JASA ARTICLE

Question intonation in conversational speech: Chungcheong and Gyeongsang varieties of Korean

Hae-Sung Jeon,[1,a)] Constantijn Kaland,[2] and Martine Grice[2]
[1]School of Psychology and Humanities, University of Central Lancashire, Preston, Lancashire, PR1 2HE, United Kingdom
[2]Institute of Linguistics-Phonetics, University of Cologne, Cologne 50931, Germany

ABSTRACT:
This study analyzed unscripted, conversational speech to investigate differences in question intonation in Korean varieties spoken in Chungcheong (CC) and Gyeongsang (GS). It is traditionally assumed that, apart from differences in word-level prosody, these two varieties differ in their intonational marking of questions, the CC variety typically using a final f_o rise, and the GS variety using a final f_o fall. Hierarchical clustering of f_o contours and random forest analysis were used to inspect patterns emerging from the data. Further manual examination was conducted for the individual f_o patterns. An investigation of the effects of variety and question type (alternative, polar, and wh-questions) did not strongly support the traditional account of the dialectal difference. Nonetheless, evidence for the interaction between variety and question type predictors was shown in the distribution of f_o contours, with the GS variety showing greater variation than the CC variety. The approach taken here constitutes an effective analytical pipeline for exploring the distribution of f_o contours in speech data with a high degree of sociolinguistic variation that is generally absent in scripted speech data. This study further shows that combining data-driven and linguistically informed analyses is useful for reconciling phonetic and linguistic approaches to analyzing intonation.
 https://doi.org/10.1121/10.0037191

(Received 23 October 2024; revised 19 June 2025; accepted 26 June 2025; published online 24 July 2025)

[Editor: Molly Babel] Pages: 684–696

I. INTRODUCTION

The standard practice for investigating cross-dialectal phonetic differences has long been that of recruiting speakers of target dialects and recording them reading carefully designed materials. This approach is essential for systematically comparing how speakers of different dialects read written texts. However, to ensure applicability of phonological models for conversational speech and to improve our understanding of speech communication, it is crucial to examine unscripted interactions (Wagner *et al*., 2015; Cangemi *et al*., 2023). A wide range of conversational speech corpora across languages and varieties (e.g., Koutsombogera and Vogel, 2018; Saltlux, 2021) can serve as invaluable resources for researching language in real-life situations.

However, for prosody research, speech data collected outside experimental and laboratory settings poses significant challenges. In particular, conversational speech comes with a great deal of f_o variation owing to factors such as speakers' indexical information, speakers' emotional state, and discourse structure (e.g., Beckman, 1997; Xu, 2011; Hirschberg *et al*., 2020), to name a few. The uncontrolled segmental content of utterances also leads to difficulties in f_o extraction. For analyzing intonation, controlled data often involves utterances that consist of sonorants with minimal segmental perturbation in f_o tracks, whereas natural speech rarely consists of sonorant-only utterances. Although the alternation between sonorants and obstruents may be suboptimal for analyzing intonation, this alternation is optimal for segmenting the speech stream. Furthermore, manual annotation of prosodic events in conversational speech, such as prosodic boundaries and their tones, is time-consuming and prone to researchers' biases (Kaland, 2023a).

This study attempted to tackle the challenges in analyzing unscripted conversational speech while investigating intonational differences between Chungcheong (CC) and Gyeongsang (GS) varieties of Korean. While the intonational model of standard Seoul Korean has been much researched (e.g., Jun, 1998; Jeon, 2015), there has been little cross-dialectal investigation using conversational speech. The CC and GS varieties were chosen because they are claimed to differ in their question intonation. In the literature, the CC variety is traditionally described as marking questions with an f_o rise at the end, whereas the GS variety is described as using a fall, regardless of its modality (see Sec. I B). For the present investigation, question utterances were selected from a large open-access corpus of topic-guided conversational speech across Korean dialects (Saltlux, 2021). The analysis revealed an interaction between variety and question type (alternative, polar, and wh-interrogatives). Cross-dialectal differences were observed primarily in the distributional patterns of f_o contours rather than in a one-to-one mapping between a particular variety and its default pattern, with the GS variety showing greater variability compared to the CC variety. While the observed cross-dialectal differences were

[a)]Email: HJeon1@uclan.ac.uk

attributable to a subset of speakers, these findings provide a useful foundation for subsequent investigations employing controlled datasets.

The aims of the study were twofold. The first aim was to test the workflow using data-driven and statistical methods, agglomerative hierarchical clustering (e.g., Kaufman and Rousseeuw, 1990; Kaland, 2021) and random forests (Breiman, 2001), for unscripted speech data and complementing these methods with manual analysis. The second aim was to assess the dialectal difference. Specifically, the possibility of an interaction between the varieties and question types (alternative, polar, and wh-questions) was tested.

In this article, the intonational representation in Korean (Sec. I A), the CC and GS varieties (Sec. I B), and statistical methods (Sec. I C) are first discussed. Section II presents data annotation and analysis methods. Section III presents the results of the statistical analysis (Sec. III A) and a manual analysis on the individual f_o pattern (Sec. III B). Section IV discusses the findings. Conclusions follow in Section V.

