Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Prosodic accommodation in Seoul Korean Accentual Phrases

Jiseung Kim

University of Michigan
jiseungk@umich.edu

Abstract

The goal of this study is to examine prosodic accommodation, specifically to test accommodation of prosodic boundaries with native speakers of Seoul Korean. Sixteen native speakers of Seoul Korean participated in a sentence completion task where they were asked to complete a target sentence after reading (in the baseline condition) or listening to (in the test condition) a context sentence. In both cases, participants completed the target sentence by speaking. The auditory context sentences had artificially manipulated prosody. The manipulation lowered the f0 of the phrase-final syllables that were associated with the Accentual Phrase (AP)-final rise, which is a characteristic intonational property of Seoul Korean. Four f0 values – f0 maximum, minimum, mean, and range – were extracted from the AP-final syllables of the participants' responses, and were compared between the baseline and test conditions. Preliminary results of the four speakers analyzed to date show evidence of convergence for three speakers. The results suggest that effects of accommodation may be manifested in the pitch contours associated with a prosodic boundary.

Index Terms: prosodic boundaries, accommodation, Seoul Korean

1. Introduction

Speech accommodation refers to a situation in which the speech patterns of interacting speakers become more similar, or converge, over the course of a conversation. In part because accommodation has been argued to be a vehicle for dialect acquisition and sound change ([1], [2], [3]), it has received extensive investigation in recent years at various linguistic levels, including syntactic structure ([4], [5], [6]), lexical items ([7], [8], [9]), and phonetic features ([10], [11]). The goal of this study is to investigate accommodation at the prosodic level, specifically prosodic boundaries, in order to gain a better understanding of the ways in which prosodic properties might be similarly malleable.

The results of the few previous studies that have explored prosodic accommodation have been mixed. D'Imperio et al. [12] investigated patterns of accommodation in terms of the alignment of pitch accents across two dialects of Italian. They found that the speakers of Neapolitan Italian accommodated to the speakers of Bari Italian even though those adjustments would lead to masking of the different alignment patterns of the rising pitch accents in their native dialect. Ní Chiosáin [13] also looked at accommodation in terms of prominence – specifically, the phonetic correlates for the lexical stress in two varieties of Irish (Gaelic), although the study did not find statistically significant evidence of accommodation. These two studies examined phonological or phonetic cues to prominence, rather than boundary information. Cole & Shattuck-Hufnagel [14] tested both prominence- and boundary-related information in an imitation study. Their preliminary results suggested that the phonological structures of model speech can be reproduced by the subjects, given that the pitch accents and intonational phrase boundaries that appeared in the model speech were mostly retained in the imitation speech. However, speakers were less accurate in imitating the phonetic details – i.e., pause duration and irregular pitch periods – of the model speech.

The present study examines whether speakers would demonstrate effects of accommodation to the f0 contours associated with a prosodic boundary, specifically the Accentual Phrase (AP) in Korean. The AP in Seoul Korean is an intonationally defined prosodic unit that contains one or more words and is differentiated from Intonational Phrase by the lack of pause or final lengthening ([15], [16]). It is marked by the underlying tonal pattern of THLH where T is a H tone when the initial segment of an AP is an aspirated or tense consonant and a L tone otherwise ([15], [16]). A corpus study on Seoul Korean found that 84% of all APs in the corpus had a final H tone, and the AP-final tone was actively used in speech processing ([17]). It can thus be concluded that the rising (LH) intonation that co-occurs with the phrase-final syllable is a characteristic feature of Seoul Korean AP.

The current study synthetically manipulated the AP boundary and investigates whether how speakers of Seoul Korean would accommodate to this artificially-created prosody. The experiment may provide insight into the process of accommodation involving an artificially manipulated prosody in a language outside the Indo-European family.

The motivation for using artificial manipulation instead of speech from a naturally occurring dialect is that the Seoul dialect is widely recognized as the standard variety in Korean ([18]). With other regional dialects of Korean being heavily stigmatized, it is unlikely that speakers of Seoul Korean would identify themselves with a dialect other than Seoul. Considering that group membership may influence the direction of speech accommodation, as noted in previous studies ([19], [20]), testing accommodation between the Seoul dialect and a different dialect would not be left without a bias. The artificial manipulation employed in this study only affected prosody; all other – segmental, lexical, and syntactic – aspects of the original Seoul Korean speech remained untouched, so that participants would not readily identify the stimuli as a particular dialect of Korean. In addition, using artificial manipulation allowed for appropriate control over unwanted confounding effects that could be introduced in a comparison of two regional dialects.

