Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

A cluster analysis of Korean IP-final intonation

Hae-Sung Jeon[1], Constantijn Kaland[2], and Martine Grice[2]

[1]School of Psychology and Humanities, University of Central Lancashire, UK
[2]Institute of Linguistics, University of Cologne, Germany
HJeon1@uclan.ac.uk, ckaland@uni-koeln.de, martine.grice@uni-koeln.de

Abstract

In the Korean Tones and Break Indices (K-ToBI) system, Intonational Phrase (IP) final f0 contours are represented using tones at two levels, Low and High. However, functional analyses have shown the need for two additional levels, Mid and Top. These four levels are also used in some speech synthesis applications. This paper uses hierarchical clustering to investigate the f0 contours on IP final syllables in spontaneous dialogues. Our aim is to develop a classification scheme for analysing surface-level intonational variation in Korean. The clustering results show differences in register for the IP final f0 contours, supporting the usefulness of a phonetic analysis with more than two levels. The IP's position in a sentence seems to affect the IP boundary tone; the low boundary tone for the sentence-medial IP may not be as low as that for the sentence-final IP. The mapping between meaning and boundary tones is not straightforward, and the effect of post-positional particles must be considered when investigating intonational meanings in Korean. Given the variability related to morphosyntactic and sociolinguistic factors, implementing a fine-grained classification scheme that goes beyond two levels is desirable for further exploration of large corpora.

Index Terms: Korean, intonation, boundary tone, contour clustering, spontaneous speech

1. Introduction

To advance prosodic research for real-world application, it is crucial to investigate the use of prosody in speech data collected outside the laboratory [1]. Bearing this in mind, the present study aims to make methodological contributions for analysing pitch events at the Intonational Phrase (IP) boundary in Korean. This study takes a data-driven approach, using a large corpus of spontaneous speech data in which sociolinguistic variation is prevalent [2].

The K-ToBI (Korean Tones and Break Indices) [3] has been widely used for analysing Korean intonation. However, these analyses have been centred on the standard variety, Seoul Korean, and there has been little corpus-based phonetic analysis of dialectal intonation. It is therefore unclear whether the IP boundary tone inventory in K-ToBI is adequate for capturing sociolinguistically more diverse speech. To address these issues, we used the contour clustering application [4] to classify f0 contours based on acoustic similarity and have extended the dialectal area to Chungcheong which in the central area of South Korea. The primary aim was to develop a classification scheme for the IP-final surface-level f0 events for further investigation in sociolinguistic variation.

We first review the existing work on IP boundary tones (§2.1) and provide an overview of the central varieties of Korean (§2.2), followed by information about the dataset and methodology (§3). Section 4 describes the process and interpretations of the contour clustering. Section 5 discusses the findings. Section 6 presents conclusions.

2. Background

2.1. Intonational Phrase Boundary Tones

The K-ToBI system was developed within the Autosegmental-Metrical framework [5]. In this framework, intonation is modelled as strings of High and Low targets and the intonational contour between two successive targets is assumed to be interpolated. K-ToBI provides an inventory of nine IP boundary tones in Seoul Korean, i.e., L%, H%, LH%, HL%, LHL%, HLH%, HLHL%, LHLH%, and LHLHL% [3, 6]. They represent the f0 contour shapes, but their linguistic or paralinguistic functions and their interaction with morphosyntax remain unexplored. In Korean, post-positional particles play important roles, signalling grammatical functions and modality [7, Chap. 9]. As the IP boundary tone is often associated with the post-positional particle(s), the function of boundary tones cannot be elucidated without considering the particle effect. This property of Korean poses challenges in establishing categories for the IP boundary tones. For instance, the sentence ending suffix *-ta* spoken with L% or LHL% shows that the speaker notices something new (e.g., A: /mɑsit*ɑ/ 'it is tasty') whereas H% is used when the speaker highlights the newness pertaining to the already-shared information (e.g., B: /igɛ tʌ mɑsit*ɑ/ 'this one is tastier') [8]. Thus, one may argue for two phonological categories *L% and LHL%* vs *H%* based on their pragmatic usage. However, these categories do not apply to other particles with a different lexical meaning.

