Detecting the Accentual Phrase boundaries in Seoul Korean using tonal and segmental cues

Seung Suk Lee

University of Massachusetts Amherst (USA)
seungsuklee@umass.edu

The most widely adopted model of Seoul Korean intonational phonology (K-ToBI, [1, 2, 3]), proposes that Accentual Phrases (APs)–prosodic constituents larger than Phonological Words (PWds)–can be identified via characteristic tonal events at their junctures [1]. It also proposes that APs are demarcated by a particular kind of perceived juncture, which is transcribed with the Break Index (BI) level of 2 in the model [2]. For example, (1) can be interpreted as (1a) or (1b) depending on where this juncture is located. However, the parsing of APs via non-tonal BI cues has been less emphasized in the literature compared to the tonal marking [4]. Moreover, less is known about what acoustic properties, tonal or non-tonal, might underlie the percept of AP juncture.

Experimental work has shown that listeners are sensitive to the tonal fall from the AP final H tone and the AP initial L tone and use it as a cue for segmentation [5, 6, 7]. APs can also start with an H tone if they start with an Aspirated or Fortis obstruent but in this paper, we focus exclusively on the L-initial APs. The fall associated with the AP-level juncture was shown to be larger and steeper than the fall that may happen within the AP [7, 8]. In (1a), the tonal fall from [ɾim] to [man] (across AP boundaries) should be larger than the fall from [gu] and [tʃʰʌn] (within the same AP), for example. Previous work has also found the AP serves as the domain of segmental allophony of Lenis obstruents: Lenis is voiceless AP-initially but optionally voiced AP-medially [1], which indicates that it can signal whether it is in the AP left edge. The /k/ (in bold) in /ku/ is voiceless when the syllable is AP initial in (1b) but voiced in (1a). Experimental work showed that Korean listeners were sensitive to the allophonic realization of Lenis in the absence of tonal cues [9]. In other words, the AP juncture can be instantiated both tonally and non-tonally and listeners are sensitive to the tonal change between syllables and the segmental details of Lenis in finding prosodic constituents, we don't know how these cues are distributed and how robustly they separate the AP initial syllables ($[_{\text{AP}}\sigma$) from the AP non-initial syllables ($[_{\text{AP}}\ldots\sigma$) in spontaneous speech data.

In this paper, I investigated how PWd initial syllables ($[_{\text{PWd}}\sigma$) and PWd non-initial syllables ($[_{\text{PWd}}\ldots\sigma$) are distributed in the acoustic cue space defined by the tonal and segmental cues, in two speech corpora of Seoul Korean. The first corpus was taken from [10] and it was transcribed in K-ToBI, but it did not have enough Lenis tokens to investigate segmental cues. The second corpus was taken from a spontaneous speech corpus [11] that was larger in size, but it was not transcribed in K-ToBI. Two (one teenage female and one male in his forties) out of forty speakers from the second corpus were investigated as a first pilot. The tonal cue was parameterized as the maximal change in F0 from the previous syllable, normalized as the fraction of range for each utterance. This value was negative when the tonal change was falling from previous syllable. The finding that the tonal fall is larger at the AP juncture than within the AP suggests that the distribution for $[_{\text{PWd}}\sigma$ would be negatively larger than the distribution for $[_{\text{PWd}}\ldots\sigma$ because only $[_{\text{PWd}}\sigma$ can be AP initial under the strict layer hypothesis [12]. The segmental cue was parameterized by combining three common lenition measurements via Principal Components Analysis [13]: percentage of voiced interval [14, 15], the difference between the maximum and minimum rate of change in intensity [16], and the speech-rate-normalized closure duration [9]. It was expected that $[_{\text{PWd}}\sigma$ would have a larger value for this cue compared to $[_{\text{PWd}}\ldots\sigma$, for the same reason as above, since only $[_{\text{PWd}}\sigma$ can be AP initial. The separability of the cue was evaluated by measuring the overlapping area between the kernel density estimate curve of $[_{\text{PWd}}\sigma$ and $[_{\text{PWd}}\ldots\sigma$ which were normalized independently (the hatched areas in figures). The fact that the second corpus is not transcribed gives us a good testing ground whether the acoustic cues can be used to separate $[_{\text{PWd}}\sigma$ from $[_{\text{PWd}}\ldots\sigma$.

The tonal cue separated $[_{\text{PWd}}\sigma$ from $[_{\text{PWd}}\ldots\sigma$ in the expected way in the first dataset (Fig 2): the distribution of $[_{\text{PWd}}\sigma$ was further left compared to that of $[_{\text{PWd}}\ldots\sigma$. This was confirmed in Fig 1, which shows the distribution of $[_{\text{AP}}\sigma$ and $[_{\text{AP}}\ldots\sigma$, which were transcribed in [10]. The distributions were remarkably similar between the two figures since 69% of the PWd formed an AP on their own. However, this was not replicated in the second dataset (Fig 3) which might be since it was not the case in the larger dataset that most [PWdσ were also [APσ. [3] showed that contrastive focusing can cause dephrasing which would increase the proportion of [PWdσ that are not [APσ. It also indicated that the tonal cue is realized with extreme variation and overlap between the categories in a spontaneous speech [4]. On the other hand, [PWdσ was better separated from [PWd…σ on the segmental cue (Fig 4), which shows that the Lenis onsets in [PWdσ were more often realized with larger values than the Lenis onsets in [PWd…σ. This study showcased the concern addressed in [4] that while the prosodic constituents are often defined and claimed to be segmented with respect to their tonal markings, it is not entirely straightforward how this can be done, especially in a larger corpus containing more realistic speech. It also showed that the segmental cues such as the one presented here measuring the allophonic variation of Lenis might be more robust in separating the prosodic categories. Further studies are needed to explore what other robust segmental or tonal cues exist that might allow better detection of AP boundaries.