A. Intonational representation in Korean

The basic word order in Korean is subject-object-predicate. Sentence modality such as declarative, interrogative, and imperative is often marked by post-positional particles (Sohn, 1999). However, this morpho-lexical marking is not obligatory. When morpho-lexical ambiguity between a declarative and an interrogative occurs, intonation is often considered a key disambiguation cue, a rise indicating interrogativity (Yoon, 2010).[1]

Prior to discussing question intonation, we provide a brief survey on the intonational analysis of Korean. The most widely used transcription system is Korean Tones and Break Indices (K-ToBI) (Jun, 2000). K-ToBI was developed for Standard Seoul Korean within the Autosegmental-Metrical framework (Pierrehumbert, 1980). In this framework, intonation is modelled as combinations of high (H) and low (L) targets, and the f_o contour between successive targets is taken to be interpolated. A hierarchical intonational structure is assumed (see Jeon, 2015 for survey); the highest-level prosodic unit is the intonational phrase (IP), which consists of one or more accentual phrases (AP). The AP is demarcated by f_o. In the central varieties, including both standard Seoul and CC varieties, the AP-initial f_o is determined by the segmental type; while a fortis or aspirated consonant and a voiceless fricative trigger high f_o, other segments, including a lenis consonant and a vowel, trigger low f_o. IP-medial APs frequently carry a f_o rise at their right edge (Jun, 1998, 2000). While the f_o pattern over the AP has a demarcating function (Jeon and Nolan, 2013), it does not seem to be related to the utterance meaning (Jun, 1998, 2000). On the other hand, the IP boundary tones can differentiate utterance modality and deliver pragmatic meanings. K-ToBI provides an inventory of nine IP boundary tones, i.e., L% (level or fall), H% (rise), LH% (late rise), HL% (rise-fall), LHL% (late rise-fall), HLH% (fall-rise), HLHL% (fall-rise-fall), LHLH% (rise-fall-rise), and LHLHL% (rise-fall-rise-fall). The IP boundary tones are typically associated with the IP-final syllable, but they may begin earlier. While K-ToBI presents the boundary tones with a broad description of their possible meanings (e.g., L% is commonly used in stating facts and in declaratives), their usage in interaction and their communicative functions has not yet been thoroughly explored.

Some studies in Korean pragmatics have examined how IP boundary tones interact with morphosyntax in delivering utterance meaning in a given context (e.g., Kim, 2010; Kim, 2015; Yun, 2023). For instance, in a polar question, the plain question sentence-final particle *-ni* is often used with H%, but the suppositive/committal question particle *-ji*, which can be translated into a tag question in English, is not. In contrast, in a wh-question, *-ji* is often used with H%, while *-ni* is used with HL% (see Park, 2012). These studies offer useful insights into the complex mapping between morphosyntax, intonation, and pragmatic meanings. However, they often classify f_o contours according to the K-ToBI inventory from the data annotation stage and do not present phonetic variants. For instance, although H% is used for both a rise from the bottom of a speaker's f_o range to the mid-level and a steeper rise to a high level, these two phonetic forms may play different functional roles, the former indicating continuation and the latter questioning. Adhering to a single label for f_o contours with different functions may keep us from understanding potentially fine-grained intonational meanings and sociolinguistic variation. In the present study, therefore, we analyze phonetic forms, the initial exploration being based on automatic classification of f_o contours using cluster analysis, i.e., solely based on their numerical similarity (Sec. I C).

B. CC and GS varieties of Korean

The CC variety is classified as one of the central varieties spoken in broad regions including CC, Gyeonggi, Hwanghae, and Gangwon provinces, as well as in the Seoul metropolitan area in South Korea. The central varieties do not have lexical stress, lexical pitch accent, or lexical tones (Brown and Yeon, 2015). While dialectal differences in the central regions were prevalent in the past, the present-day CC variety is considered similar to standard Seoul Korean (Jeon, 2013). While central varieties are traditionally described as using rising intonation for polar questions and falling intonation for wh-questions in line with the cross-linguistic tendencies (Cruttenden, 1997, p.155; Kohler, 2004), recent empirical studies have reported mixed findings. In read speech, polar questions seem to be dominantly marked by a rise, but wh-questions show both rising and falling intonation (see Yun, 2023).

On the other hand, the GS variety is known for its "strong accent" (Jeon, 2013). The Gyeongsang Province is in the southeastern part of the Korean peninsula, and it is divided into South Gyeongsang and North Gyeongsang. South and North GS varieties have lexical pitch accent systems that differ from each other ("snow" /nǔn/ with a rise for South GS vs /nún/ with a high pitch for North GS; Lee and Zhang, 2014). With the ongoing dialectal leveling in South Korea (Jeon, 2013), there is some evidence that the phonetic forms of GS pitch accents have changed over time (Lee and Jongman, 2015; Lee, 2018). For instance, Lee and Jongman (2015) showed that, compared to speakers born before 1952, South GS 'speakers who were born after 1985 have reduced the phonetic differences across the pitch accent categories LL, LH, and HL.

For utterance-final intonation, the GS variety is traditionally known for its frequent use of a final f_o fall for all sentence modalities. Unlike the central varieties, GS varieties differentiate sentence modality by sentence final particles (e.g., statements with *-ta*, polar questions with *-na*, wh-questions with *-no*, and commands with *-la*; Lee, 1988) and, therefore, intonational disambiguation may not be required. Empirical studies support the frequent use of question-final falls in GS, although rises are observed and the intonation is affected by factors such as the utterance-final particle type, focal structure, and word order (Hwang, 2006). If the dominant use of question-final falls is further substantiated by conversational speech data, then the GS variety would represent a further language in which a final fall is used in questions (especially in polar ones), such as those spoken in the Sudanic belt of Africa which have "lax" endings (Rialland, 2009). Many other languages such as Greek, Hungarian, and Romanian, and varieties of Italian and Portuguese have final falls but with the rise being elsewhere in the IP on a pitch accent (Grice *et al.*, 2000; Savino, 2012; Frota, 2002).

There has been little empirical work directly comparing GS and CC intonation. Furthermore, previous studies are limited in scope; they analyzed speech from only one or two speakers from each region. For instance, Han and Oh (1999) analyzed spontaneous conversation on the radio by two GS speakers and two Seoul speakers. They reported that GS speakers used LH% in the IP-final syllable more frequently compared to Seoul speakers. However, this study did not take sentence modality into account. Differences between the South and North GS varieties have also been reported. For example, while falling intonation was used in both South and North GS for polar questions in Kim (2003), subtle cross-dialectal differences were noted. North GS speakers used two distinctive utterance-final intonation patterns, (1) a sustained rise followed by a steep fall in the utterance-final syllable and (2) a gentle fall beginning earlier in the phrase, while South GS speakers used only the gentle fall. Kim (2003) also showed that both patterns were observed for wh-questions across South and North GS speech. However, it is not clear whether the reported differences were attributable to cross-dialectal differences in lexical pitch accent or in question marking.