While thoroughly examining the intonational realisation of numerous Korean particles would pose serious methodological challenges for researchers for its complexity, some studies, such as Lee [9] and Oh & Kim [10] incorporated the particle effect in their study design. These studies accounted for the particle types in discussing results and proposed variants of K-ToBI with additional Mid and Top tones (Table 1). Although three or more tones can be associated with an IP-final syllable, only monotonal and bitonal tones are presented here due to limited space. (See [10] for a full comparison.)

Table 1: *Mono- and bi-tonal IP boundary tones across studies*

Jun [3]	Lee [9]	Oh & Kim [10]
L%	L%	L%, LL%
H%	M%, H%	M%, H%
HL%	ML%, HL%	ML%, HL%, TL%
LH%	LH%, LM%	LH%, LM%

10.21437/SpeechProsody.2024-207

Lee [9] developed his inventory based on laboratory speech data from five Seoul Korean speakers. They read sentences in different types (e.g., declarative, interrogatives) delivering different attitudes or emotions, such as being friendly, irritated, surprised, etc. In the results, L% was frequent for declaratives and imperatives. While both M% and H% were observed for polar questions and echo questions, H% was more frequently associated with polar questions and M% with echo questions [also see 11].

Oh & Kim [10] analysed 4853 sentences from spontaneous speech and developed their inventory based on pragmatic functions for improving speech synthesis. They argued for a distinction between M and H, because, for instance, an H% (or LH%) tone may turn a statement into a question while an M% (or LM%) tone does not. Oh & Kim also differentiated L% (a level or a gently falling tone) and LL% (a rapidly falling tone); L% indicates the speaker's certainty or confidence while LL% tends to occur when speaker's attitude does not need to be delivered, e.g., in a monologue or when reading noun phrases.

While Lee [9] and Oh & Kim [10] did not explicitly differentiate linguistic and paralinguistic functions [cf. 12], Park [13] classified intonational functions as *informational* (e.g., related to the propositional content and often determined by the sentence-ending particle type), *affective* (e.g., expressing the affective stance of the speaker towards the interlocutor or situation) and *structural* (e.g., indicating the relationship between the utterances or discourse structure), following Gussenhoven [14, Chap. 5]. Park used the two-level convention with L and H tones and suggested that the sentence-medial monotonal boundary tones commonly have the informational use while sentence-finally, they are more likely to have affective or structural use. On the other hand, complex boundary tones which consist of multiple tones mainly project speakers' affective stance.

To summarise, there have been debates on the number of tonal levels required, the effect of particles, and pragmatic functions of the IP boundary tones in Korean. For West Germanic or Romance languages, phonological categories may be established by examining intonational contrasts associated with different modality (e.g., question vs answer) and/or focal structure [cf. 15]. Using similar methods across Korean dialects requires more deliberation, partly because they differ in the types and use of particles and lexical prosody [e.g. 16]. For dialects without a lexical stress or lexical pitch accent, the prosodic cues to the focal structure seem to be spread over an utterance [17]. This makes direct comparison of local prosodic properties across dialects difficult, while the IP-final tones play pragmatic functions across dialects.

2.2. Central varieties of Korean

Six major dialectal regions (northwest, northeast, central, southwest, southeast and Jeju Island) are identified in Korea [18]. The central region covers a broad area, including Seoul, Gyeonggi, Chungcheong, Gangwon, and Hwanghae Provinces. Although dialectal differences were pronounced in the past, recent studies report dialectal levelling in the central region and native Korean speakers classify Chungcheong, Seoul, and Gyeonggi as one dialectal region [19]. These varieties do not have lexical stress, lexical pitch accent or lexical tones [18]. That is, the broad descriptions about Chungcheong, Seoul, and Gyeonggi varieties are similar, but systematic empirical studies are required to assess their phonetic properties.

Given the complexity in intonational analysis (§2.1), we first explored the surface f0 variation in the IP-final syllable. The central question was whether the two-level (i.e., using only L and H) or multilevel descriptions would be more suitable for further investigation on the potential categorical nature and communicative functions of the IP boundary tones.

