcoming prosodic phrase. Also, Fletcher & colleagues (2002) examined the relationship between the boundary tones and forward- or backward-looking functions in Australian English. Their study revealed that H-H% tones were used to show forward-looking functions such as continuation while H\*L-H% tones were used to show backward-looking functions such as requesting information.

This function of phrase accents and boundary tones regarding discourse information, however, has not been studied much in the second language acquisition of prosody. One study, Huang & Jun (2011) investigated the use of boundary tones by Mandarin Chinese learners of English, and indicated that the learners not only put more prosodic phrase boundaries than the native speakers, but also marked their IPs with High boundary tones (H%) more frequently than the native speakers. Similar patterns were found in the speech

by Korean learners of English (Choe, 2016; Lee, 2005). For example, Lee (2005) examined the use of phrase accents and boundary tones in the sentences with coordinate or subordinate conjunctions. She found that Korean learners put a prosodic phrase boundary where there was a conjunction and then marked their ips with H- tones and IPs with H% tones. In addition, Choe (2016) studied the effect of sentence position (either sentence-internal phrases or sentence-final phrases) on the realization of tonal patterns in Korean learners' English read speech. The results revealed that the learners put prosodic phrase boundaries more often than the native speakers and that the learners marked sentence-internal IPs with High boundary tones (L-H% or H-H%) significantly more often than the native speakers did. Both of the results in Lee (2005) and Choe (2016) suggested that Korean learners appropriately understood the meaning of High boundary tones (i.e., forward-looking functions) although they chunked their sentences into prosodic phrases more frequently than the native speakers. The results also proposed that the Korean learners were able to successfully realize High boundary tones to mark sentence-internal phrase boundaries. Other research on the realization of boundary tones by Korean learners of English, however, showed that the learners did not fully understand the meanings of different boundary tones and failed to use the native-like boundary tones (Lee, 2008; Park et al., 2000). Specifically, Park and colleagues (2000) asked the learners to read the same sentences, each of which had two different intentions: one with certainty and the other with uncertainty. The results showed that in the native speakers' speech, L-L% boundary tones were used to mark the end of the sentence with certainty, while L-H% tones were used to mark the end of the sentence with uncertainty. However, the Korean learners used L-L% for both versions regardless of how sure the speaker was about the statement. On the other hand, Lee (2008)'s study presented that Korean learners successfully marked their sentence-internal IPs with either L-H% or H-L% tones to execute forward-looking functions. However, the learners had hard time to mark the end of interrogative sentences with H-H% tones, but rather used L-H% tones for questions. Altogether, whether the learners successfully understand the discourse information of boundary tones and realize the appropriate boundary tones at the right place may depend on the different discourse functions of boundary tones. That is, based on the four studies, Korean learners of English seem to successfully understand and use the High boundary tones to encode a forward-looking function. However, they did not successfully acquire which boundary tone to use for more "sentence-level" discourse information such as types of sentences and/or the degree of certainty.

As in the aforementioned studies, almost all of the research regarding the acquisition of intonation by Korean learners of English was based on the read speech data, and most of them even focused on the prosodic structures of isolated sentences. Nonetheless, in a real-life situation, second language learners would have much more opportunity to have spontaneous speech than read speech as well as their fluencies and proficiency levels would often be judged by their spontaneous speech. In addition, considering the different intonational patterns between read and spontaneous speech, the realization of intonation in the second language spontaneous speech needs to be examined. The previous research on the acoustic differences between read and spontaneous speech indicated that spontaneous speech had its distinctive intonational patterns, such as

F0 ranges, the degree of downstepping, and the use of pitch accents and boundary tones (e.g., De Ruiter, 2015; Face, 2003; Kim, 2019; Lieberman et al., 1985). For example, Face (2003) analyzed the spontaneously produced 150 declarative sentences from a Spanish corpus and indicated that sentence-final falling tones did not as frequently occur in spontaneous speech as in read speech. De Ruiter (2015) asked the same set of German speakers first to make a story based on picture books and then to read the text of the same picture book. The study then compared the intonational contours in the spontaneous and read speech, and found that speakers marked sentence-final IPs with Low boundary tones in read speech while they used High boundary tones for sentence-final IPs in spontaneous speech to show their continuity. In total, speakers tend to use rising tones (or more precisely High boundary tones) more often in their spontaneous speech than in read speech. In spontaneous-speech situations, a speaker is more likely to put prosodic phrase boundaries to think about what to say or to retrieve the upcoming lexical words. A speaker could also have more chances to show the listeners that he or she is not done yet for the currently-speaking phrases and/or that the upcoming phrase is highly related to the current one. Due to these unique situations for spontaneous speech, a speaker should use High boundary tones at the end of prosodic phrases to execute the forward-looking function of High boundary tones.

