

The Role of Pitch Range Variation in the Discourse Structure and Intonation Structure of Korean

Eunjong Kong

Department of Linguistics
The Ohio State University, USA
ejkong@ling.ohio-state.edu

ABSTRACT

This study explores pitch range variation in Korean spontaneous narratives and read transcripts of the same narratives. There are two research goals. One is to examine whether pitch range variation helps mark discourse segment boundaries and signal the hierarchy of discourse segment purposes. Another is to see whether categorical differences in pitch range encode the contrast between the two intonationally marked units in the Korean prosodic hierarchy. The narratives were prosodically annotated for Accentual and Intonational Phrase boundaries and the F0 maximum was measured in each Accentual Phrase. Comparing F0 maxima across adjacent phrases shows that pitch range is reset at discourse segment boundaries in spontaneous speech, and also that size of the pitch range reset reflects the hierarchy of discourse segment levels. By contrast, there is no systematic difference in pitch range resetting values between higher Intonational Phrases as compared to lower Accentual Phrase boundaries.

1. INTRODUCTION

A coherent discourse has a discourse purpose, which organizes its subparts in terms their relationships of dominance or satisfaction-precedence to each other and to the larger discourse purpose.[3] Prior research has shown that pitch range is used to demarcate discourse segment boundaries and to mark the hierarchy of relationships among segments. In general there is pitch downtrend within a segment so that a discourse segment begins with an expanded pitch range and ends with a compressed pitch range. Moreover, the degree of expansion at segment boundaries reflects the segments' relationship in the discourse hierarchy. This has been shown for several languages, including English, Japanese, and Mandarin Chinese (see, e.g., [1], [2], [11]). This paper examines spontaneous and read narratives of Korean to see whether pitch range functions in a similar way in a fourth language that differs prosodically from Chinese in lacking lexical tone, and from English and Japanese in lacking any linguistic specification of "pitch accent".

In English and Japanese, pitch downtrend and resetting are also reliable phonetic markers of the intonational hierarchy. This is because a major component of the downtrend is "downstep", a step-wise compression of pitch range related to the linguistic specification of accent. Downstep is blocked at intonational phrase boundaries, so that the size of pitch range reset across clause boundaries can cue whether a following clause is included in the same Intonational Phrase as the preceding clause or begins a new Intonational Phrase. Although Korean lacks any linguistic specification of pitch

accent, it is possible that downtrend works similarly to mark the contrast between the Intonational Phrase and the Accentual Phrase, a lower-level prosodic grouping that is marked by rising pitch at the beginning and/or the end.

This paper explores these two aspects of pitch range control in spontaneous narratives and in read versions of the same narrative texts, because it is possible that the two speech styles might differ in which aspect of pitch range variation is emphasized, as suggested by [1].

2. METHOD

2.1. Speech materials

Three female speakers of Seoul Korean were recorded. Each speaker produced a spontaneous 3-4 minute monologue describing a place (either the speaker's hometown or a recently visited city). The narratives were transcribed in Korean hangul orthography, and then each speaker was recorded reading the text of the monologue she had produced. The interval between the two recordings was less than a week. The transcribed text included cue phrases more characteristic of spontaneous speech, such as fillers *e* and *eom* and sentence-initial conjunctions *georigo* 'and' and *geondae* 'by the way'.

2.2. Discourse analysis

Each spontaneous monologue was segmented using [2]. This is a reduced variant of the model of intentions proposed by [3], providing for only three levels of disjuncture between adjacent clauses: DSP0 (no discourse segment boundary), DSP1 (separate discourse segments sharing related purposes within the larger discourse segment purpose), and DSP2 (two segments with unrelated discourse purposes or "topics"), as in:

[gong-weon-deul-i geu-leo-gu-seon] DSP2 [geo-gi sin-si-nae-ti ta-un-ta-un-e ba-lo gang-i iss-geo-deun-yo]

[the parks are so then] DSP2 [there Cincinnati downtown-LOC right river is]

'So there are parks. There is a river in downtown Cincinnati.'