3. Methodology

3.1. The corpus and data

The Chungcheong speech data were selected from a Korean dialect corpus which includes speech recording, transcripts, and basic demographic information about speakers [2]. Data from three pairs of speakers (2 male pairs and 1 female pair in their 20s) were selected. Because the familiarity between speakers and their relationship affects morphosyntactic marking and speech styles in Korean [7, Chap. 9], speaker pairs who appeared to know each other well (i.e., using informal language and referring to each other without honorifics) were selected.

Each sound file contained a dialogue between two speakers (18–20 minutes long). Each pair had a conversation about literature, travelling and sports. The dialogue consisted of long stretches of each turn, rather than rapid exchanges. Speakers recorded their dialogue on their device such as an iPad using online meeting platforms. No further technical information was available. The sampling rate was 16 kHz and the sound quality was suitable for prosodic analysis.

3.2. Annotations

A forced aligner [20] was used for phoneme-level segmentation of speech data. Then annotators, whose first language was Korean, carried out the audiovisual analysis and annotation using Praat [21] to identify Accentual Phrase (AP) boundaries, IP boundaries, and IP-final syllables following K-ToBI criteria [3]. Author H.-S. Jeon who is extensively trained in prosodic analysis of Korean finalised all annotations.

3.3. Contour clustering

We used Contour Clustering [4] which performs hierarchical agglomerative clustering (HAC) of time-series f0 measures. An alternative is *k*-means clustering [cf. 22]. HAC has the advantage over *k*-means clustering that the clustering process can be inspected using dendrograms before deciding on the number of clusters. (For other clustering methods, see Section 1.6 in [4]). First, the f0 contours were smoothed using an in-application tool (20 time points per syllable, f0 range: 50–400 Hz, f0 fit: 0.8, and smoothing bandwidth: 1). In total, 1080 contours were submitted to clustering (in Hz, f0 standardised for each speaker). The contour clustering application requires users to specify the optimal number of clusters informed by built-in evaluation criteria, Minimum Description Length (MDL) based on information cost [23] and within/between cluster variance (W/B). They identified different numbers of optimal clusters (7 by MDL vs 3 by W/B). We opted for the larger number of clusters; the seven clusters (§4.1) were perceptually distinctive, and they were judged to be potentially functionally contrastive from each other.

We took a zooming-in approach; after initial clustering, we identified three clusters with notable within-cluster variation and carried out another round of clustering to examine their subclusters (§4.2) determined by the evaluation method

described above. Sample sounds in each (sub)cluster were subject to audio-visual examination. In the output plots, the solid line shows the mean, and the grey area shows the standard deviation, with 0 being the centre of the speaker's range. All statistical analyses were conducted with R [24] and RStudio [25]. The package *tidyverse* [26] was used for data processing.

4. Analysis and results

4.1. Initial seven clusters

The initially identified seven clusters (Fig. 1) showed differences in register, i.e., where the f0 events occur in the speaker's range. Zero on the y-axis represents the middle of the speaker's f0 range; Clusters (Cs) 2, 3 and 6 were in the mid-high range and the wide grey area indicated wide variation. These clusters are examined in more detail in Section 4.2.

Cs 1 vs 4 may be phonetically labelled as *M or L* vs *extra-L,* respectively. In the speech samples, the IP-final f0 contour shapes for C1 varied between a rise and a fall, but the end point was above the bottom of the speaker's f0 range. Figure 2 shows an example; f0 falls to the IP-final syllable *-deun* ('touch' + present tense particle) which modifies a noun 'romance' following 'like'. Here, *-deun* is IP-final but utterance-medial; the boundary tone indicates continuity. On the other hand, the f0 contour shapes for C4 varied less as shown in the narrower grey area in Figure 1. They tended to be level at the low register throughout the IP-final syllable, corresponding to the canonical L% tone associated with the utterance-finality.