The purpose of the current study was to investigate boundary tone patterns in the spontaneous speech produced by Korean learners of English. Specifically, the current study aimed to compare the patterns of boundary tones in native speakers' spontaneous speech with those in learners' spontaneous speech, and then to understand whether Korean learners successfully acquired the discourse meanings and the way of using boundary tones. To achieve this goal, the present study collected spontaneous speech data by asking the speakers to summarize video clips, and then analyzed their prosodic structures using Mainstream American English MAE-ToBIs transcription conventions.

## 2. Methods

### 2.1. Participants

Twelve Korean learners of English (hereafter KL) and four native speakers of English (hereafter NS) participated in the current study. All were female speakers. The participants in NS group were the instructors who taught English at universities in Korea. They were in their 30s and 40s (average age=39.75). Three of them are from the United States of America (1 from California, 1 from Michigan, and 1 from Virginia), and one is from Canada<sup>1</sup> (British Columbia), so all of them are considered to speak North American English. The participants in KL group were undergraduate students at a university located in Seoul. The two learners' data were excluded since their spontaneous speech data either mostly consisted of lists of words rather than complete sentences or were too short to analyze the prosodic structures of the speech. None of the 10 learners reported to have spent more than 11 months in English-speaking countries.

Their English proficiency level was determined by the self-reported TOEIC scores. The average TOEIC score of the 10 Korean speakers were 871, ranging from 735 to 970. These TOEIC scores led us to decide that their English was slightly above an intermediate level but did not yet reach a near-native level. Therefore, the Korean learners in the current study could be considered as high intermediate learners. Each participant in both groups received a 10,000 won gift certificate for their participation.

### 2.2. Procedures

Each participant was asked to watch a video and then to tell the experimenter the story of the video. Each participant was asked to do this with five different video clips. Each of the five video clips did not exceed 5 minutes. Two were cartoons (1 from Tom and Jerry, and 1 from the Simpsons); 1 was an animated film (from Monster University); 1 was the movie for children (from Charlie and the Chocolate Factory), and 1 was a sitcom (from Friends). These videos were chosen since the selected parts contain relatively straightforward and simple stories enough for the English learners to understand the contexts quite easily even without fully understanding the languages used in the videos. For the same reason, the cartoons did not contain any verbal conversations and the other ones were selected to minimize the characters' verbal conversations. This was also to minimize the possibilities that the participants remember and use the same words and expressions from the movie and then to make the participants use their own words. More importantly the videos with a few verbal conversations could let the speakers use their own prosodification to summarize the stories.

The experiment was conducted in a quiet meeting room or in a quiet office in a university. The participants used a head-worn microphone (Shure SM35-XLR) and their speech was digitally recorded to a Marantz PMD 661 MKIII.

### 2.3. Analysis

All the recorded speech was first dictated by a research assistant, who was not phonetically trained. The research assistant was asked not only to dictate the whole speech, but also to report the audible pauses and other miscellaneous information such as fillers, self-corrections, speech errors, coughing, laughing, and the experimenter's interruption. An incomprehensible word or phrase was left as a blank so that the experimenter could fill it up later. These dictations and information on the participants' speech were recorded as a Microsoft Word format.

After all the dictation processes were completed, the author determined the prosodic structures and information of all the participants' speech using Mainstream American English MAE-ToBIs transcription conventions (Beckman & Ayers, 1997; Beckman & Hirschberg, 1994; Beckman, et al., 2005; Veilleux et al., 2006). Pitch tracks of the speech data were examined with the Praat software (Boersma & Weenink, 2014). The author faithfully followed the criteria in ToBI transcription convention (e.g., Veilleux, et al., 2006) to determine the types of tones—pitch

---

<sup>1</sup> Although one study argued that different dialects of American English showed different prosodic patterns, the results of the study indicated that the difference was mainly about the speech rate and the places of the pitch accents (Clopper & Smiljanic, 2011). Since the current study deals with the use of boundary tones, the four native speakers were considered to produce utterances without cross-dialectal differences regarding boundary tone patterns.