[sin-si-nae-ti-neun do-si-ga o-lae-doen do-si-yeo-ga-ji-gu] DSP1 [yu-leob-pung-i-geo-deun-yo i-mi-ji-ga]

[Cincinnati-NOM city-NOM old city because] DSP1 [Europe looks like image-NOM]

'Because Cincinnati is an old city, it looks European.'

[di-ge do-lo-do gu-bul-gu-bul iss-go] DSP0 [job-go geu-lae-yo] ([very road winding is] DSP0 [narrow is])

'The road is winding and narrow.'

Two labelers independently segmented each narrative. Disagreements were few and resolved by discussion to make a consensus analysis. The read versions were assumed to have the same discourse structure as the spontaneous monologues.

2.3. Intonation analysis

All six recordings were then annotated for prosodic structure using the K-ToBI labeling conventions [6]. These labeling conventions posit two tonally marked units: the Accentual Phrase (AP: defined by a tonal pattern of “LHLH”) and the Intonation Phrase (IP: marked by pause and final syllable lengthening as well as by final boundary tones L%, H%, etc.). Two labelers independently annotated the recordings, marking prosodic boundary locations at both AP and IP levels. Disagreements were few (less than 5% for any recording). However, since disagreement may reflect somewhat ambiguous marking, boundaries where labelers disagreed (IP in one label and AP in the other or vice versa) were excluded in evaluating the results.

2.4. Acoustic measurements

Pitch range was measured by extracting the maximum F0 (HiF0) in each AP or IP, excluding values on any final syllable labeled as an IP boundary. When the maximum F0 fell in the first syllable of a phrase beginning with a tense or aspirated consonant, the HiF0 value was extracted at the intensity maximum in the vowel to avoid segmental influences from the preceding consonant.

To look at the effects of discourse structure, pitch range reset was assessed at each clause boundary. Degree of reset was measured by comparing the HiF0 across the two neighboring prosodic units on either side (P-2 and P-1 for the preceding units, and P+1 and P+2 for the following units).

An initial examination of the HiF0 patterns at different discourse segment boundaries showed that cue phrases are typically realized in a lower pitch range, even when they introduce a discourse segment at juncture level DSP2. This replicates findings from previous research on English and Japanese (see [1], [4], [5], [8], and [10]). In order to look at the effects of discourse segmentation independent of this effect of cue phrases, we eliminated the HiF0 values for cue phrases in a second set of calculations, so that P+1 would be the phrase after the cue phrase. Phrases identified as cue phrases include fillers, conjunctions and deictic expressions such as *kogi* ‘there’ signaling location in the discourse structure rather than location in space.

To look at the effects of prosodic structure, pitch range reset was assessed at each AP and each IP boundary. Degree of reset was measured by subtracting the HiF0 of the phrase before the prosodic boundary from the HiF0 of the phrase after the boundary, ignoring differences at boundaries where the labelers disagreed, as stated above.

3. RESULTS

3.1. Pitch range variation in discourse structure

Fig.1 shows one speaker’s mean HiF0 values averaged over each of the four positions that were defined to assess the

degree of pitch range reset at different types of discourse segment boundary. The two panels to the left show the means excluding the cue phrases, and the two panels to the right include the cue phrases. The contrast between the two sets of graphs is due to the low pitch range on cue phrases at the beginning of higher-level discourse segments.