For C5, f0 contours were distributed in the range higher than expected for the canonical H% tone in read speech (e.g., a rise associated with a question). In the speech samples, the 'extra high' contours were perceptually distinctive from the canonical H% tone with more emphatic impression (Fig. 3). The relationship between the 'extra high' and meaning is further discussed in Section 4.3. Finally, C7 showed falling contours from high to a low-mid register (HL or HM).

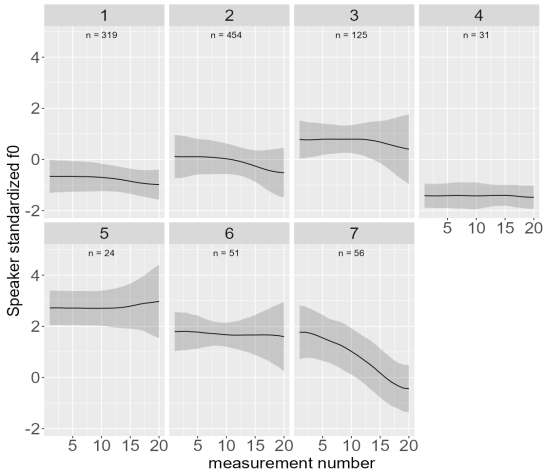

Figure 1: *Initial seven clusters. The x-axis shows the normalised time and the y-axis shows the normalised f0.*

4.2. Zooming in for subclusters

Further clustering was carried out for each of Cs 2, 3 and 6 (Fig. 1) to identify their subclusters (SCs). For the SCs, in audio-visual examination of the speech samples, the relationship between the boundary tones and intonational functions was not clear due to the presence of different post-propositional particles. Therefore, this section reports only phonetic f0 shapes, using one or two tones to describe them.

For C2 (Fig. 1), four SCs were identified (Fig. 4); the averaged contours showed ML for SC1, M or MH for SC2, ML for SC3, and HL for SC4. For C3 (Fig. 1), there were three SCs (Fig. 5); SC1 showing H, SC2 showing MH, and SC3 showing HM. The contours distribute in the speaker's high f0 register. For C6 (Fig. 1), the three SCs were at the high register (Fig. 6). Here, SC1 showed H, a slight fall from a high onset, but the final f0 is still at a high register. SCs 2, 3, 4 are low in frequency (only 4–8 tokens for each SC) and they may be outliers.

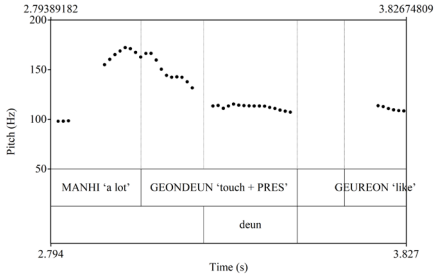

Figure 2: *Example for C1 (Fig. 1), for -deun in '..touched a lot, like, romance' (male speaker).*

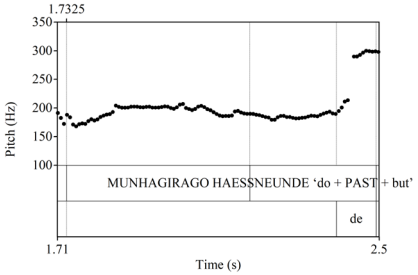

Figure 3: *Example for C5 (Fig. 1) for -de 'but' in '..regarded as literature, ..but..' (same speaker as in Fig 2).*

4.3. Boundary tones and meanings

To control the particle effect (§2.1), we selected one frequent particle, connective *-go* 'and' (n = 135) to examine its clusters. (In fact, the most frequent particle was *-neun* (n = 155) which was not used due to its ambiguity. It could be either a topic particle or a present tense particle.) For *-go* 'and', four clusters were identified (Fig. 7, C1: M, C2: H, C3: L, and C4: HM).