The major hypothesis of this study is that prosodic convergence will take place, in that speakers of Seoul Korean will accommodate to the novel prosody of the model speaker.

 doi: 10.21437/SpeechProsody.2016-81

Under this hypothesis, the participants' post-exposure production of the target sentences would show the features of the artificial prosody that are not present in their native dialect.

2. Method

Sixteen speakers of Seoul Korean participated in the study. All participants were female speakers in their 20s or early 30s who had not spent more than one year outside Seoul. They received monetary compensation for participating in the study. All experiments were conducted in a sound-proof studio located in downtown Seoul.

The experiment was a sentence completion task. There were two blocks in each of the two conditions. Each condition had Block 1 and 2, and the sentences were presented in differently randomized orders. The difference between Block 1 and Block 2 were themes, although they expressed similar types of comparison between number values. For example, the sentences in Block 1 were about different ages of siblings, and the sentences in Block 2 were about different times of arrival (see Table 1).

Block 1: different ages of siblings
Context sentence "*Ahn Mina is four years older than Ahn Yuri."* anmina-nun anyuri-pota nesalina naiga-manta Ahn Mina-Subj Ahn Yuri-than four-year age-more
Target sentence *"Therefore Ahn Mina is (Ahn Yuri's older sister)."* kuromuro anmina-nun (anyuri-pota unnita) Therefore Ahn Mina-Subj (Ahn Yuri-than oldersister)
Block 2: different times of arrival
Context sentence *"Ahn Mina had come to the Ahnam station at eight and Ahn Yuri was coming to the Ahn station at nine."* anmina-nun yeotolshi-e anamyeok-e wa-isso-ko Ahn Mina-Subj eight-Top Ahnam st.-Top come-Past-Conj anyuri-nun ahopshi-e anamyeok-e oko-isso-ta. Ahn Yuri-Subj nine-Top Ahnam st.-Top come-Past
Target sentence *"Therefore Ahn Mina (had come to Ahnam Station earlier than Ahn Yuri)."* kuromuro anmina-nun (anyuri-pota Therefore Ahn Mina-Subj (Ahn Yuri-than anamyeok-e moncho wat-ta) Ahnam st.-Top earlier come-Past

Table 1. *Example of two pairs of sentences from the experiment blocks 1 and 2.*

The first two blocks of experiment were baseline conditions for all participants. The baseline condition was designed to establish the speakers' reference productions of the accentual phrase of Seoul Korean (i.e., their natural productions before any accommodation might take place). The speakers silently read a short sentence that provided the context for a following incomplete target sentence (e.g., the context sentence in Table 1). The participants were then asked to complete a related target sentence by saying the full completed sentence out loud (e.g., the target sentence in Table 1; the parenthetical portion of the sample sentence was not provided for participants).

The last two blocks were test conditions that followed a short break. In the test conditions, participants listened to the context sentence and the first two words of the target sentence over headphones. In this sentence completion task, they repeated the first two words and then produced the remainder of the target sentence. Before each of the four blocks, participants were given five pairs of context and completed target sentences as a practice. These practice sentences were shown on the computer monitor only. Unlike the test sentences, there was no accompanying audio recording. After the experiment, all participants completed an exit survey that asked for their language background and provided ratings on the model speaker.

For the auditory stimuli used in the test condition, a Seoul speaker was recorded producing the context and incomplete target sentences. The manipulation lowered the f0 peak in the AP-final rise (LH) by 20% (about 40-50 Hz) with appropriate interpolation that affected AP-final syllables for an interval of about 200ms around the f0 peak. Prior to the experiment, a series of pilot investigations involving naïve Korean listeners as well as phoneticians were completed to establish the appropriate degree of manipulation. The f0 lowering was done at every IP-medial AP boundary using Praat's pitch extraction algorithm and a Praat script; a token-by-token inspection was also conducted to ensure natural-sounding speech. Because the auditory stimuli were based on the recordings of the author, another Seoul Korean speaker conducted all experiments and interacted with the participants.