The present study investigates the effects of variety (CC, GS), question type (alternative, polar, wh-questions, following Yoon's (2010) analysis of conversational speech) and their interaction on f_o contour patterns. While the mapping between question types and f_o contour patterns is unlikely to be one-to-one, cross-dialectal differences are expected in the distributions of f_o patterns. Overall, the GS variety is expected to show more variation compared to the CC variety for three reasons. First, GS speakers are expected to produce falls more frequently compared to CC speakers. Second, the GS data are likely to contain a higher level of variation related to sub-regions in GS and the language change. Third, the lexical pitch accent in GS may affect question-final intonation. Although the effect of lexical pitch accent on utterance-level intonation in GS Korean has not been thoroughly studied, there is some evidence showing their interaction (Hwang, 2006). Furthermore, an interaction between the variety and question type predictors is expected, given that CC speakers may prefer rises for polar questions and falls for wh-questions, while GS speakers are more likely to produce falls across question types.

C. Contour clustering and random forest

To explore the distribution of f_o contours in conversational speech data, this study used agglomerative hierarchical clustering (Kaufman and Rousseeuw, 1990; James *et al.*, 2013) in conjunction with random forest analysis (Breiman, 2001). This approach is data-driven; automatic analysis is carried out purely based on the physical properties of the speech signal. The Contour Clustering application (see Kaland, 2023a for theoretical background) used here was developed to account for phonetic variation without imposing pre-defined intonational categories on data (Ladd, 2022). It offers options for normalizing between-speaker variability and conducts bottom-up hierarchical agglomerative cluster analysis, which starts with all observations in separate clusters. The clustering process merges clusters based on their numerical similarity as expressed by a distance metric and a linkage criterion (James *et al.*, 2013). The outcome of the clustering is a tree structure representing the merging process (dendrogram), with a specific height in the tree corresponding to the number of clusters. A low number of clusters allows users to see broad grouping, while increasing the number of clusters allows more fine-grained grouping. The clustering outcomes do not suggest phonological categories, but they are useful for examining the distribution of surface f_o contours in the dataset. Cluster analysis is a promising method to explore intonation (Kaland, 2023a), particularly when the prosodic properties of the language under investigation are not well-known. Recent studies have combined it with other statistical techniques such as random forest or generalized additive mixed modelling (GAMM) (e.g., Cole *et al.*, 2023; Kaland and Grice, 2024; Steffman *et al.*, 2024).[2]

Random forest analysis (Breiman, 2001) can provide an account of robust predictors for f_o contour clustering outputs. Random forest analysis is a machine-learning technique of classifying data by constructing a large number of decision trees using randomly selected training datasets and random subsets of predictors. The decision trees are used to assign values to predictors that are related to the classification

https://doi.org/10.1121/10.0037191

outcome. Random forests are useful for determining important variables given a small number of observations, and they may show higher accuracy compared to other classification techniques (Speiser *et al.*, 2019).

Clustering and random forests were successfully used by Kaland and Grice (2024), who analyzed IP-final f_o patterns in unscripted task-oriented dialogues in Papuan Malay. Participants were given seven shapes (triangles, a square, etc.) in pictures on a card. One of the shapes was indicated with an arrow, and participants in dyads were asked to find out whether the indicated figure was the same or different between their cards. Kaland and Grice (2024) carried out cluster analysis to examine f_o patterns associated with the same disyllabic words in phrase-final position ($n = 324$ from 35 speakers). They also manually labelled the speech data for whether the IP-final constituent was turn-medial or -final, topic continuation, and also information structure (information status and contrastivity, i.e., whether the word was mentioned before and whether it contrasted with another word, respectively), word class (adverb, conjunction, demonstrative, etc.), and syllable structures. These labels were used for a random forest analysis to determine the predictors' importance. Results showed that f_o contour variation, in particular f_o direction and target level, was best explained by whether the IP was turn-medial or -final.

The present study used fully topic-guided (but not task-oriented) dialogues in two dialectal groups. The present data were expected to show greater variation because the two dialects are reported to have different question-final intonation and the speaker's choice of lexical items was not controlled.

To summarize, this study first aimed to combine statistical methods and manual examination of f_o contour variation in conversational speech data. The cluster analysis classified all f_o contours over IP-final two or three syllables based on their numerical similarity. The clustering could be affected by both the f_o level relative to speakers' range f_o and the movement direction (rise, fall, flat). Then, the random forest analysis was carried out for the clustering output to identify the important predictors for the classification. Subsequently, f_o contours associated with the IP-final syllable were classified focusing on the movement direction (e.g., falls, rises) regardless of the f_o level. The second aim was to assess the dialectal difference (CC vs GS) and the potential interaction between variety and question type (alternative, polar, and wh-questions). The differences in the distribution of f_o patterns are expected and an interaction between variety and question type predictors in the random forest analysis.

II. DATA AND ANALYSIS

We combined contour clustering and random forest analysis for f_o contours measured over two or three IP-final syllables (henceforth "multisyllabic domain") to examine (1) the distribution and clustering of f_o contours based on their acoustic similarity and (2) whether the clustering is related to the predictors, variety (CC, GC), question type (alternative, polar, and wh-questions), and the interaction between variety and question type. Contour clustering and random forest were carried out in parallel, as described in the supplementary materials. While these analyses classified f_o contours purely based on numerical similarities and the classification would be affected by statistical distribution in the dataset (i.e., a small cluster can be merged into a larger cluster), the manual analysis approximated the traditional analysis of the IP-boundary movement.