In the audio-visual examination, it was difficult to establish clear mapping between f0 contours in a particular cluster and meaning even though the particle type was controlled. For C2 (Fig. 7), the high f0 delivered excitement or emphasis; an utterance in C2 could be interpreted as a question in isolation, but the question interpretation is unlikely in the given context. If a speaker wants to ask a question ending in *-go* 'and', then an extra high f0 or a large f0 jump between the IP penultimate and final syllables would be required to override the lexical meaning. For C3 (Fig.7), although the f0 contours were in the low register, because of the lexical meaning of *-go* 'and', there was a strong sense of continuity rather than finality.

5. Discussion

5.1. IP boundary f0 contours

This study examined surface f0 contours associated with IP-final syllables. We identified (Figs. 1, 4, 5 and 6): M, L (C1);

M, ML, MH (C2 and its SCs); H, MH, HM (C3 and its SCs); extra-L (C4); extra-H (C5), H (C6) and HL, HM (C7). While the results do not suggest phonological categories, they support a need for phonetic analysis differentiating Low, Mid, High, Top ('extra high') and possibly Bottom ('extra low'). The speaker-normalisation procedure used here led to representations which are similar to the INTSINT system [27] using three levels for absolute Top, Mid, and Bottom tones relative to the speaker's f0 range in addition to High, Same, and Low tones relative to the preceding syllable. Two broad issues identified in this study are discussed below.

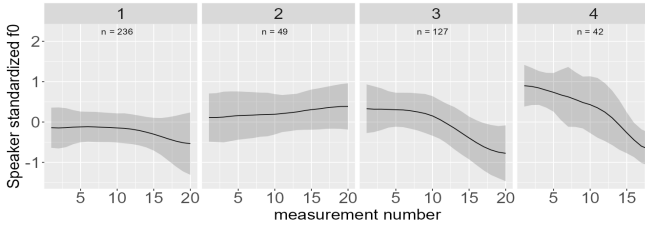

Figure 4: *Four subclusters for Cluster 2 in Fig 1.*

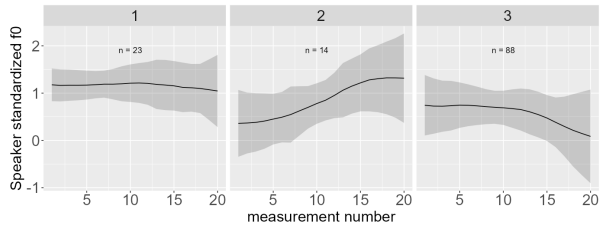

Figure 5: *Four subclusters for Cluster 3 in Fig 1.*

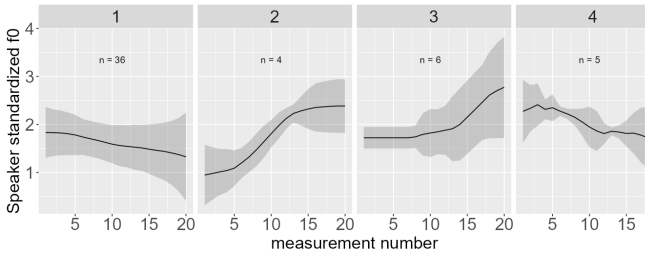

Figure 6: *Four subclusters for Cluster 6 in Fig 1.*

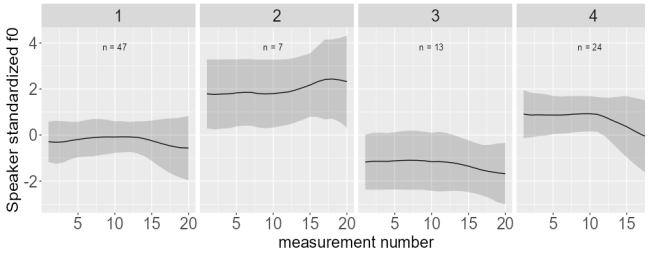

Figure 7: *Four clusters for -go 'and'*

First, there may be intonational differences between sentence-medial and -final IPs. Both Cs 1 and 4 (Fig. 1) would be classified as L% using the current K-ToBI inventory. However, the audio-visual examination revealed more within-cluster variation for C1 with differences in the register from C4. Therefore, in a phonetic inventory with multilevel tones, Cs 1 vs 4 would be differentiated as *M or L* vs *extra-L*. Our examination suggests that C1 may be more likely to be associated with sentence-medial IPs, but further investigation is required to clarify the positional effect. To our knowledge, the positional effect has received little attention; this may be partly because in previous studies, only idealised speech with intended prosodic structure was analysed.

