Determining prominence and prosodic boundaries in Korean by non-expert rapid prosody transcription

Hie-Jung You

Department of Linguistics
University of Illinois, Urbana-Champaign
You11@uiuc.edu

Abstract

This paper examines how non-expert listeners perceive prominence and prosodic boundaries in Korean using the Rapid Prosody Transcription (RPT) method, developed by Mo, Cole and Lee [9] for American English. While prominence is used to mark prosodically salient or "highlighted" words and phrases, prosodic boundaries demarcate units or "chunks" of speech to mirror the hierarchical relations among prosodic structures. Confirming the findings of earlier studies on American English, non-expert transcribers of Korean show agreement rates that are well above chance and, show higher agreement rates for prosodic boundaries than prominence. Korean listeners not only perceive prosodic boundaries at sentence-level boundaries corresponding to Intonation Phrase (IP) boundaries in the Korean Tones and Break Indices (ToBI) system, but also at clausal-level boundaries. The findings suggest that cues to boundaries are more salient than cues to prominences in Korean. For the perception of prominence, Korean listeners seem to orient to different cues including prosodic, syntactic and lexical information.

Index Terms: speech perception, prosody, Korean prosodic phonology, Rapid Prosody Transcription

1. Introduction

Spoken corpora have become invaluable resources for investigating real-time language processing mechanisms or the range of pragmatic or discourse meaning conveyed through speech. One aspect of spontaneous speech that is distinct from read or elicited speech is prosody. Prosody is defined as the intonational and rhythmic patterns of speech, and determines the prominence relations between words and the grouping of words within an utterance into prosodic phrases. Prominence is defined as "a word that is highlighted for the listener" [11] and prosodic boundaries represent units or "chunks" of speech that serve to mirror the hierarchical relations among prosodic phrase structures [2].

Languages differ in their prosodic structures, their location of pitch accents and the way listeners structure and interpret utterances [1, 5]. In American English, "every word has stress, [but] not every word receives pitch accent" [6]. Pitch accents refer to pragmatically and semantically prominent words in an utterance that contain a stressed syllable in American English [10]. However, not all languages are like American English. A language with a typologically distinct prosodic system is Korean. To summarize some differences, new information is assigned stress prominence at the phrase level. A "new" word is initial in its Accentual Phrase (AP), and stress is realized on the AP-initial syllable [4]. Thus, stress in Korean is viewed to be phrasal and co-occurs with prominence given that f0 peaks are not associated to specific "stressed" syllables, but associated with a specific "location" within a phrase [5]. Second, boundaries and pitch-accented syllables in Korean are interrelated, which leads to predictions about the locations of prominences relative to boundaries. Third, while the pragmatic meaning of a sentence is delivered by the whole intonation contour in English, the pragmatic meaning of a sentence is delivered by the Intonation Phrase (IP) boundary tone realized on the phrase final syllable in Korean.

Prior work on prosody perception includes (1) studies of brain responses in the perception of prosody [8], and, (2) studies of (indirect) behavioral responses such as prosodic transcriptions. In the context of phonological analysis, prosodic transcription is generally performed by transcribers who are trained in phonology or phonetics. In their study on prosody perception with spontaneous speech in American English, Mo, Cole and Lee [9] introduced a method of transcription called Rapid (Naïve) Prosody Transcription with 70+ untrained listeners of American English who were asked to identify prosodic boundaries and prominent words in a real-time listening task. In line with the findings of earlier studies [11, 12], their study provided consistent agreement rates, with higher agreement rates for prosodic boundary perception than prominence.

The present study replicates Mo et al.'s [9] study using RPT for the following reasons. RPT enables the investigation of prosody perception with larger speech segments drawn from uncontrolled, conversational speech. And since transcribers are marking perceived prominence and boundaries, the resulting transcriptions can be fruitfully compared to a Tones and Break Indices (ToBI) based transcription of the same elements, to evaluate listeners' perception in relation to the phonological prosodic features of the language [5]. This study predicts that (1) ordinary Korean listeners will perceive prosodic boundaries in conversational speech, corresponding to IP boundaries in a Korean ToBI (K-ToBI) analysis, (2) ordinary Korean listeners will perceive some words as more prominent than other words in a given segment from conversational speech, corresponding to the existence of phrasal stress in Korean, (3) agreement rates between listeners' transcriptions will be high, comparable to what has been found for prosody perception in English and, (4) agreement rates for prosodic boundaries will be higher than those for prominence, due to (i) the presence of a more robust set of cues to boundaries (including cues to boundary tones) compared to stress prominence, which lacks tonal cues; and (ii) in accordance with the higher boundary agreement rates found for English.