A. Data

This study used a corpus of Korean dialects (Saltlux, 2021), including speech sound files (.wav), transcripts, and basic demographic information about speakers. The speech data were collected in different locations, such as CC, Gangwon, GS, Jeju, and Jeolla Provinces. Each sound file contains a conversation between two speakers about one topic (18–20 min long), such as literature, travelling, food, and sports. Some speakers were recorded in a quiet office using a headset with a microphone. Other speakers used their mobile devices, such as an iPad and online meeting platforms during the Covid-19 pandemic. No further technical information was available. The sampling rate was 16 kHz. The quality of recordings in the corpus varied. Some recordings included significant background noise, which made reliable acoustic analysis impossible. We examined the quality of each sound file by listening to it and inspecting its spectrogram using Praat version 6.1.16 (Boersma and Weenink, 2023). Files with no or little background noise were selected for accurate f_o measurements.

B. Speakers

Data from six pairs of speakers were selected from each of CC (four male pairs and two female pairs) and GS (two female-male pairs and four female pairs) sets. Information about the speakers' main places of residence was available; while some speakers named a specific city (e.g., Daejeon), others named a province (e.g., South CC, North GS). CC speakers' main places of residence were North CC (six speakers), Daejeon (the largest city in CC, three speakers), South CC (two speakers), or Sejong (a self-governing city at the administrative boundary between South and North CC, one speaker). GS speakers' main places of residence were Busan (South Korea's second-largest metropolis after Seoul in South GS, five speakers), Ulsan (in South GS, three speakers), South GS (two speakers), and Daegu (two speakers, close to North GS). All the speakers were in their 20 s. Because the familiarity and hierarchical relationship between speakers affects morphosyntactic marking and speech styles in Korean (Sohn, 1999), only speaker pairs who appeared to know each other well (i.e., using informal language and referring to each other without honorifics) were selected for analysis.

C. Annotations

Praat version 6.1.16 (Boersma and Weenink, 2023) was used for all annotations and acoustic analyses. A forced aligner (Yoon, 2023) was used for phoneme-level segmentation of speech data. Research assistants who were trained in prosodic annotation carried out the first-stage annotation. They used the Busan (GS) variety as their first language, and they were familiar with the central varieties. They manually corrected the phoneme-level segment boundaries, following the criteria in Turk *et al.* (2006). They marked AP and IP boundaries respectively in two interval tiers in each TextGrid. The AP boundary was identified by the presence of the AP boundary tone; the right edge of the IP was identified by the presence of the IP-boundary tone and significant final lengthening (Jun, 2000). They labeled syntactic constituents for each AP. For example, a case marker was classified as a nominative marker, a topic marker, and an accusative marker or other. For each IP, the sentence modality was annotated as a statement, wh-question, polar question, alternative question, command, suggestion, and direct quotation. This classification was based on the lexicogrammatical properties of the target utterance. For instance, "is there any place you want to go for your holiday?" was classified as a polar question, although the response was "I want to go to Paris" without an explicit yes-no response. The question type classification (polar, wh-, and alternative questions) followed Yoon (2010). There were 255 IPs produced as questions out of 6011 IPs in total. Further, the final or penultimate particle was annotated in Romanized form. The APs are not analyzed in the present study. The analysis results for particles are available on the Open Science Framework (OSF).

H.-S. Jeon, who is trained in prosodic analysis of Korean, checked all first-stage annotations. One interval tier for each sound file was created. Then a single interval was created to mark the beginning of the final two or three syllables in each IP and the IP offset. When an IP-final AP was disyllabic, the beginning of the penultimate syllable was marked. When an IP-final AP had three or more syllables, the beginning of the final three syllables was marked. Each interval contained information about the question type (wh-, polar, and alternative) and the IP-final particle type.

D. Contour clustering

For multisyllabic domain analysis (with the analysis window over two or three syllables), time-series f_o measures were taken using the Contour Clustering application (settings: fit 0.9, time step 10 ms, $50-475$ Hz range, 30 measurement points) (Kaland, 2023a). While other clustering methods are available (Kaland *et al.*, 2024), the application facilitates semi-automatic analysis and visualization of the clustering outputs. The Contour Clustering application (Kaland, 2023a) applies f_o interpolation and extrapolation, i.e., all missing f_o points were filled in by linear interpolation and missing points at the edges were filled in by linear extrapolation using a constant based on the first/last available f_o point. Thereafter, the f_o contour was smoothed using kernel density estimation (KDE) (Silverman, 1986). The outcome of the KDE method was multiplied by the smoothing factor. This provides additional control over the smoothing accuracy, with lower smoothing values leading to more accurate but less smooth tracking of the measured f_o values and higher smoothing values leading to less accurate tracking and smoother contours. Not all f_o values could be tracked; this was shown in the sampling plots, and unavailable data (NAs) were written to the output file. The application automatically rejects observations with NAs. After the data cleaning procedure, 185 f_o contours out of 255 question IPs were left for analysis. The f_o values measured in Hertz were converted to ERB, and they were normalized for speakers using z-transformation to account for differences in f_o level and range (Rose, 1987). The ERB scale was preferred for comparing falls and rises on perceptual grounds (Jeon and Heinrich, 2022).

Complete linkage hierarchical clustering was carried out using Contour Clustering (2025-03 version; Kaland, 2023a) based on Euclidean distances. These settings were chosen as the default, as they perform reasonably well across languages (Kaland, 2023a,b). The number of clusters varied over the course of a series of cluster analyses (henceforth "rounds"). This process requires users to adjust various parameters (e.g., the range of the number of clusters) and generate a clustering plot for each round. The range of number of clusters (i.e., the minimum and maximum numbers of clusters expected) was set from two to eight. The ideal number of clusters was determined by outputs of clustering and random forest (Sec. III A). In the clustering and evaluation process, users must note that outcomes are solely based on the numerical properties of the f_o contours. For linguistic analysis, it is recommended that users check plots and data distribution across outcome clusters for each round (Kaufman and Rousseeuw, 1990; Scitovski *et al.*, 2021).