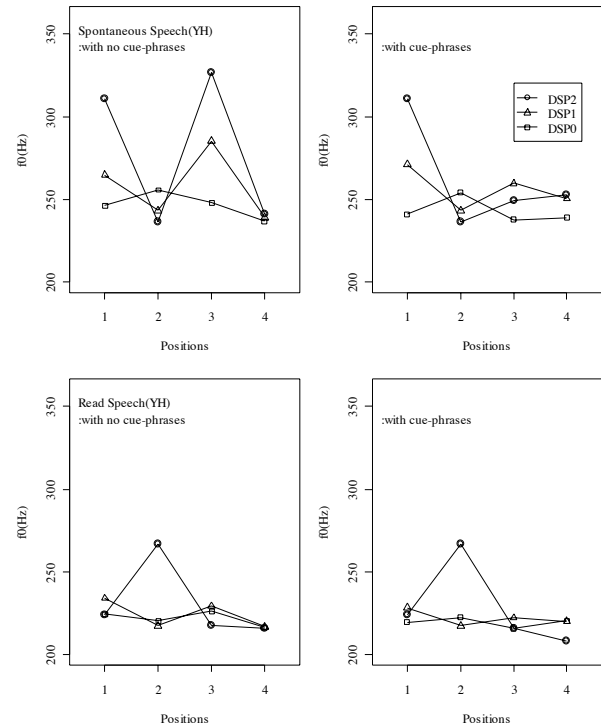


Figure 1: Mean F0 values in prosodic phrases adjacent to the discourse segment boundaries (P-2, P-1, P+1 and P+2). N = 6, 40 and 57 for DSP2, DSP1 and DSP0.

The pattern in the upper left graph shows how pitch downtrend versus reset functions as a discourse boundary marker in the spontaneous narrative. There is a lowering of HiF0 across P-2 and P-1, and across P+1 and P+2, with a reset at boundaries other than DSP0, which is larger at DSP2 than at DSP1. The other two speakers showed the same pattern, and had similar pitch range values overall, so their data were pooled with this speaker’s data in further analyses. A two-way ANOVA for each speech type showed a significant main effect of position both for the spontaneous narratives [$F(3,884)=11.03$, $p<0.0001$] and for the read texts [$F(3,896)=11.01$, $p<0.0001$]. Three planned post-hoc t-tests confirm that the HiF0 differs between P-2 and P-1, between P-1 and P+1, and between P+1 and P+1 ($p<0.0001$ for all 3 comparisons for both styles).

The ANOVAs also showed a significant main effect of DSP level for both styles [$F(1,886)=4.096$, $p<0.05$, for spontaneous; $F(1,898)=3.976$, $p<0.05$ for read]. There were also significant interactions between position and DSP for both styles [$F(7,880)=8.269$, $p<0.0001$, for spontaneous; $F(7,892)=7.305$, $p<0.0001$ for read]. Post-hoc Tukey-b tests between each pair of DSP levels (DSP0 and DSP1, DSP1 and DSP2, DSP0 and DSP2) showed somewhat different patterns

between the two styles. The HiF0 values differed significantly between DSP0 and DSP2 for spontaneous speech ($p<0.05$), but not for read speech. No other differences were significant, suggesting that a large pitch range reset at the discourse segment boundary is a consistent marker of segmentation only for level DSP2 in spontaneous speech. For other contrasts, each speaker differentiated DSP levels by other pitch range manipulations, which differed from speaker to speaker. For example, YH seems to have used downtrend within a DS more than resetting across the DS boundary.

Results for read speech were also complicated by inter-speaker differences in this style. A one-way ANOVA showed no main effect of speaker for spontaneous speech and but a significant main effect in read speech [$F(2,897)=59.66$, $p<0.0001$]. A possible inference from all of these results together is that speakers differ in how well they mark discourse structure in their read speech styles, and that the pitch range patterns of spontaneous speech reflect discourse structure more effectively and consistently across speakers.