accents, phrase accents, and boundary tones—and the break indices. Among the detailed criteria in ToBI conventions that the author followed, two—one about High boundary tones and the other about the break indices—would be specifically mentioned here. First, regarding the differences between L-H% and H-H%, the High phrase accent (H-) is known to trigger “an ‘upstep’ (a local raising of the pitch range to the end of the phrase)” (Beckman & Ayers, 1997:18). Therefore, to differentiate H-H% from L-H%, the author marked H-H% when there was the second upstep like in the middle figure in Figure 1 or when there was a sharp rising in which the F0 value went much above the highest F0 value of the current IP as in the bottom figure in Figure 1. With respect to the breaks, the break index 4—for Intonational Phrase (IP)—was marked where there were (1) one or more pitch accents, (2) a phrase accent and a boundary tone, and (3) a final lengthening followed by a pause that was longer than 200 ms of silence (Krivokapić, 2007). On the other hand, an Intermediate Phrase (ip) by the break index 3 indicates less disjoint from the following phrase, so both tonal and durational cues for ips are known to be weaker than those for IPs (Veilleux, et al., 2006). That is, the break index 3 was marked where there were one or more pitch accents and a phrase accent but no following pause. Regarding the durational cues, however, Veilleux and colleagues argued that it is necessary to mark the break index 3 where there was an obvious pitch accent and final lengthening, which was weaker than the final lengthening at the end of IP. Note that the current study showed more ip boundaries than other speech data, which was due to the frequent elongation in spontaneous speech, and the example of the break indices 3 are shown in the top and the middle figures in Figure 1.

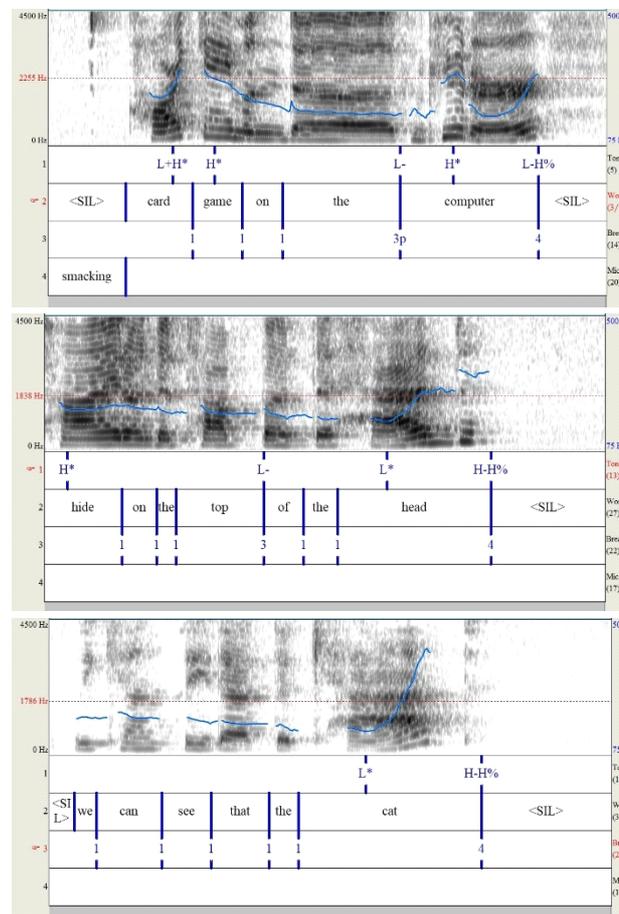


Figure 1. Prosodically-transcribed speech samples using MAE-ToBI transcription conventions

### 3. Results

Before analyzing the prosodic phrase structures and boundary tones of the Korean learners' and the native speakers' spontaneous speech data, the durations of all 70 spontaneously produced stories (hereafter a story indicates the spoken speech data for each video) of the two groups were measured and then compared. The duration of a story was calculated from the initiation of a vocal pulse of the participant's first word to the end of a vocal pulse of the participant's last word regardless of the different numbers and duration of story- internal pauses and disfluencies across stories. The statistical analysis revealed that the native speakers' speech was much longer ( $M=113.56$ ,  $SD=68.93$ ) than the Korean learners' one ( $M=70.06$ ,  $SD=34.89$ ), and a Mann-Whitney U test indicated that this difference was significant [ $U=266.0$ ,  $p=.002$ ].

The native speakers' longer speech naturally led us to assume that they told the stories with more words than the learners. However, since the duration in the current study contained pauses and disfluencies within a story, the longer speech might also be due to more pauses or disfluencies. Therefore, the next analyses were to find out the details of the two groups' spontaneous speech. Spontaneous speech is known to contain frequent disfluencies due to lexical retrieval, self- correction, fillers, and repetitions, and these so-called incomplete phrases or utterances were more frequently found in the learners' one (e.g., Bhat & Yoon, 2015; Foster et al., 2000). Due to these characteristics of spontaneous speech, simply