Figure 1 shows a sample context sentence (without the first two words of the target sentence). The f0 contours of the IP-medial AP-final syllables ('nun', 'ta', 'na') before and after the manipulation are are affected by the manipulation. This sentence can be translated as *Ahn Mina is four years older than Ahn Yuri*.

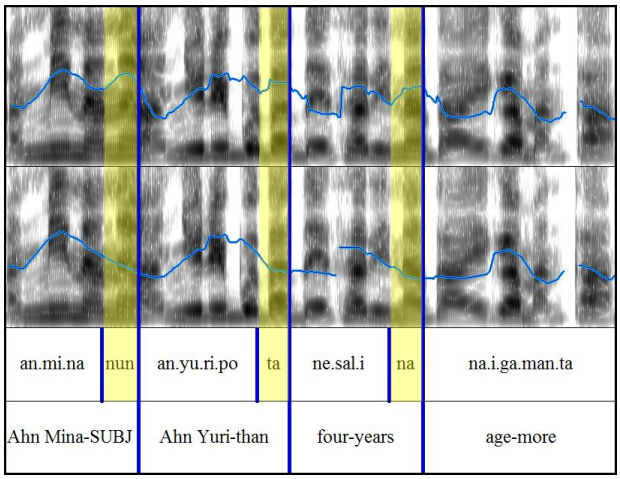

Figure 1. *f0 tracks of a sample context sentence before (top) and after (bottom) the artificial manipulation.*

Each participant produced 96 target sentences in the baseline condition and 96 target sentences in the test condition. There were 28 filler tokens in each condition, leaving a total of 164 sentences to be analyzed for each participant. The target utterance consisted of four APs (Block 1) or five APs (Block 2), and thus each utterance contained three or four IP-

medial AP boundaries. Excluding the cases where the speaker produced an IP boundary instead of an AP boundary as a mistake, an average of 467 instances of IP-medial AP boundaries was analyzed per participant.

For the acoustic analysis, four f0 measures – f0 maximum, minimum, mean, and range – were taken in the final syllable of all IP-medial APs from the baseline and test productions. Comparison of the baseline and test measures determines whether participants accommodated to the manipulated stimuli by producing lower f0 at AP boundaries after exposure to the artificial prosody. For the statistical analysis, a set of one-way Analyses of Variance (ANOVA) was performed for each participant to test whether the difference between the f0 measures in the baseline and test conditions is statistically significant.

3. Results

Analyses have been completed for four of the 16 participants; those data are presented here. For each participant, the difference in the four f0 measures between the baseline and test conditions was statistically tested. If the f0 levels in AP-final syllables from the test condition were significantly lower than the f0 levels from the baseline condition, it can be interpreted as the effect of artificial prosody.

For S8, the f0 maximum, mean, and range were significantly lower in the test production, compared to the baseline production ($F(1,480) = 67.73$, $p < .001$; $F(1,480) = 45.51$, $p < .001$; $F(1,480) = 47.69$, $p < .001$, respectively). The f0 minimum was marginally significant ($F(1,480) = 3.62$, $p = .058$). For S13, all four f0 measures were significantly lower in the test condition than in the baseline condition ($F(1,480) = 328.90$, $p < .001$; $F(1,480) = 142.10$, $p < .001$; $F(1,480) = 317.40$, $p < .001$; $F(1,480) = 7.00$, $p < .01$; for f0 maximum, minimum, mean, and range, respectively). However, S15's measures did not show any significant difference between the baseline and the test productions ($F(1,442) = 0.16$, $p > .10$; $F(1,442) = 0.69$, $p > .10$; $F(1,442) = 0.004$, $p > .10$; $F(1,442) = 1.60$, $p > .10$). S22 showed significant decrease in the f0 maximum and mean from the baseline to the test condition ($F(1,466) = 10.39$, $p < .01$; $F(1,466) = 10.1$, $p < .01$), but no significant effect for f0 minimum and range ($F(1,466) = 2.77$, $p > .1.0$; $F(1,466) = 3.53$, $p = .061$).

Figure 2 illustrates the lowering of f0 maximum between the baseline and test conditions for S8, S13, S22 but not S15. Figures from 3 through 5 show the comparisons between the baseline and test conditions for f0 minimum, mean, and range, respectively.