E. Random forest

A series of random forest analyses was carried out for each round of clustering using R version 4.4.0 (R Core Team, 2024), R Studio Build 561 (R Studio Team, 2023), and the package "ranger" (Wright and Ziegler, 2017). The response variable was the f_o clusters for each clustering round (between two and eight clusters), and the predictors were speaker (24 speakers), variety (two levels: CC, GS), question type (three levels: polar, wh, alternative), and the interaction between variety and question type. In addition, a random control predictor was added as a reference for a plausibility check in interpreting the importance of other predictors by sampling a random number from one to ten (with replacement) for the total number of observations (185 times). Given its randomness, the control predictor is not expected to improve the prediction on clustering outputs and should, therefore, have a low importance value (<0). Variable importance values of predictors that lie around or below the control predictor indicate that the predictors concerned do not affect the response variable, which was the clustering outcome (see Kaland *et al.*, 2021 for the same procedure).

The number of trees in the analysis was increased in steps of 10 000 starting from 10 000 trees. This procedure enables an assessment of the stability of variable importance values. If repeating the random forest analysis with a certain number of trees shows different rankings across the rounds, the number of trees is too low. The variable importance of the predictors reached a stable ranking around 50 000 trees, which was taken as the final number of trees throughout the clustering rounds. The number of randomly preselected predictors in the tree-building process was set to the square root of the total number of predictors in the analysis ($\sqrt{5}$), and variable importance mode was set to "permutation." These settings are recommended for analyses with correlating predictors (Strobl *et al.*, 2008; Strobl *et al.*, 2009).

F. Manual analysis

H.-S. Jeon carried out manual classification of the f_o contours associated with the IP-final syllable. The results of clustering and random forest (Sec. III A) highlighted the differences in the f_o level over multisyllabic domains (i.e., IP-final two or three syllables) across the clusters while we were also interested in the IP-final f_o patterns specifically associated with the final syllable. It was possible to carry out further contour clustering only for IP-final syllables; however, in a preliminary analysis of a data subset, the clustering output largely differentiated f_o levels rather than the contour shape (Jeon *et al.*, 2024). The contour shape was classified as fall (F), rise (R), mid (M), and short rise (sR) (Fig. 1)[3]. The sound files and TextGrids were examined using Praat version 6.1.16 (Boersma and Weenink, 2023), and the classification was based on the visible f_o track (range 60–500 Hz, raw autocorrelation) and perceived intonation. While the four-way classification was feasible for the present data, using the IP tone inventory of K-ToBI (Jun, 2000) was problematic for two reasons. First, there were cases with a flat stretch of f_o near the mid-level of the speaker's f_o range as noted in the literature (Lee, 1999; Kim, 2003; Lee, 2007). Using two tonal targets, H and L, would implicate some kind of phonological categorization of these cases, which we intended to avoid. Therefore, a flat stretch of f_o around the mid-level of the speaker's f_o range was annotated as M. Second, while the timing of the f_o turn is crucial for differentiating some IP boundary tones (e.g., H% for a rise vs LH% for a late rise), the turning point was often difficult to locate, owing to consonantal perturbation. Finally, the sR label was used for a f_o rise followed by a f_o dip at a latter part of the IP-final syllable (i.e., a rise-fall-rise), a short rise constituting the last portion of the IP-final f_o. The short rise could be interpreted as a complex HLH% boundary tone in the K-ToBI inventory. However, the short rise seemed to differ in its phonetic shape from the schematic representation of HLH% and their canonical realizations in Seoul Korean (Jun, 2000), which shows a large excursion for the fall-rise portion.

III. RESULTS

A. Statistical analysis results

The number of question IPs was similar across CC ($n = 95$) and GS ($n = 90$). The majority were polar questions ($n = 105$) and wh-questions ($n = 75$). There were only a few alternative questions ($n = 5$). The frequency distribution across speakers was uneven. For CC, three speakers produced 62% of the question IPs (19–20 IPs from each speaker); others produced only two to eight question IPs. For GS, three speakers produced 46% of the question IPs (10–22 IPs from each speaker); others produced one to nine question IPs.

The optimal number of clusters identified based on minimal description length (MDL) (Rissanen, 1978) was five. The MDL is based on a measure of information cost (Kaland and Ellison, 2023). However, the six-cluster output was chosen for interpretation to allow examination of falling contours, which were expected for GS and of our interest. The six-cluster output (Fig. 2) captured useful f_o variation not shown in the outputs with fewer clusters. However, for the four-cluster output (see OSF supplementary materials A), falls were not clearly visible in the clusters while one of the clusters ($n_0 = 51$; n_0 refers to one large cluster, which was split into n_1 and n_2 in the subsequent round) showed wide variance, possibly including falls. In the five-cluster output, the large cluster ($n_0 = 51$) was divided into two clusters ($n_1 = 46$, $n_2 = 5$), but one of them ($n_1 = 46$) still showed wide variance. However, for the six-cluster output (Fig. 2), the large cluster ($n_1 = 46$) from the four-cluster round was divided into clusters 1 ($n = 30$) and 4 ($n = 16$), with cluster 4 showing a falling averaged f_o contour.