Second, the interplay between particle types, intonation, and context deserves further investigation. Even when only one particle type, the connective *-go* 'and' was examined (§ 4.3), there was no clear one-to-one mapping between an f0 cluster and its meaning. High f0 (C2 in Fig. 7) could deliver emphasis or excitement, but the f0 contours across the clusters seemed to signal continuity; for this connective particle, a strong rise which ends in extra-high pitch or with a large pitch jump between the two final syllables could be interpreted as a question marker. Accordingly, we could assume four functional categories, (1) question marking (surface LH or T), (2) emphatic continuity (H, C2 in Fig. 7), (3) continuity (M, L, and HL, respectively for C1, 3 and 4 in Fig. 7), and (4) finality (L or extra-L). But this impressionistic classification requires empirical validation, and this mapping may not be applicable other particles.

6. Conclusions

This paper investigated the surface f0 contours associated with IP-final syllables in Chungcheong Korean. The results suggest that to further investigate the functions and sociolinguistic and/or individual differences in intonation, using a phonetic model with more than two tonal levels would bring benefits. For instance, the f0 contour distributions and the meaning-intonation mapping discussed in Section 4.3 may be different across Korean dialects. Using the two-level model (restricted to H and L) without accompanying descriptions on variation poses a risk of neglecting the potential semantic or pragmatic contrasts (e.g., the difference between continuity and finality in C1 and C4 respectively, see Fig. 1). It is also important to further examine the f0 movement slope and its functions, for example, in the distinction between L% (a level tone or gentle fall) and LL% (a steep fall) [10], and between HL (a steep fall) and HM (a shallower fall, see C7 in Fig. 1 and SC3 in Fig. 5). By carrying out further fine-grained phonetic, morphosyntactic and functional analysis, we can enrich the current phonological model by clarifying whether and/or when to implement downsteps or upsteps and also contribute to developing an intonational model linked to other linguistic domains [28].

Finally, it should be noted that the present study analysed only the IP-final syllables. The schematic representations in the K-ToBI [3] may seem to suggest that the IP boundary tones are associated with the final syllable. However, the boundary tone can begin earlier [13] and, in fact, the K-ToBI cites examples for the tones beginning from the penultimate syllable. In this sense, applying the contour clustering to domains larger than the final syllable could provide us with further insights on the acoustic and functional classification of the boundary tones [cf. 29].

7. Acknowledgements

This study was supported by the Academy of Korean Studies grant (AKS-2023-R040) and the German Research Foundation (DFG), Project-ID 281511265, CRC 1252 "Prominence in Language". We would like to thank Jaehyung Park, Sunran Shin, and Hyunjung So for their assistance for data annotation.

8. References

[1] F. Cangemi, M. Grice, H. Jeon, and J. Setter, "Contrast or context, that is the question," *Proc. of the 20th ICPhS*, pp. 1360–1364, Aug. 2023.

[2] Saltlux, *Korean Dialect Data*, ver. 1.5. available at: https://aihub.or.kr/, 2021.

[3] S.-A. Jun, "K-ToBI (Korean ToBI) labelling conventions (Ver. 3.1)," *UCLA Working Papers in Phonetics*, vol. 99, pp. 149–173, 2000.

[4] C. Kaland, "Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours," *Journal of the International Ph*onetic *Association*, vol. 53, no. 1, pp. 159–188, 2023.

[5] J. Pierrehumbert, "The Phonetics and Phonology of English Intonation," unpublished PhD thesis, MIT, 1980.

[6] S.-A. Jun, "Korean Intonational Phonology and prosodic transcription," In *Prosodic Typology*, Jun, S-.A., ed. Oxford, UK: Oxford University Press, pp. 201–229, 2005.

[7] H.-M. Sohn, *The Korean Language*, Cambridge, UK: Cambridge University Press, 1999.