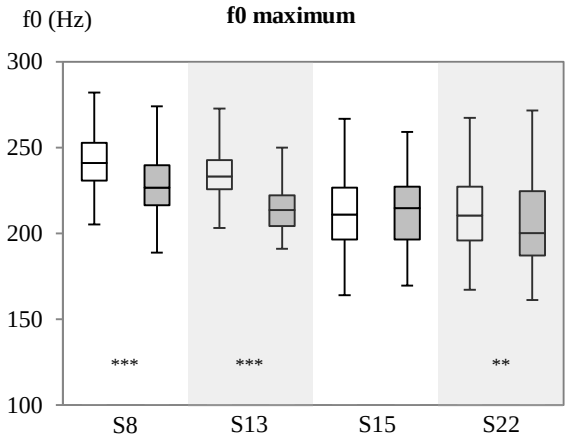

Figure 2. *f0 maximum difference in the baseline (white boxes) and test (grey boxes) conditions for four participants. (***p < .001, **p < .01)*

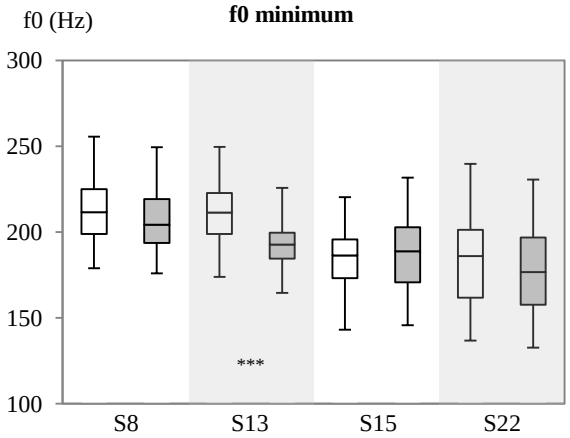

Figure 3. *f0 min difference in the baseline (white boxes) and test (grey boxes) conditions for four participants. (***p < .001)*

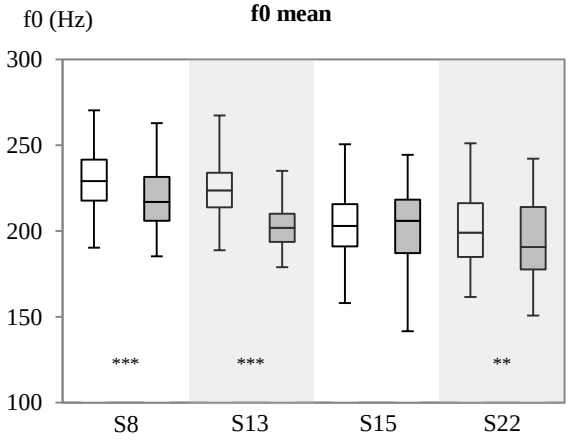

Figure 4. *f0 mean difference in the baseline (white boxes) and test (grey boxes) conditions for four participants. (***p < .001, **p < .01)*

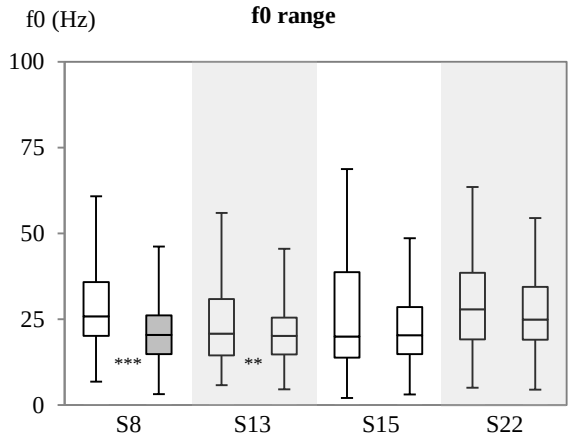

Figure 5. *f0 range difference in the baseline (white boxes) and test (grey boxes) conditions for four participants. (***p < .001, **p < .01)*

Significant f0 lowering in the test condition was found in f0 maximum and f0 mean for three participants (S8, S13, S22), and in f0 range for two participants (S8, S13), and in f0 minimum for one participant (S13). This is consistent with the AP-final rise (LH) in the test condition becoming more similar to the artificially lowered AP-final rises of the model speaker.

Figure 6 illustrates an example of a sample target sentence produced by S13, who showed significant effects of convergence. The AP-final syllables 'ro', 'nun', and 'ta' were examined in the acoustic analysis. This sentence can be translated as *Therefore Ahn Mina is Ahn Yuri's older sister*.