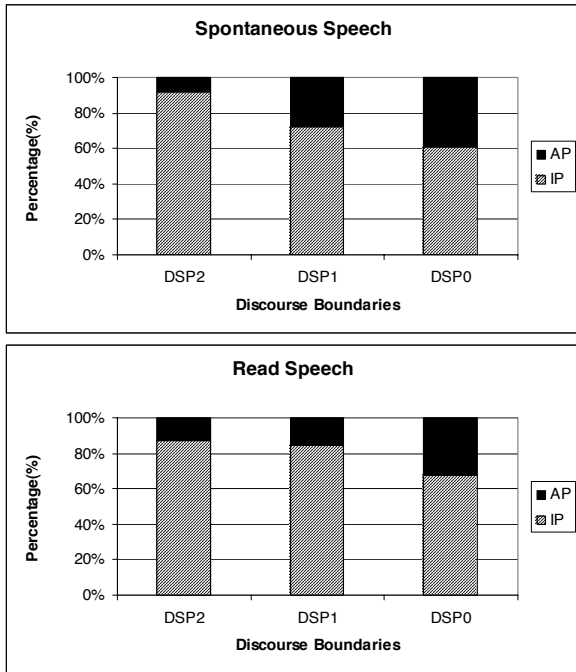


Figure 2: the proportion of IPs and APs in each level of Discourse Segment Purpose (DSP2, DSP1 and DSP0).

Fig. 2 shows the distribution of IPs and APs at each level of the discourse hierarchy. The solid part of each bar is the percentage of APs at that level discourse boundary, while the striped part is the percentage of IPs. A Chi-square test shows that the distribution of IPs versus APs differs across the DSP levels for both types of speech [$\chi^2=26.9078$, $df=2$, $p<0.0001$, for spontaneous; $\chi^2=13.8856$, $df=2$, $p<0.0005$ for read speech].

3.2. Pitch range variation in intonation structure

Fig.3 shows histograms of one speaker's HiF0 differences across boundaries of each prosodic unit type in the two

narrative styles. The solid line is the distribution of HiF0 differences at IP boundaries, while the dotted line is the distribution of HiF0 differences in AP boundaries. HiF0 differences at AP boundaries, interpreted as a measure of degree of reset, were distributed in a similar pattern to those in IP boundaries. No differences in distribution patterns were found between spontaneous narratives and read versions of the same text. The other two speakers also showed the same pattern of overlapped distributions shown as in Fig.3. The results suggest that the degree of pitch range reset does not systematically vary according to the hierarchical level of prosodic units. Thus, the degree of pitch range reset at different prosodic boundaries does not categorically mark the prosodic hierarchy of Korean.

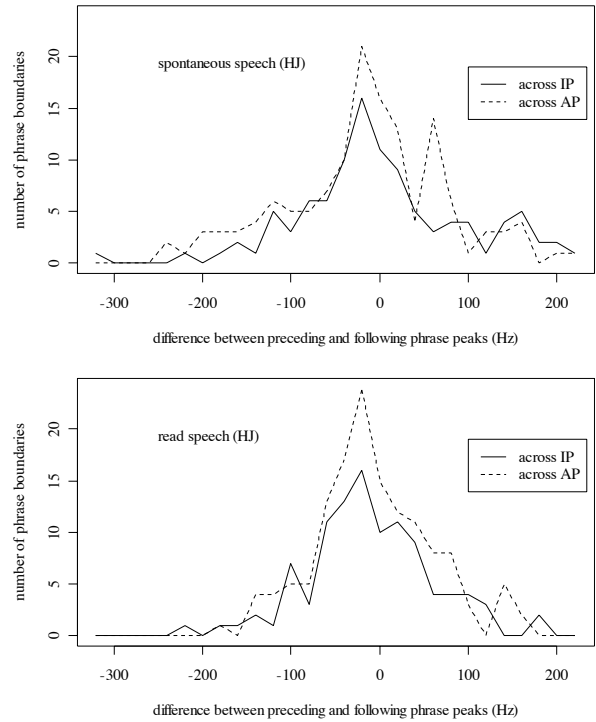


Figure 3: the histograms of HiF0 differences across IP boundaries and across AP boundaries in spontaneous narrative and read versions of the same narrative text.

4. CONCLUSION

This study examined pitch range variation in Korean spontaneous narratives and read transcripts of the same narratives. HiF0 values from prosodic units were assessed to find out whether pitch range trends mark the discourse hierarchy and the intonation hierarchy.