counting the number of words might contain all the fillers, repetitions, errors, and self-corrections, and so might mislead how fluent the speech was. For a similar reason, it is hard to clearly define an “utterance,” which is the largest speech domain with one or more IPs. Moreover, since fillers should be considered as a part of an utterance (Candea et al., 2005), the current study would use a “sentence” instead of an utterance. A “sentence” was determined only by its syntactic structure, which contains a subject and a verb and has a complete meaning regardless of subtle grammatical errors. From the dictation and the impression by the experimenter, common syntactic characteristics were noted in the obtained spontaneous speech data. One is that the speakers frequently put pauses and fillers to retrieve the appropriate words, to remember the contents of the video, and to correct speech errors. Therefore, the stories often had sentence fragments, and the current analyses did not consider these fragments as a “sentence.” Another criterion for determining a sentence was about the use of coordinate conjunctions between sentences. Many speakers in the current study, regardless of the groups, frequently put coordinate conjunctions such as “and” and “so” when they started a new sentence. Although the two clauses connected by a coordinate conjunction should be considered as one sentence, the speakers in the current study often started most of the sentences with one of these conjunctions. These patterns led us to assume that the speakers in the current study used these words as fillers rather than to grammatically connect two clauses into one sentence. Therefore, the two clauses connected a coordinate conjunction were determined as two different sentences in the current analyses.

A Mann-Whitney *U* test was conducted with the number of sentences per speaker and story as a dependent variable and the group as an independent variable. The test yielded a significant group difference for the number of sentences [ $U=126.0, p<.001$ ]. Specifically, the native speakers produced significantly more sentences ( $M=23.30, SD=15.33$ ) than the Korean learners ( $M=9.08, SD=4.00$ ). Therefore, it is possible to argue that the native speakers produced much more sentences than the Korean learners, which caused longer stories.

The next analyses were to compare the numbers of prosodic phrases in the spontaneous speech of the two groups. The ips and IPs consisting of only fillers, self-corrections, or repetitions were noted but excluded from the later analyses. Mann-Whitney *U* tests were conducted with the group as an independent variable and the number of ips and IPs as dependent variables. The number of ips and IPs were calculated within a story and a speaker. The results showed the significant group differences for the number of ips [ $U=154.0, p<.001$ ] and IPs [ $U=319.5, p=.019$ ]. In detail, the native speakers produced their spontaneous speech with significantly more ip boundaries and IP boundaries ( $M=17.20, SD=14.92$  for ips;  $M=39.60, SD=21.08$  for IPs) than the Korean learners ( $M=4.92, SD=4.08$  for ips;  $M=28.36, SD=14.60$  for IPs). The detailed comparison indicated that the Korean learners were less likely to put ip boundaries but rather to put IP boundaries when they chunked their speech. This less frequent use of ips is quite outstanding especially compared with the results from Choe (2016)’s prosodic phrasing in read speech. When the Korean learners read a passage, they seemed to put ip- and IP-boundaries at the similar rate.

Another finding was that the native speakers produced significantly more ips and IPs than the Korean learners. Combined with the analysis about the number of sentences, this result was because

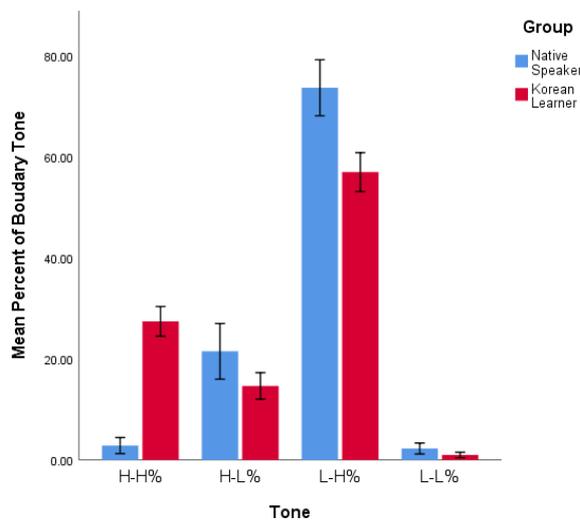
the native speakers produced so many sentences that their speech contained many ips and IPs. To confirm whether the native speakers’ frequent chunking was not due to disfluencies but due to the larger number of sentences, the total number of ips and IPs was divided into the number of sentences per speaker and story. This analysis indicated that the average number of prosodic phrases per sentence for the NS group was 2.63 ( $SD=0.58$ ), while that for the KL group was 3.61 ( $SD=0.85$ ). The result confirmed that the native speakers’ more prosodic phrasing was because they produced much more sentences than the learners. It also suggested that the Korean learners frequently chunked their spontaneous speech into different prosodic phrases similar to the patterns in the Korean learners’ read speech (Choe, 2016).