2. Methodology and analysis

2.1. Materials

To collect spontaneous Korean speech, two native Korean speakers (S1 and S2) were recruited to answer a list of questions. The interviews were recorded and lasted a total of seven minutes each. Four sound files were extracted from the

recordings. Each sound file is about 30 seconds long and they were taken from different parts of the interview. The sound files have been transcribed and a printed transcript was given to each subject. The transcription did not contain any punctuation marks and words have been separated by spaces. Speech errors and disfluencies have been noted and transcribed.

2.2. Procedure

A total number of twenty listeners have been divided into two groups, with each group consisting of ten transcribers. Both groups received the same four transcripts of the four audio files. Group 1 was asked to annotate prosodic boundaries in the first two excerpts and prominence for the last two transcripts. Group 2, on the other hand, was instructed to annotate prominence in the first two transcripts and prosodic boundaries in the last two transcripts. Each subject completed the task while listening to the recording. To mark prominence, listeners were asked to underline words that they perceived to be highlighted in relation to surrounding words. To mark prosodic boundaries, listeners were instructed to insert a vertical line between words where they perceived a boundary between chunks. A fragment of a transcript annotated for prominence (example 1a) and prosodic boundaries (example 1b) is given below.

(1)
a. *e nay-ka manyakey taythonglyengila-myun*
uh I-nom if president-be-cond
cangayin-tul wihan pep-ul com
disabled-pl for law-acc little
ilehkey mantulko sipheyo
like.this make-comp want-hon

'Uh if I were the president I want to pass a law for people with disabilities like this'

b. *e nay-ka manyakey taythonglyengila-myun* |
uh I-nom if president-be-cond
cangayin-tul wihan pep-ul com
disabled-pl for law-acc little
ilehkey mantulko sipheyo
like.this make-comp want-hon

'Uh if I were the president I want to pass a law for people with disabilities like this'

In example 1a, "people with disabilities" and "law" were marked as prominent; and, a prosodic boundary was inserted between the conditional clause "uh if I were the president" and the main clause "I want to pass a law for people with disabilities like this" in 1b.

2.3. Analyses

To analyze listeners' transcriptions, a reliability test for prominence and boundary labels is performed for all annotations pooled over speakers and transcribers. The findings are compared with the findings for American English [9]. The second analysis examines one of the four excerpts, excerpt 1, more closely. The third analysis looks at speaker variation to compare differences in agreement rates across speakers (S1 and S2).

3. Results

To gather a greater variety of annotations and to examine the same excerpt for both boundaries and prominence, all transcribers were given the same four excerpts although with a different distribution of tasks in the two groups.

Table 1. *Distribution of excerpts and tasks*

	Ex. 1-S1	**Ex. 2-S2**	**Ex. 3-S2**	**Ex. 4-S1**
Group 1	Boundary	Prominence	Boundary	Prominence
Group 2	Prominence	Boundary	Prominence	Boundary

Note that excerpts 1 and 4 were taken from the speech material produced by speaker S1 and excerpts 2 and 3 were produced by speaker S2.

3.1. Reliability test

Fleiss' kappa coefficient was used to test the reliability of the method of RPT [3]. Following Mo et al. [9], I chose Fleiss' kappa statistic because it gives a single coefficient to measure agreement rates between multiple transcribers. To interpret kappa values, Landis and Koch [7] have come up with a statistic interpretation suggesting that kappa scores < 0 reflect poor agreement, 0.01-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement and 0.81-1.00 almost perfect agreement. The kappa scores for boundaries and prominence are given in figure 1.