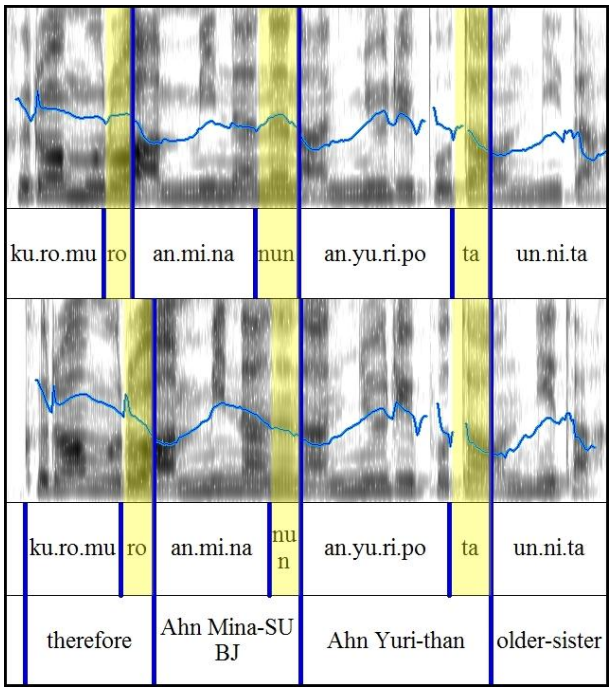

Figure 6. *f0 tracks of a target sentence produced by S13, in the baseline condition (top) and the test condition (bottom).*

Note that the difference in the pitch levels of AP-final syllables ('ro', 'nun', 'ta') between the baseline and the test conditions. The trend of convergence was not across-the-board, however. S15 did not show a significant difference in any of the four f0 measures.

4. Discussion

The hypothesis tested in this study was that the native speakers of Seoul Korean would converge to the Seoul Korean stimuli that had undergone the artificial manipulation on the pitch levels of AP-final syllables. The manipulation lowered the f0 by 20% at every AP-final syllable, rendering the sentences lacking the AP-final rise that is a characteristic feature of Seoul Korean prosody. It was predicted under the hypothesis that the f0 values extracted from participants' AP-final syllables would be lower after they were exposed to the manipulated prosody.

For three out of four participants whose data have been analyzed to date, there was statistically significant lowering of the f0 values in the test productions compared to the baseline productions. That is, those three participants' pitch levels associated with the AP-final syllables have been modified in the same way that the model speaker's had been manipulated. The overall results can be interpreted as support for convergence at the prosodic level. This outcome is in line with previous research on prosodic convergence ([12], [21], [22]), and demonstrates that the pitch levels associated with a prosodic boundary may be modified in the process of convergence.

There could be several reasons for lack of convergence in the production of S15. It could be that this speaker detected the atypical prosody but nonetheless retained her prosody throughout the experiment. There might be a sociolinguistically motivated reason for this maintenance. In this case, relating the f0 analysis to the analysis of the exit survey could shed light on this issue. It is also possible that the manipulated prosody was not perceptually salient for this speaker, or that this speaker was converging on a different correlate that has not yet been analyzed.

The next steps for this study will be to (1) analyze the data from the remaining speakers, (2) include additional measures in the acoustic analysis, such as the temporal alignment between the AP-final rise and the AP-final syllable, and (3) explore the exit survey results in relation to the results of the acoustic analysis.

5. Conclusion

The goal of this study was to examine accommodation of prosodic boundaries. The major finding is that three out of four participants showed significant effects of accommodation, by adjusting their f0 when they were exposed to the model speaker whose f0 was artificially lowered. The results mostly support the hypothesis that the speakers of Seoul Korean would converge to the artificially manipulated prosody. In line with previous studies that reported cases of prosodic convergence, the results of this study demonstrated that the effects of accommodation may be observed in the phonetic details relevant to a prosodic boundary.

6. References

[1] P. Trudgill, "Linguistic accommodation: Sociolinguistic observations on a sociopsychological theory," *In Papers from the Parasession on Language and Behavior*, ed. Roberta Hendrick, Carrie Masek, and Mary Frances Miller, Chicago: Chicago Linguistics Society, pp.218-237, 1981.