The next analyses aimed to examine the frequency of phrase accents and boundary tones in spontaneous speech by the two groups. As the total numbers of prosodic phrases varied with speakers and stories, the percentage of the use of each boundary tone was calculated within a story and used for the analyses. The group effect was tested in Mann-Whitney *U* tests with the percentages of 2 phrase accents and 4 boundary tones (H- and L- for ips and H-H%, H-L%, L-H%, and L-L% for IPs) as dependent variables. The statistical tests yielded the significant effect of group only on the four boundary tones of intonational phrases [ $U=984.5, p<.001$  for H-H%;  $U=332.0, p=.029$  for H-L%;  $U=182.0, p<.001$  for L-H%;  $U=331.5, p=.011$  for L-L%]. Table 1 below shows the average and the standard deviation of each phrase accent and boundary tone for each group, and Figure 2 shows the significant group differences for four boundary tones.

**Table 1.** The mean and standard deviation of the percentage of boundary tones for the NS and the KL group.

Boundary tones	Group	Mean	Standard deviation
H-	NS	67.66	19.16
	KL	73.85	28.87
L-	NS	31.62	18.74
	KL	26.15	28.87
H-H%*	NS	2.86	3.39
	KL	27.43	10.34
H-L%*	NS	21.50	11.76
	KL	14.65	9.24
L-H%*	NS	73.66	11.85
	KL	56.97	13.54
L-L%*	NS	2.28	2.31
	KL	0.99	1.90

An asterisk(\*) after a boundary tone indicates a significant difference between the group.



**Figure 2.** Mean percentage of four boundary tones by group (KL vs. NS) with 95% confidence interval error bars.

Table 1 and Figure 2 showed that the most frequently used boundary tone for intonational phrases was L-H% for both NS and KL groups. Also, the results argued that the native speakers rarely used H-H% and L-L% tones to mark their IPs, but they marked 95% of their IPs with either L-H% and H-L% tones. However, the usage of boundary tones for the Korean learners was slightly different in that the second most frequently used boundary tones for the Korean learners were H-H%, which were rarely used in the native speakers' spontaneous speech.

The previous research on the use of boundary tones in Korean learners' read speech revealed that Korean learners could successfully understand the forward-looking function of High boundary tones for sentence-internal phrases and were able to appropriately utilize these tones. The Korean learners, however, were reported to have difficulty in marking sentence-final prosodic phrases with appropriate boundary tones to express "sentence-level" discourse information such as sentence types and the speakers' certainty about the contents (Choe, 2016; Lee, 2005; Lee, 2008; Park, et al., 2000). These results lead us to the last analyses which were to examine the relationship between the frequency of boundary tones and the position of IPs within a sentence. The position of prosodic phrases was divided into two groups: sentence-internal vs. sentence-final phrases. The position within a sentence for ips were coded but excluded from the analyses due to their insignificant group differences. Because of the inter-story differences in the numbers of IPs, the analyses used the percentage values calculated by dividing the number of IPs with a specific boundary tone (H-H%, H-L%, L-H%, or L-L%) by the total number of IPs in a specific sentence position (sentence-internal vs. sentence-final). Two-way ANOVAs were conducted to examine the effect of sentence position and group on the frequency of different boundary tones. The dependent variables were the percentages of boundary tones calculated for each story within the factors of interest (the speaker and the sentence position). Since the percentages of H-H% and L-L% violated the homoscedacity, Mann-Whitney *U* tests were conducted to compare the average percentages of the two groups (NS vs. KL) and the two sentence-positions (sentence-internal vs. sentence-final). The two-way ANOVA results revealed a significant

interaction between group and position for the percentage of L-H% tone [ $F(1, 136)=22.59, p<.001$ ], but not for the percentage of H-L% tone. However, the effects of group [ $F(1, 136)=7.01, p=.009$ ] and position [ $F(1, 136)=41.44, p<.001$ ] on the percentage of H-L% were significant. The results of Mann-Whitney *U* tests indicated significant group differences on the average H-H% percentage of both sentence-internal [ $U=956.0, p<.001$ ] and sentence-final [ $U=978.0, p<.001$ ] IPs and the average L-L% percentage of sentence-final IPs [ $U=300.5, p<.001$ ]. The averages and standard deviations of 4 boundary tones for 2 groups and 2 sentence positions were shown in Table 2.

**Table 2.** The mean and standard deviation of the percentage of boundary tones for the groups (NS vs. KL) and the sentence-positions (sentence-internal vs. sentence-final).

Sentence position	Boundary tones	Group	Mean	Standard deviation
Sentence-internal	H-H%*	NS	0.27	0.91
		KL	19.98	11.44
	H-L%	NS	9.43	7.16
		KL	6.72	6.88
	L-H%	NS	60.10	19.30
		KL	63.01	13.73
L-L%	NS	0.44	0.98	
	KL	0.39	1.42	
Sentence-final	H-H%*	NS	5.26	6.65
		KL	47.43	17.94
	H-L%*	NS	22.12	9.82
		KL	15.86	11.11
	L-H%*	NS	60.62	13.61
		KL	34.44	18.38
	L-L%*	NS	2.68	3.58
		KL	0.37	1.85

An asterisk(\*) after a boundary tone indicates a significant difference between the group.