(1) /pʰa.ɾan.kɨ.ɾim.man.**k**u.tʃʰʌn.wʌn.i.ja/ - parentheses indicate PWd boundaries

a. [(pʰa.ɾan) (gɨ.ɾim)AP] (BI: 2) [AP(man) (**g**u.tʃʰʌn) (wʌn i.ja)]
Gloss: blue painting 19000 *won* be
Translation: The blue painting is 19000 *won* (about $15).

b. [(pʰa.ɾan) (gɨ.ɾim - man)AP] (BI: 2) [AP(**k**u.tʃʰʌn) (wʌn i.ja)]
Gloss: blue painting - only 9000 *won* be
Translation: Only the blue painting (but not other paintings) is 9000 *won* (about $7).

Blue = [AP, Gray = [AP…σ Blue = [PWdσ, Gray = [PWd…σ

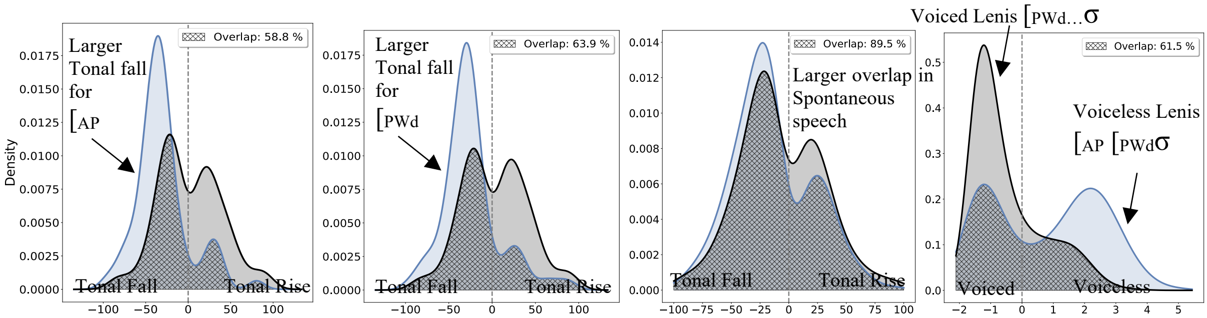

Fig. 1 Fig. 2 | Fig. 3 Fig. 4 Lenis PCA Cue

Dataset 1 | Dataset 2

Tonal cue: Maximum difference in F0 from the previous syllable in percent First Component of PCA

References

[1] Jun, S. A. (1998). The Accentual Phrase in the Korean prosodic hierarchy. *Phonology*, *15*(2), 189-226.

[2] Jun, S. A. (2000). K-ToBI (Korean ToBI) labelling conventions. *Speech Sciences*, 7(1), 143-170.

[3] Jun, S. A. (2007). The intermediate phrase in Korean: Evidence from sentence processing. *Tones and tunes: Experimental studies in word and sentence prosody*, 143-170.

[4] Jun, S. A. (2022). The ToBI transcription system: Conventions, strengths and challenges. *Prosodic theory and practice*, 151-181.

[5] Kim, S., & Cho, T. (2009). The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean. *The Journal of the Acoustical Society of America*, *125*(5), 3373-3386.

[6] Kim, S., Broersma, M., & Cho, T. (2012). The use of prosodic cues in learning new words in an unfamiliar language. *Studies in Second Language Acquisition*, *34*(3), 415-444.

[7] Tremblay, A., Cho, T., Kim, S., & Shin, S. (2019). Phonetic and phonological effects of tonal information in the segmentation of Korean speech: An artificial-language segmentation study. *Applied Psycholinguistics*, *40*(5), 1221-1240.

[8] Lee, J. Y., & Lee, H. Y. (2013). Korean Intonation Patterns from the Viewpoint of F 0 Percentage Change. *Phonetics and Speech Sciences*, *5*(1), 123-130.

[9] Yoo, Kayeon (2020). *The production and perception of domain-initial strengthening in Seoul, Busan, and Ulsan Korean*. Ph.D. Thesis, University of Cambridge.

[10] Jun, S. A., Lee, S. H., Kim, K., & Lee, Y. J. (2000). Labeler agreement in transcribing Korean intonation with K-ToBI. In *INTERSPEECH* (pp. 211-214).

[11] Yun, Weonhee, Kyuchul Yoon, Sunwoo Park, Juhee Lee, Sungmoon Cho, Ducksoo Kang, Koonhyuk Byun, Hyeseung Hahn & Jungsun Kim (2015). The Korean corpus of spontaneous speech. *Phonetics and Speech Sciences* 7:2, 103-109.

[12] Selkirk, E. O. (1986). *Phonology and syntax: the relationship between sound and structure*. MIT press.

[13] Dalcher, C. V. (2007). Statistical methods for quantitative analysis of multiple lenition components. *ICPhS*.

[14] Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*, *54*, 35-50.

[15] Davidson, L. (2018). Phonation and laryngeal specification in American English voiceless obstruents. *Journal of the International Phonetic Association*, *48*(3), 331-356.

[16] Kingston, J. (2008). Lenition. In *3rd Conference on Laboratory Approaches to Spanish Phonology* (pp. 1-31). Cascadilla Proceedings Project.