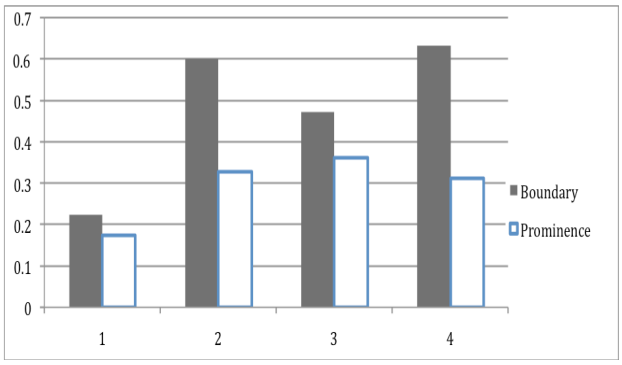

Figure 1: *Fleiss' kappa scores for boundaries and prominence*

For excerpts 2 and 4, values for boundary are almost twice as high compared to the values achieved for prominence. While kappa scores for excerpt 3 are a little lower and much lower for excerpt 1, transcribers' agreement rates are considerably higher for boundaries than for prominence in all four excerpts.

The values for prominence reflect similarly high inter-labeller agreement rates for prominence in excerpts 2, 3 and 4. The kappa score for excerpt 1 is comparatively lower. The boundary kappa scores for excerpt 4 are almost three times higher than those for excerpt 1. It is noticeable that kappa scores may vary to a great extent even within the same speaker. Excerpt 2 and 3 also show a greater variation within a single speaker compared to the results for prominence, which are very similar for S2. Excerpt 1 shows lower agreement rates for the annotation of both prominences and boundaries. This

raises the question why the results for excerpt 1 are so much lower compared to the other three excerpts.

3.2. Comparison of findings between American English and Korean

Table 2 compares the findings of the study on American English by Mo et. al [9] with the findings of the present study on Korean. The present study delivers a total of eight kappa (K) scores, four boundary and four prominence scores. The four scores for boundary vary to greater degree, showing a broader range (0.22-0.63) than the scores for prominence (0.17-0.36). Based on the considerably lower agreements for excerpt 1, I have created two columns for comparison, one including all excerpts and another excluding the scores for excerpt 1.

Table 2. *Comparison of Fleiss' Kappa scores in American English and Korean (*Note that M = average kappa score)*

	American English [9]	**Korean (present study) Excerpts 2-4**	**Korean (present study) All excerpts, 1-4**
Boundary	0.54 - 0.62 M = 0.58	0.47 - 0.63 M = 0.57	0.23 - 0.63 M=0.48
Prominence	0.37 - 0.42 M = 0.4	0.31 - 0.36 M = 0.33	0.17 - 0.36 M = 0.29

The average kappa score of the four scores for Korean prosodic boundaries amounts to an average Kappa score M=0.48 compared to Mo et al.'s [9] average score for boundaries M=0.58. The average kappa score for prominence in Korean is M=0.29 compared to M=0.4 in American English. Overall, annotations for prominence and boundaries in American English are more consistent than in Korean. If we exclude the kappa scores of excerpt 1, the results are very similar for boundaries, but still lower for prominence than in American English. The average kappa score for prosodic boundaries for excerpts 2, 3 and 4 is M=0.57 and for prominence M=0.33. Independent of excerpt 1, the findings show that prominence in Korean is perceived less consistently among listeners than prosodic boundaries.

3.3. Analysis of excerpt 1

Given the comparatively lower Fleiss' kappa scores for both boundary and prominence found in excerpt 1, the question arises why transcribers reach less agreement for excerpt 1. Table 3 summarizes the number of words for each excerpt to illustrate the number of possible targets or locations for prominences and prosodic boundaries. All excerpts are about the same length, that is, 30 seconds long each.

Table 3. *# of words*

#	**Excerpt 1-S1**	**Excerpt 4-S1**	**Excerpt 2-S2**	**Excerpt 3-S2**
#of words	36	61	42	45

As the counts illustrate, excerpts 2 and 3 produced by S2 differ only in three words whereas excerpt 1 is almost half as long as excerpt 4 even though they are from the same speaker S1. Examining its syntactic structure, excerpt 1 consists of a single sentence, with one subordinate clause and one main clause containing two relative clauses. Thus, there are fewer clausal boundaries than in excerpt 4. In contrast to excerpt 4, excerpt 1 has more pauses and words are stretched out. Prosodic boundaries are more easily perceived if the pause between chunks is larger. Thus, pauses may serve as strong boundary cues. However, if the words are stretched out, then pauses may be more ambiguous in identifying prominent words. After listening several times to excerpt 1, a slower speech rate was observed for excerpt 1, which aligns with the previous observation that words are stretched out. In addition, excerpt 1 contains some disfluencies, repeats and restarts, but no fillers. This is interesting because fillers would also account for lower agreement rates causing ambiguity in perceiving prominence and boundaries. Thus, syntactic, lexical and phonetic cues can account for the lower Fleiss' kappa values found in excerpt 1.