The results argued that the most frequently used boundary tone for sentence-final IPs produced by the native speakers was L-H%, while that produced by the Korean learners was H-H%. Also, the results suggested that the native speakers used either H-L% or L-H% for most of the sentence-final IPs (around 83%). However, both groups marked around 60% of their sentence-internal IPs with L-H% tones. Another interesting finding from the last analyses was that the Korean learners used H-H% tones for about 20% of sentence-internal IPs. Altogether, the native speakers often used L-H% tones for more than half of the IPs regardless of phrase positions within a sentence, and they marked about 20% of their sentence-final IPs with H-L% tones. In contrast, the Korean learners used slightly different strategy to mark IPs depending on their sentence positions: the learners frequently used L-H% tones for sentence-internal IPs similar to the native speakers, but rather used H-H% tones for sentence-final IPs.

#### 4. Discussion

The purpose of the current study was to investigate the prosodic structures in spontaneous speech produced by the high-intermediate Korean learners of English. More specifically, the current study was designed to examine which boundary tones the Korean learners used to mark their intonational phrases and to figure out whether or not

the pattern of their using a specific boundary tone was different from the native speakers' pattern. By analyzing the prosodic structures of spontaneous speech by the Korean learners and the native speakers, the current study revealed three main findings as follows: (1) the Korean learners' speech was significantly shorter and contained significantly fewer sentences, but they more frequently put prosodic phrase boundaries (both ip and IP boundaries) in their spontaneously produced speech, (2) both of the Korean learners and the native speakers used High boundary tones (H%) the most to mark their IP boundaries, and (3) the most frequently used boundary tone for the native speakers was L-H% and that for the Korean learners was H-H%, and these patterns were consistent regardless of the IP-position within a sentence. In this section, the implication and the possible explanation about these findings will be discussed in terms of the acquisition of boundary tones in spontaneous speech by Korean learners of English.

The first finding that the Korean learners' frequent prosodic phrasing with a few sentences gives us an idea about the general characteristics of the learners' speech, which is not the main focus of the current study though. Specifically, the frequent prosodic phrasing pattern represented the typical characteristics of learners' speech since this result was consistent with the previous findings in Huang & Jun (2011) and Choe (2016), in that the learners put prosodic phrase boundaries more often than the native speakers when they asked to read sentences or passages. Another characteristic of learners' speech could be shown where the Korean learners' spontaneous speech was about 62% shorter with 39% fewer sentences compared to the native speakers' one on average. However, since the current study analyzed the durations of the entire stories including pauses, fillers, and other miscellaneous non-speech information, we cannot argue that the learners produced definitely shorter stories with fewer sentences. Therefore, to specify the overall characteristics of the learners' spontaneous speech, more precise analyses such as the number and the duration of silent pauses as well as speech rates should be provided. Before our moving on the next finding, it should be noted that the learners might produce more words than the current result considering the strict criteria of defining a "sentence" in the current study. As the current study counted the number of "sentences" only which satisfied both grammatical (i.e., a subject and a verb) and semantic (i.e., complete meaning) criteria, some of the learners' speech could not be counted as "sentences."

As for the use of boundary tones, the finding that the native speakers marked about 77% of the IPs with High boundary tones (3% for H-H% and 74% for L-H%) in their spontaneous speech showed the intonational characteristics of spontaneous speech. Specifically, let us think about the situation where a speaker needs to tell a story. In this situation, one piece of information that a speaker wants to convey is to inform the listener where his or her story ends. In other words, if the ongoing sentence is not the end of a story, a speaker should include the information that he or she has not finished telling the story yet. The frequent use of coordinate conjunctions such as "and" or "so" in the current data can be an evidence for the speakers' intention to connect the currently-speaking sentences with the following sentences. Considering this natural characteristic of spontaneously produced storytelling, it is possible to assume that a speaker would mark their prosodic phrases with High boundary tones to entail the tones' forward-looking function, which is to show the continuity and relatedness with the

upcoming prosodic phrase. Not only the overall frequency of High boundary tones, the frequency of H% sentence-final IPs exaggerates the distinctive characteristics of the spontaneous "story-telling" speech. Specifically, the comparison of the current result with the use of boundary tones in the read speech in Choe (2016) showed us that when the native speakers read a passage, they tended to mark less than 10% of their sentence-final IPs with H% (see Figure 4 in Choe (2016)). However, the native speakers in the current study marked about 66% of their sentence-final IPs with H% (5% for H-H% and 61% for L-H%). Altogether, the patterns of the native speakers' using High boundary tones in their spontaneous speech, consistent with the previous research (e.g., De Ruiter, 2015), proved that speakers tend to frequently mark their prosodic phrases with rising tones in spontaneous speech for their forward-looking function.