[2] P. Trudgill, *Dialects in contact*, Blackwell Publishing, 1986.

[3] M.E. Babel, M. McAuliffe, and G. Haber, "Can mergers-in-progress be unmerged in speech accommodation?", *Front Psychol*. 24,4:653, 2013.

[4] W.J.M. Levelt, and S. Kelter, "Surface form and memory in question answering", *Cognitive Psychology*, 14, pp.78-106, 1982.

[5] M.J. Pickering, and V. Ferreira, "Structural priming: a critical review", *Psychological bulletin*, 134(3), pp.427-459, 2008.

[6] H. Branigan, M.J. Pickering, J. Pearson, and J. McLean, "Linguistic alignment between people and computers", *Journal of Pragmatics*, 42(9), pp.2355-2368, 2010.

[7] S. Brennan, "Lexical entrainment in spontaneous dialog", *Proceedings of ISSD*, pp.41-44, 1996.

[8] S. Garrod, and A. Anderson. "Saying what you mean in dialog: A study in conceptual and semantic co-ordination", *Cognition*, 27, pp.181-218, 1987.

[9] M.J. Pickering, and S. Garrod. "Toward a mechanistic psychology of dialogue", *Behavioral and Brain Sciences*, 27, pp.169-225, 2004.

[10] J.S. Pardo, "On phonetic convergence during conversational interaction", *The Journal of the Acoustical Society of America*, 119(4), 2382-2393, 2006.

[11] M.E. Babel, "Evidence for phonetic and social selectivity in spontaneous phonetic imitation", *Journal of Phonetics*, 40(1), pp.177-189, 2012.

[12] M. D'Imperio, R. Cavone, and C. Petrone, "Phonetic and phonological imitation of intonation in two varieties of Italian", *Front. Psychol*. 5:1226, 2014.

[13] M. Ní Chiosáin, "Effects of synchronous speech task on length and prosody in interdialectal nonprestige varieties", *Language Variation and Change*, pp.51-62, 2007.

[14] J. Cole, and S. Shattuck-Hufnagel, "The phonology and phonetics of perceived prosody: What do listeners imitate?", *Proceedings of Interspeech 2011*, pp.969-972, 2011.

[15] S-A. Jun, *The Phonetics and Phonology of Korean Prosody*, Ph.D. dissertation, The Ohio State University, Columbus, Ohio, 1993.

[16] S-A. Jun, *The Accentual Phrase in the Korean prosodic hierarchy*, Phonology, 15.2, pp.189-226, 1998.

[17] S-H, Kim. *The role of prosodic phrasing in Korean Word Segmentation*, PhD Dissertation, UCLA, 2004.

[18] K-M, Lee, and S.R. Ramsey, *A History of the Korean Language*, Cambridge University Press, 2011.

[19] R.Y. Bourhis, "Language in ethnic interaction: a social psychological approach", In H. Giles & B. Saint-Jacques (eds.), *Language and Ethnic Relations*, Oxford: Pergamon Press, pp.117-142, 1979.

[20] H. Giles, R.Y. Bourhis, and D.M. Taylor, "Towards a theory of language in ethnic group relations", In H. Giles (ed.), *Language, Ethnicity and Intergroup Relations*, London: Academic Press, pp.307-348., 1977.

[21] F. Bonin, C. De Looze, S. Ghosh, E. Gilmartin, C. Vogel, A. Polychroniou, H. Salamin, A. Vinciarelli, and N. Campbell, "Investigating fine temporal dynamics of prosodic and lexical accommodation", *Proceedings of Interspeech 2013*, pp.539-543, 2013.

[22] C. De Looze, S. Scherer, B. Vaughan, and N. Campbell, "Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction", *Speech Communication*, 58, pp.11-34, 2014.

[23] B. Vaughan, "Prosodic synchrony in co-operative task-based dialogues: a measure of agreement and disagreement", *Proceedings of Interspeech 2011*, pp.1865-1868, 2011.

[24] J. Jun, J. Kim, H. Lee, and S-A. Jun, "The prosodic structure and pitch accent of Northern Kyungsang Korean", *J East Asian Linguis*t, 15, pp.289-317, 2006.

[25] D. Lee, and S. Davis, "On the Pitch-Accent System of South Kyungsang Korean: A Phonological Perspective", *Language Research*, 45, pp.3-22, 2009.