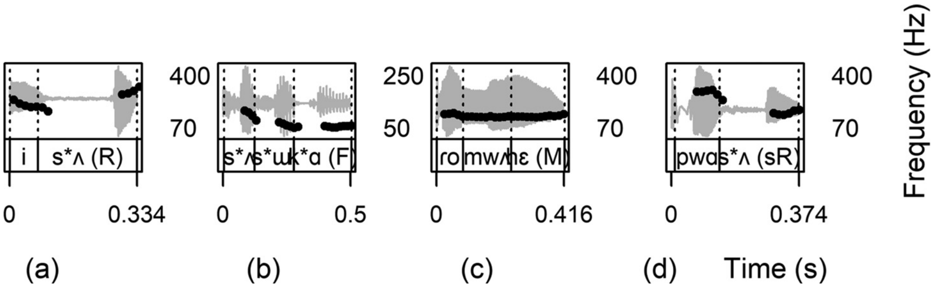

FIG. 1. Sample f_o contours for (a) a rise (R), [is*ʌ] *to be* + INT, (b) a fall (F), [s*ʌɯlk*ɑ] PAST + INT, (c) a mid-tone (M) [ɾomwʌhɛ], *by what-do-*INT, and (d) a short rise (sR) [pwɑs*ʌ], *see* + PAST + INT. Figure created using "praatpicture" Version 1.5.0 (Puggaard-Rode, 2024).

https://doi.org/10.1121/10.0037191

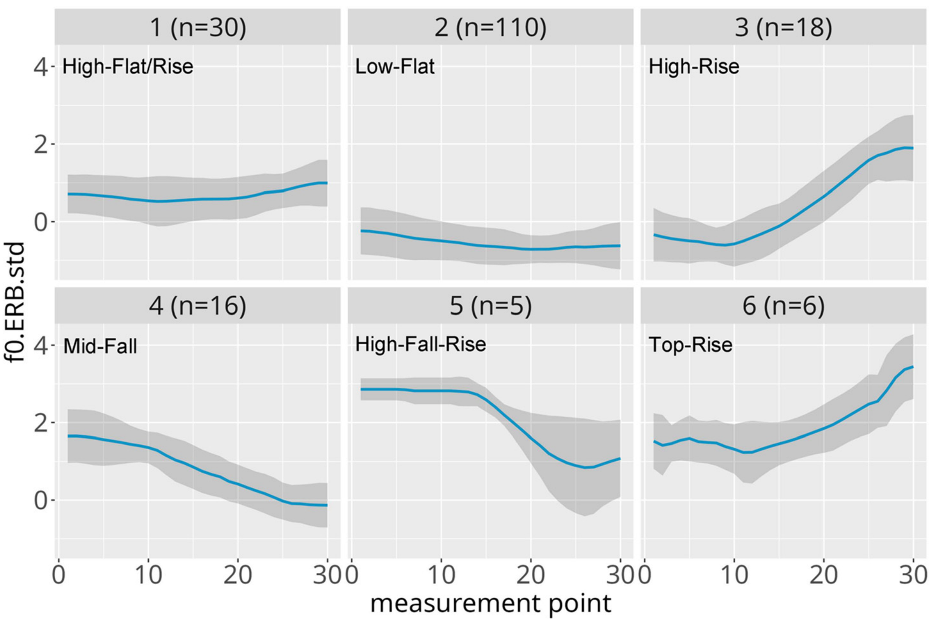

FIG. 2. Contour clustering output with six clusters showing the averaged f_o contour and standard deviation for each cluster. The x axis shows the normalized time and the y axis shows the normalized f_o, with zero indicating the middle of the speaker's f_o range. Each cluster is labelled for the level (e.g., low/mid/high/top) relative to speakers' f_o range and the f_o movement pattern (rise/fall/flat) at or towards the end of the averaged contour.

For interpreting the clustering output (Fig. 2), we focus on the latter half of the averaged contour as f_o patterns associated with the IP boundary because the measurement was taken over two or three syllables. Each cluster in Fig. 2 is labelled for the level (e.g., low/mid/high/top) relative to speakers' f_o range and the f_o movement pattern (rise/fall/flat) at or towards the end of the averaged contour. The latter part of the averaged contours showed distinctive movement patterns across the six clusters. While each cluster was labelled for the averaged contour of the latter portion for convenience, it is important to examine standard deviations (Fig. 2); for instance, cluster 1 could include a rising, flat, or falling contour, while a flat contour followed by a gentle rise represented as the average contour may have been frequent.

For cluster 1 (high-flat/rise), the averaged contour shows a gentle rise towards the end to a high level, but its slope is not as salient as what is shown in other clusters with a fall or a rise. Therefore, it was labeled as high-flat/rise. For cluster 2 (low-flat), the latter portion of the f_o contour is flat at the low level. For cluster 3 (high-rise), the f_o contour shows a steep rise to the high level. For cluster 4 (mid-fall), the f_o contour gently falls towards the end with a short flat stretch. For cluster 5 (high-fall-rise), the f_o contour falls from a high level, and there is a small rise at the end; the contour begins from a high level, but there was a high deviation towards the end. Therefore, a manual examination was carried out for each IP for cluster 5. In all tokens, the large deviation was due to consonantal perturbation, and there was a f_o rise in the IP-final syllable. For cluster 6 (top-rise), the f_o contour rises to a level higher than what was shown for clusters 1 (high-flat/rise) and 3 (high-rise).

The frequency distribution of f_o contours across the clusters was non-uniform (Fig. 2); the majority were in cluster 2 ($n = 110$, out of 185 contours). That is, the majority of f_o contours were near the middle of the speaker's f_o range.

Importantly, the random forest outputs (Fig. 3) identified the variety × question type interaction, speaker, question type, and variety as important predictors. The output of the four-cluster round corroborates that the four clusters were not optimal; the random number predictor was identified to be important (>0). For the five-cluster round, the order of importance was variety × question type (highest importance) > speaker > question type > variety (lowest importance), and the random number (<0) worsened the prediction. For the six-cluster round, the order was variety × question type (highest importance) > variety > speaker > question type (lowest importance). The random number (<0) worsened the prediction. On the other hand, having seven