In addition to the native speakers' frequent use of High boundary tones, the Korean learners also marked about 84% of their IPs with High boundary tones (27% for H-H% and 57% for L-H%). Comparing this number with the native speakers' one (i.e., 77% of their IPs with H%) led us to argue that the Korean learners used similar strategy to show the continuity and relatedness in their spontaneously produced story telling. That is, the Korean learners seemed to understand the forward-looking function of H% and to be able to realize these tones to mark the IPs in their spontaneous speech. However, the direct comparison of the use of H% in between the Korean learners and the native speakers indicated that Korean learners more often marked their IP boundaries with H% tones than native speakers. The first language transfer might cause this overuse of H% by the Korean learners. Specifically, Seoul Korean is well known to have an Accentual Phrase, which is embedded under an IP (e.g., Jun, 1998, 2005). An Accentual Phrase in Seoul Korean is demarcated by either Low-High-Low-High or High-High-Low-High. As Seoul Korean speakers often mark their Accentual Phrases with High tones no matter which of the two tonal patterns is used, it is possible to argue that the Korean learners transfer their native language's prosodic structure to the spontaneous story telling in their second language. This possibility could be confirmed when the prosodic structures, especially the use of boundary tones, in spontaneously produced story telling in Korean are examined.

Another reason why the Korean learners more often used H% is because they often marked their IPs with High phrase accents (H-) followed by High boundary tones (H%) instead of Low phrase accent (L-), which was often used by the native speakers. More specifically, this H-H% tone was the most frequently used boundary tone for the IPs in sentence-final position (i.e., almost half of the sentence-final IPs). Then, what made the Korean learners to frequently mark their sentence-final IPs with H-H% tones? One possibility of the H-H% overuse could be due to the learners' lack of confidence about their language performance. In the literature, the H-H% was known to mark interrogative sentences (e.g., Ladd, 1996; Ohala, 1983). As second language learners, spontaneously telling a story in their target language could be just a big burden. While conducting this experiment, the participants often wanted to get confirmation from the experimenter about whether they were using the right vocabulary and making grammatically correct sentences. Although the experimenter could not verbally interact with the participants, she sometimes gave a nod and a glance for the participants to move on. Considering this situation, the learners'

intention to get confirmation about their language performance could be expressed in somewhat “question-like” boundary tones due to unsureness.

Another possibility could be because of the learners’ intention of expressing continuity. Previous research suggested that H-H% tone in declarative sentences indicated the meaning of a strong continuation to the upcoming phrase in Australian English (Fletcher, et al., 2002). Considering Fletcher and colleagues’ argument, the possible explanation is because the learners intended to strongly inform the experimenter of their continuation to the upcoming phrases. In other words, when the learner thought she needed to complete a syntactic sentence, but to continue the story, the learner might use her own strategy to differentiate the boundary within a syntactic sentence (which the listener might be easily noticed as one sentence) from the boundary at the end of a syntactic sentence while keeping the forward-looking function.

## 5. Conclusion

In sum, the current study investigated the prosodic structures, especially the use of boundary tones, of high-intermediate Korean learners’ spontaneous speech. By comparing the learners’ patterns with the native speakers’ ones, the current study examined how far the Korean learners acquired the function and realization of boundary tones. The findings from the current study proposed that high-intermediate Korean learners successfully understood the forward-looking function of High boundary tones, and then were able to appropriately make use of them. However, the fact that these learners used steep rising tones to mark sentence-internal IPs more often than the native speakers, who often marked their sentence-internal IPs with more gradual rising tones, suggested that the learners only understood the general meaning and function of rising tones at the end of prosodic phrases, but not fully acquired how to fine-tune the degree of steepness of the rising tone.