3.4. Speaker variation

Figure 2 illustrates the mean interval between prominent words and boundaries for each speaker pooling each speaker's data over all transcribers. Overall, the mean intervals between boundary labels (8-9.5 words) and prominence labels (7.6-10.3 words) are similar and, mean intervals for prominence labels are slightly longer than for boundary labels. The mean intervals for excerpts 2 and 3 produced by S2 are between 8-10 words for both boundary and prominence. Compared to that, the mean intervals for excerpt 1 and 4 produced by S1 show greater variation for the same speaker. Considering the transcriptions labeled for S1, prosodic phrase intervals in excerpt 1 may sometimes have two prominent words and, some prosodic phrase intervals do not even contain a single prominent word in excerpt 4. This contrasts with the prediction supported by linguistic models of prosody that prosodic phrases may lack prominences [2]. Moreover, this pattern of variation not only reflects variation in how speakers produce prosodic information, but also in how listeners perceive those prosodic cues.

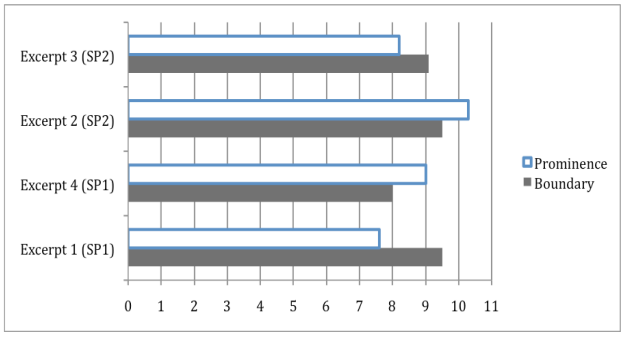

Figure 2: *Mean interval between prominence and boundary labels by excerpts. Each excerpt is pooled across transcribers.*

4. Discussion

Based on the findings of this study, hypothesis 1, that ordinary Korean listeners will perceive prosodic boundaries in conversational speech corresponding to IP boundaries in a K-ToBI analysis, is partly confirmed in that listeners identify prosodic boundaries. However, what is surprising is that the

high scores obtained for boundaries do not entirely correspond to IP boundaries in a K-ToBI transcription. High boundary scores are detected both at sentence-level boundaries, which would correspond to IP boundaries in K-ToBI, but also at clausal-level boundaries. This is interesting because IP-final boundary tones convey sentence types and pragmatic meanings, which predict sentence boundaries, but not necessarily clausal boundaries [5, 6]. Moreover, listeners perceive prominences, but based on the findings for prominence scores and the closer examination of excerpt 1, it can be concluded that prosodic information is not the only information for Korean transcribers to perceive a word as prominent.

In general, the reliability tests conducted in this study reveal that listeners agree in annotating prosodic boundaries and prominence at levels well above chance based on the interpretation of kappa statistics introduced by Landis and Koch [7]. To repeat the interpretation of kappa values, kappa scores < 0 reflect poor agreement, 0.01-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement and 0.81-1.00 almost perfect agreement. This study has shown that non-expert transcribers are consistent in identifying prosodic boundaries and prominent words in Korea. The findings from the inter-transcriber reliability test indicate that Korean listeners show fair agreement for prominences and moderate agreement for boundaries reflecting agreement rates that are at levels above chance confirming the findings of previous work on English as predicted in hypothesis 3 [9, 12]. Hypothesis 4 also turned out to be true as Korean transcribers score higher for prosodic boundaries than for prominence. Boundaries are easier to mark because their prosodic cues such as pauses are more salient to the listener than for prominence. Furthermore, pragmatic meaning in Korean is expressed through the final boundary tone whereas in American English it is the whole intonation contour that conveys pragmatic meaning. Therefore, Korean listeners might be more sensitive to final boundary tones than to prominent words even though prominent words are marked with stronger prosodic cues than non-prominent words in Korean [5].