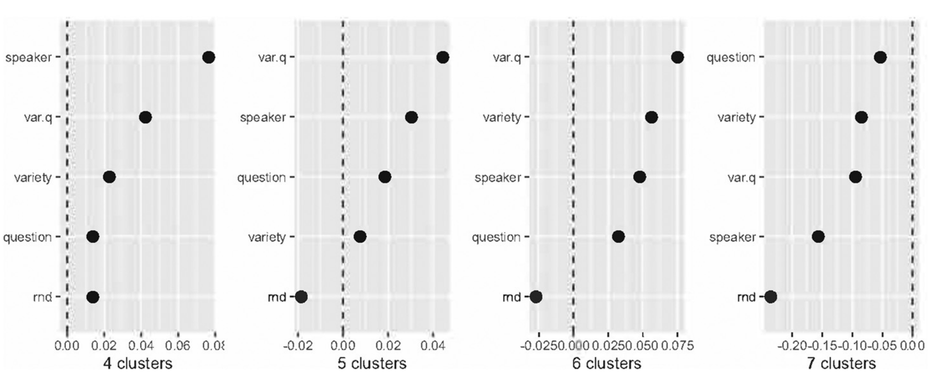

FIG. 3. Random forest outputs for four-, five-, six-, and seven-cluster rounds: var.q, variety × question type interaction term; rnd, random number.

clusters seemed superfluous; all predictors worsened the prediction of the cluster grouping (variable importance for all predictors <0).

The importance of the speaker predictor seemed to be a by-product of the uneven distribution of contours. The speaker distribution across the clusters is available as supplementary materials (see OSF). The largest cluster 2 ($n = 110$, low-flat) (Fig. 2) contained f_o contours from all speakers except for two GS speakers, and therefore, the speaker would not be a determining predictor. On the other hand, smaller clusters included contours from a few speakers, showing cluster-speaker association. Two speakers (out of 24) contributed to only one cluster, and 12 speakers contributed only to two clusters.

The frequency of f_o contours in each cluster by variety and question type is provided in the OSF. There were only a few alternative questions ($n = 5$) in cluster 1 (high-flat/rise) for CC, cluster 2 (low-flat) for CC, and cluster 4 (mid-fall) for GS. Both polar and wh-questions were found in both varieties for cluster 1 (high-flat/rise), cluster 2 (low-flat), cluster 3 (high-rise), and cluster 4 (mid-fall). For cluster 5 (high-fall-rise), only polar questions were observed for both varieties, but the falling portion in the f_o contours was caused by consonantal perturbation as discussed earlier. For cluster 6 (top-rise), both polar and wh-questions were observed for CC, but only polar questions were observed for GS. This distribution for cluster 6 (top-rise) suggests a variety × question type interaction. However, this cannot be confirmed due to the low number of contours ($n = 6$).

Finally, instances of multitonal IP-boundary tones, LHL%, HLH%, LHLH%, LHLH%, and LHLHL% in the K-ToBI inventory (Jun, 2000) did not seem to have frequently occurred. The clustering outcomes (Fig. 2) showed differences in the f_o level and movement pattern across the clusters and the latter portion of the averaged contour included only a rise (clusters 1, 3, 5, and 6), a flat stretch (clusters 1 and 2) and a fall (cluster 4). While this finding does not indicate that the multitonal boundary tones did not occur at all, probably the contrast between monotonal (e.g., L%, H%) and multitonal tones was not salient or necessary for grouping f_o contours based on acoustic similarity and frequency.

It is postulated that the level differences in Fig. 2 were linked to a range of factors such as the AP-initial segment type, presence of a pitch accent in GS, the uncontrolled focal structure in the utterance, and speaker-specific behaviors. The findings informed the following analysis focusing on the IP-final syllable. For example, it appeared to be necessary to annotate flat f_o stretches represented in the averaged f_o contour for the largest cluster 2.

B. Manual analysis of IP-final syllables

The most frequent f_o pattern for the IP-final syllable was a rise (R) (CC, $n = 82$; GS, $n = 51$) followed by a fall (F) (CC, $n = 10$; GS, $n = 11$). Some IP-final syllables had a flat f_o stretch at the mid-level (M) (CC, $n = 3$; GS, $n = 8$). The short rise (sR) (a rise followed by a dip in the latter part of the syllable, creating a short rise towards the end) was observed only for GS ($n = 19$). Chi-square test results showed that the variety ($\chi^2 = 28.57$, df = 3, $p < 0.001$) and question type ($\chi^2 = 16.05$, df = 6, p = 0.01) predictors were related to the f_o pattern classification.

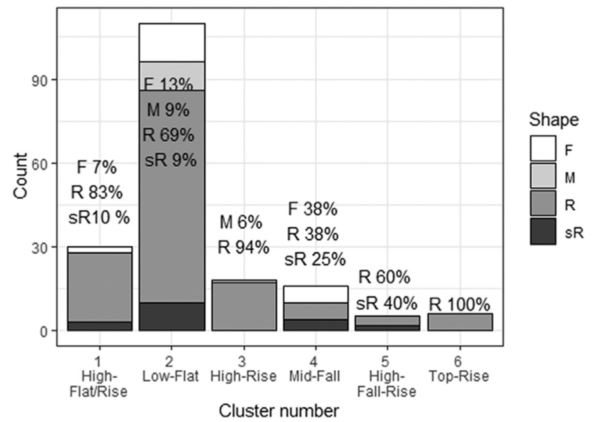

FIG. 4. Count and proportion of manually annotated IP-final f_o contour shapes, fall (F), mid (M), rise (R), and short rise (sR) for each cluster. The x axis shows semi-automatically classified clusters 1–6.

Figure 4 presents the relationship between the manual annotations and the clustering outcome. Because the f_o contour level over the IP-final two or three syllables was an important determiner for the clustering outcome, it does not always correspond to the manual annotations for the IP-final syllable focusing on the f_o movement direction (e.g., R is not always in the clusters with an averaged contour with a final rise). Rises and short rises were observed across the six clusters. On the other hand, IP-final falls were found across only three clusters (cluster 1, high flat/rise; cluster 2, low-flat; and cluster 4, mid-fall). The flat mid-level tones were found in cluster 2 (low-flat) and cluster 3 (high-rise).