## References

- Anderson-Hsieh, J., Johnson, R., & Koehler, K., (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42(4), 529-555.
- Beckman, M. E., & Elam, G. A. (1997). *Guidelines for ToBI labelling* (Version 3.0). Columbus, OH: The Ohio State University Research Foundation.
- Beckman, M. E., & Hirschberg, J. (1994). *The ToBI annotation conventions*. Columbus, OH: Ohio State University.
- Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The Original ToBI System and the evolution of the ToBI framework. In S. A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 99-116). Oxford, UK: Oxford University Press.
- Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 5-70.
- Bhat, S., & Yoon, S. Y. (2015). Automatic assessment of syntactic complexity for spontaneous speech scoring. *Speech Communication*, 67, 42-57.
- Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer (Version 5.3.66) [Computer Program]. Retrieved from <http://www.praat.org/>
- Candea, M., Vasilescu, I., & Adda-Decker, M. (2005, September). Inter- and intra-language acoustic analysis of autonomous fillers. *Proceedings of Disfluency in Spontaneous Speech Workshop* (pp. 47-52).
- Choe, W. K. (2016). Acquisition of prosodic phrasing and edge tones by Korean learners of English. *Phonetics and Speech Sciences*, 8(4), 31-38.
- Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics*, 39(2), 237-245.
- Couper-Kuhlen, E. (2001). Intonation and discourse: Current views from within. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 13-34). Malden, MA: Blackwell Publishers.
- De Ruiter, L. E. (2015). Information status marking in spontaneous vs. read speech in story-telling tasks- Evidence from intonation analysis using GToBI. *Journal of Phonetics*, 48, 29-44.
- Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 45-89). Hillsdale, NJ: Lawrence Erlbaum.
- Face, T. L. (2003). Intonation in Spanish declaratives: differences between lab speech and spontaneous speech. *Catalan Journal of Linguistics*, 2, 115-131.
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied linguistics*, 21(3), 354-375.
- Fletcher, J., Stirling, L., Mushin, I., & Wales, R. (2002). Intonational rises and dialog acts in the Australian English map task. *Language and Speech*, 45(3), 229-252.
- Hewings, M. (1995). Tone choice in the English intonation of non-native speakers. *International Review of Applied Linguistics*, 33(3), 251-303.
- Holden, K. T., & Hogan, J. T. (1993). The emotive impact of foreign intonation: An experiment in switching English and Russian intonation. *Language and Speech*, 36(1), 67-88.
- Huang, B. H., & Jun, S. A. (2011). The effect of age on the acquisition of second language prosody. *Language and Speech*, 54(3), 387-414.
- Jun, S. A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, 15(2), 189-226.
- Jun, S. A. (2005). Korean intonational phonology and prosodic transcription. In S. A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 430-458). Oxford, UK: Oxford University Press, UK.
- Kim, J. E. (2019). Acoustic characteristics of read and spontaneous speech in Seoul Korean with between-age variability. *Korean Linguistics*, 85, 61-76.
- Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, 35, 162-179.
- Ladd, R. (1996). *Intonational phonology*. Cambridge, UK: Cambridge University Press.
- Lee, J. K. (2005). Korean speaker’s edge tone patterns of English conjunctive utterances. *Speech Sciences*, 12, 141-152.
- Lee, J. K. (2008). An intonational assessment of English communicative competence of Korean college students with the 7th curriculum. *Modern English Education*, 9(1), 256-282.
- Lieberman, P., Katz, W., Jongman, A., Zimmerman, R., & Miller, M.

- (1985). Measures of the sentence intonation of read and spontaneous speech in American English. *The Journal of the Acoustical Society of America*, 77(2), 649-657.
- Munro, M. J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition*, 17, 17-34.
- Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49(1), 285-310.
- Ohala, J. J. (1983). Cross-language use of pitch: an ethological view. *Phonetica*, 40(1), 1-18.
- Park, S. B., Skrypiczajko, G., & Kim, K. H. (2000). A study on the realization of intonational tunes depending on the difference of meaning in English: In comparison of English native speakers with Korean speakers. *Speech Sciences*, 7(2), 97-112.
- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 271-311). Cambridge, MA: MIT Press.
- Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
- Shokier, V. (2008). Evidence for the stable use of uptalk in South Ontario English. *Working Papers in UPenn Linguistics*, 14(2), 16-21.
- Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, 32(1), 25-38.
- Tomlinson Jr, J. M., & Tree, J. E. F. (2011). Listeners' comprehension of uptalk in spontaneous speech. *Cognition*, 119(1), 58-69.
- Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, 28, 1-30.
- Veilleux, N., Shattuck-Hufnagel, S., & Brugos, A. (2006). *The ToBI tutorial*. Cambridge, MA: MIT OpenCourseWare. Online: <http://anita.simmons.edu/-tobi/index.htm>
- Veilleux, N., Shattuck-Hufnagel, S., & Brugos, A. (2006). 6.911 Transcribing prosodic structure of spoken utterances with ToBI. from MIT OpenCourseWare. Retrieved from <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/>

- **Wook Kyung Choe**, Corresponding author  
Assistant Professor, College of General Education  
Dong-A University,  
37 Nakdong-Daero 550 beon-gil, Saha-gu  
Busan 49315, Korea  
Tel: 051-200-1022  
Email: [reginal004@dau.ac.kr](mailto:reginal004@dau.ac.kr)  
Fields of interest: Phonetics, Phonology