[8] H. R. S. Kim, "A high boundary tone as a resource for a social action: The Korean sentence-ender *-ta*," *Journal of Pragmatics*, vol. 42, pp. 3055–3077, 2010.

[9] H.-Y. Lee, "An acoustic phonetic Study of Korean nuclear tones [in Korean]," *Malsori*, vol. 37, pp. 25–39, 1999.

[10] S.-S. Oh and S.-H. Kim, "Modality-based sentence-final intonation prediction for Korean conversational-style text-to-speech systems," *ETRI Journal*, vol. 28, Art. no. 6, pp. 807–810, 2006.

[11] H.-Y. Lee, "H and L are not enough in intonational phonology," *Procs. of the 15th ICPhS*, pp. 2277–2280, Aug. 2003.

[12] D. R. Ladd, *Simultaneous Structure in Phonology*, New York: Oxford University Press, 2014.

[13] M.-J. Park. *The Meaning of Korean Prosodic Boundary Tones*, Leiden, the Netherlands: Brill, 2012.

[14] C. Gussenhoven, *The Phonology of Tone and Intonation*, Cambridge, UK: Cambridge University Press, 2004.

[15] F. Cangemi and M. Grice "The importance of a distributional approach to categoriality in Autosegmental-Metrical accounts of intonation", *Laboratory Phonology,* vol. 7, no. 1: 9, pp. 1–20, 2016.

[16] J. Jun, J. Kim, H. Lee, and S.-A. Jun, "The prosodic structure and pitch accent in Northern Kyungsang Korean," *Journal of East Asian Linguistics*, vol. 15, pp. 289–317, 2006.

[17] H.-S. Jeon and F. Nolan, "Prosodic marking of narrow focus in Seoul Korean," *Laboratory Phonology*, vol. 8, Art. no. 1, 2017.

[18] L. Brown and J. Yeon, "Varieties of contemporary Korean," In *The Handbook of Korean Linguistics*, L. Brown and J. Yeon, eds., John Wiley & Sons, Inc, pp. 459–476, 2015.

[19] L. Jeon, "Drawing Boundaries and Revealing Language Attitudes: Mapping Perceptions of Dialects in Korea," Unpublished MA thesis, University of North Texas, 2013.

[20] T.-J. Yoon, *Korean Forced Aligner*, available at: https://tutorial.tyoon.net/, nd, last accessed 13 September 2023.

[21] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]," ver. 6.1.16, 2023.

[22] J. Cole, J. Steffman, S. Shattuck-Hufnagel and S. Tilsen, "Hierarchical distinctions in the production and perception of nuclear tunes in American English", *Laboratory Phonology*, vol. 14, no. 1, pp. 1–51, 2023.

[23] C. Kaland and T. M. Ellison, "Evaluating cluster analysis on f0 contours: An information theoretic approach on three languages," *Proc. of the 20th ICPhS*, pp. 3448–3452, Aug. 2023.

[24] R Core Team, *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing, 2022.

[25] R Studio Team, *RStudio: Integrated Development for R*, RStudio, PBC, Boston, MA, available at: http://www.rstudio.com/, 2020.

[26] H. Wickham, M. Averick, J. Bryan, and W. Chang, L. D. McGowan, and R. François, et al., "Welcome to the tidyverse," *Journal of Open Source Software*, vol. 4, no. 1686, (Version 1.3.2), 2019.

[27] D. J. Hirst, "A multi-level, multilingual approach to the annotation and representation of speech prosody," In *Prosodic Theory and Practice*, J. Barnes and S. Shattuck-Hufnagel, eds., MA: MIT Press, pp.117–149, 2022.

[28] K. Schweitzer, M. Walsh, S. Calhoun, H. Schütze, B. Möbius, A. Schweitzer and G. Dogil, "Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation," *Speech Communication*, vol. 66, pp. 65–81, 2015.

[29] H. Seeliger and C. Kaland, "Boundary tones in German wh-questions and wh-exclamatives–a cluster-based approach," *Procs of Speech Prosody*, pp. 27–31, May 2022.