As for the effect of discourse structure, we found that pitch range downtrends and pitch range resets mark discourse segment boundary locations. Also, the degree of reset across discourse boundaries systematically differentiates the levels of the discourse hierarchy in spontaneous narratives. These findings replicate previous research on English, Japanese and Mandarin Chinese. Read speech differed from spontaneous narratives in manipulating degree of pitch range reset as a

marker of the discourse segment hierarchy. Discourse hierarchy was not effectively reflected in degree of pitch range reset in read speech. Further, pitch trends in this speech type revealed significant differences among speakers. All these findings suggest that the spontaneous speech style more effectively reflects the hierarchical discourse structure through use of consistent pitch range control than the read speech style. As for the effects of intonation structure, we found that pitch range reset is not a reliable marker to differentiate hierarchical prosodic units of Korean intonation. In both speech styles, pitch range reset was not categorically used to distinguish IP boundaries from AP boundaries. This indicates that pitch range reset is not necessarily triggered at a hierarchically higher level of grouping units in Korean intonation structure, whereas it is in English and Japanese. This finding supports the description of the IP in the K-ToBI labeling conventions in Korean intonation, which excludes pitch range reset from reliable phonetic cues for IP boundary identification.

5. ACKNOWLEDGEMENTS

I owe thanks to Mary Beckman, Keith Johnson, Soyoung Kang and the discussants in Phonies for their helpful comments on this study. I also thank Jennifer Venditti, Kiyoko Yoneyama and Janice Fon for their insightful discussions.

6. REFERENCES

- [1] Ayers, G. M. "Discourse functions of pitch range in spontaneous and read speech", *OSU Working Papers in Linguistics*, 44:1-49, 1994.
- [2] Fon, Y-J.J. *A cross linguistic study on syntactic and discourse boundary cues in spontaneous speech*, doctoral dissertation, The Ohio State University, 2002.
- [3] Grosz, B. G. and Sidner, C.L., "Attention, intentions and the structure of discourse", *Computational Linguistics*, 12:175-204, 1986.
- [4] Hirschberg, J., and Litman, D. "Now let's talk about now: identifying cue phrases intonationally", *In the Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, 163-171, 1987.
- [5] Hirschberg, J., and Litman, D. "Empirical studies on the disambiguation of cue phrases", *Computational Linguistics*, 19:501-530, 1993.
- [6] Jun, S.A. "K-ToBI (Korean ToBI) Labelling Conventions", <http://www.humnet.ucla.edu/humnet/linguistics/people/jun/ktobi/K-tobi.html>, 2000.
- [7] Kang, H.S. "Acoustic and intonational correlates of the informational status of referring expressions in Seoul Korean." *Language and Speech*, 39:307-340, 1996.
- [8] Kawamori, M., Kawabata, T. and Shimazu, A. "Discourse markers in spontaneous dialogue: A corpus based study of Japanese and English", *In the Proceedings of the ACL-COLING Workshop on Discourse Relations and Discourse Marker*:93-99, 1998.
- [9] Nakatani, C.H., Grosz, B.J., Ahn, D.D and Hirschberg, J. "Instructions for annotating discourses", *Technical Report #TR-21-95*, Center for Research in Computing Technology, Harvard University: Cambridge, MA., 1995.
- [10] Venditti, J.J., and Swerts, M. "Intonational cues to discourse structure in Japanese", *In the proceedings of International Conference on Spoken Language Processing(ICSLP)*: 725-728, 1996.
- [11] Venditti, J. J. Discourse Structure and Attentional Salience Effects on Japanese Intonation. Ph.D. thesis, Ohio State University, 2000
- [12] Yoneyama, K. Koiso, H. and Fon, J. "A corpus-based analysis on prosody and discourse structure in Japanese spontaneous monologues." *Proceedings of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003)*, Tokyo, 2003.