Speaker dependent variation also plays a role as the comparison of mean intervals between prominence and boundary labels indicates, but not significantly. This might be due to the number of limited numbers of speakers in this study. Based on my findings, individual differences in, for example, speech rate may result in different findings for the same speaker as the analysis of excerpts 1 and 4 illustrate. This also suggests that prosodic cues for boundary and prominence may vary between speakers and also, for the same speaker. As pointed out earlier, this pattern of variation reflects variation both in the production and the perception of prosodic cues. Listeners, for example, seem to adjust to speaker-based differences. Yet, more research with a larger number of speakers and transcribers is needed to account for speaker and listener dependent variation to generalize this finding.

5. Conclusion

This study shows that untrained listeners agree in their immediate perception of prosodic boundaries and prominence in spontaneous speech, and that perception for boundaries is more consistent than for prominence in Korean. For the goals of this paper, RPT has been found to be a reliable and useful tool providing us with interesting findings on Korean and the study of prosody perception in general. Using non-experts' data makes larger segments of talk more accessible and also allows for comparisons between a greater number of annotators. While ToBI and K-ToBI has been developed for experts, RPT enables researchers to study how speech is perceived by ordinary listeners.

For future studies on prosody perception, it would be interesting to examine what other factors than prosodic cues such as parts of speech, frequency of words in terms of familiarity and usage and status of information as new or old as well as topicality play a role in marking a word as prominent in Korean. This study suggests that there are more cues other than higher pitch, longer duration and higher f0 that help to perceive prominence. What these are for Korean and other typologically similar or different languages will be interesting to study in the future.

6. Acknowledgements

I would like to thank Prof. Jennifer Cole for her support and comments on the multiple drafts of my paper.

7. References

[1] Beckman, M., & Pierrehumbert, J., "Intonational structure in Japanese and English", Phonology Yearbook, 3, 255-309, 1986.

[2] Cole, J., Mo, Y., & Baek, S., "The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech", Language and Cognitive Processes, 25: 7, 1141-1177, 2010.

[3] Fleiss, J. L., "Measuring nomial scale agreement among many raters", Psychological Bulletin 76, 378-382, 1971.

[4] Jun, S.-A., "A Phonetic Study of Stress in Korean", poster presented at the 13th meeting of the Acoustical Society of America, St Louis, MO, JASA 98 (5-2), 2898, 1995.

[5] Jun, S.-A., "Korean Intonational Phonology and Prosodic Transcription", in S.-A. Jun (ed.), Prosodic Typology: The Phonology of Intonation and Phrasing. Oxford University Press, 201-229, 2005.

[6] Jun, S,-A., "Prosody in Sentence Processing", in P. Li (Ed.), Handbook of East Asian Psycholinguistics, Part III: Korean Psycholinguistics. London: Cambridge University Press, 423-432, 2009.

[7] Landis, J. R, & Koch, G. G., "The measurement of observer agreement for categorical data", Biometrics, 33, 159-174, 1977.

[8] McClelland, J. L., & Elman, J. L., "The TRACE Model of Speech Perception", Cognitive Psychology, 18, 1-8, 1986.

[9] Mo, Y., Cole, J., & Lee, E., "Naive listeners' prominence and boundary perception", Proceedings of Speech Prosody 2008 (Campinas), 2008.

[10] Pierrehumbert, J. & Hirschberg, J., "The meaning of intonational contours in the interpretation of discourse", in P. Cohen, J. Morgan, and M. Pollack, (Eds.). Intentions in Communication. Cambridge, MIT Press, 1990.

[11] Streefkerk, B., L. Pols, & ten Bosch, L., "Prominence in read aloud sentences, as marked by listeners and classified automatically". IFA Proceedings 21, Institute of Phonetic Sciences, University of Amsterdam, 101-116, 1997.

[12] Yoon, T-J., Chavarría, S., Cole, J., & Hasegawa-Johnson, M., "Intertranscriber Reliability of Prosodic Labeling on Telephone Conversation using ToBI", Proceedings of Interspeech 2004 (Jeju), 2722-2732, 2004.