Figure 5 shows the potential source of interaction between the variety and question type (Sec. III A). While a rise was frequently used by CC speakers across question types, GS speakers' production seemed to be more affected by question types. For alternative questions, three CC

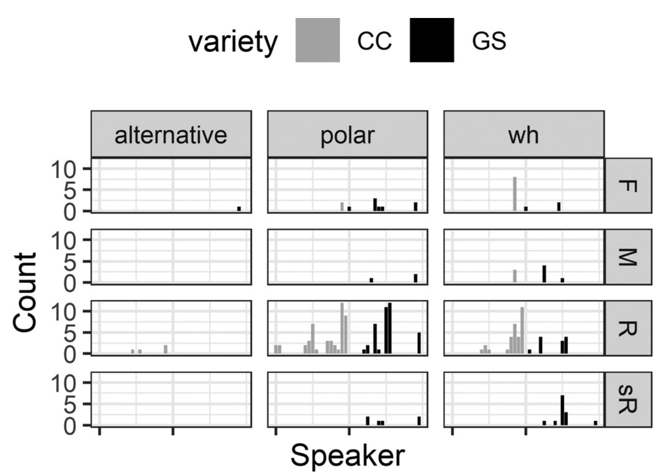

FIG. 5. Frequency of IP-final f_o contour shapes (fall, mid, rise and short rise) for varieties and question types. Each bar represents one speaker.

speakers produced them with a rise ($n = 4$), and there was only one token with a fall for GS.

For polar questions, for CC ($n = 49$), only one speaker produced a few tokens with a fall. Yet for GS ($n = 56$), the fall was observed across five speakers. For a flat f_o stretch at the mid-level, only two GS speakers produced it. A short rise was produced by five GS speakers. That is, while an IP-final rise was the most common for both varieties, half of GS speakers also produced IP-final falls. Moreover, the flat f_o stretch and short rise were observed only for GS.

For wh-questions (CC, $n = 42$; GS, $n = 56$), only one CC speaker produced a fall, while there were two GS speakers who produced a fall. The flat f_o stretch at the mid-level was produced by one CC speaker and by two GS speakers. Although a rise was most common for wh-questions across the varieties, it was produced by eight CC speakers and by four GS speakers. Finally, five GS speakers produced a short rise for wh-questions. For GS, more speakers produced IP-final syllables with a short rise compared to a rise.

IV. DISCUSSION

This study examined dyadic conversational speech data in CC and GS varieties of Korean. The CC variety does not have lexical pitch accent, stress, or tones, while the GS variety has lexical pitch accent. The traditional account suggests that, for questions, CC mainly uses a final rise and GS uses a fall. The first aim of the study was to test the use of contour clustering and random forest for exploring f_o contour variation in unscripted speech data, complemented with manual classification of IP-final f_o contour shapes. While clustering of f_o contours over the multisyllabic domain (i.e., IP-final two or three syllables) revealed grouping based on numerical similarities related to both f_o level and IP-final f_o pattern, manual analysis was akin to traditional prosodic analysis identifying the f_o pattern associated with the IP-final syllable regardless of the f_o level. The second aim was to assess the dialectal difference in question intonation between CC and GS. GS is known to frequently use falling intonation regardless of sentence modality, and its lexical pitch accent may affect scaling of IP-final f_o. Therefore, a wider variation was expected for GS compared to CC. The findings support the hypothesized interaction between variety and question type (alternative, polar, and wh-questions).

A. Clustering and manual analysis

The outcomes of semi-automatic clustering and manual analyses revealed different aspects; the clustering outcome showed an overall distribution and grouping of f_o contours mainly for their levels, and manual analysis focused on the f_o patterns regardless of their level.

Using the Contour Clustering application and built-in functions (Kaland, 2023a) first of all facilitated f_o data processing; it was useful for identifying data points to be discarded and examining the effect of consonantal perturbation on the estimated f_o contours (i.e., for cluster 5 in Fig. 2).

Second, it allowed us to explore the distribution and grouping of f_o contours over the multisyllabic domain (i.e., over IP-final two or three syllables). This process would have been labor-intensive for entirely manual analysis. As the GS variety was known for using an utterance-final f_o fall for questions, unlike CC, which mainly uses a rise, we sought a cluster of falling f_o contours. Informed by the output plots of clustering and random forest, having six clusters was deemed appropriate. The latter part of the averaged f_o contours in the six clusters were distinctive from each other for their level relative to the speaker's f_o range and the movement pattern (Fig. 2); the averaged contour for cluster 1 showed a high-flat/rise, cluster 2 showed a low-flat pattern, cluster 3 showed a high-rise with a steeper slope compared to cluster 1, cluster 4 showed a mid-fall, cluster 5 showed a high-fall-rise (but the fall portion was related to segmental perturbation), and cluster 6 showed a top-rise. The majority of f_o contours (110 out of 185) were in cluster 2; its averaged contour showed a low-flat pattern (Fig. 2), but the cluster showed that speakers mostly produced f_o contours at the mid-low range of their f_o. Random forest outputs revealed that variety (CC, GS), question type (alternative, polar, and wh-questions), and their interaction were important predictors as expected. The speaker variable also appeared to be an important predictor. The unbalanced speaker distribution, particularly for smaller clusters, is likely to be the cause. For instance, two speakers (out of 24) contributed to only one cluster, and 12 speakers contributed only to two clusters.

However, the clustering outputs presenting the averaged f_o contour and its "cluster," delimited by standard deviation, were not sufficient for detailed examination of cross-dialectal differences. For instance, an averaged high-rise contour for a cluster may represent a frequent contour, but the cluster can include flat or falling contours. The clustering outcome is influenced by frequency distribution in the dataset; an infrequent contour of a certain shape can form a cluster together with another set of contours that differ in their movement direction. Nonetheless, the clustering output highlighted differences in the f_o level relative to speakers' f_o range, which were expected in unscripted speech, while the f_o patterns near the right edge of the IP mostly seemed to include rises, falls, and flat stretches.

In the manual analysis, each of the f_o contours for the IP-final syllable was classified into a fall (F), a flat-stretch near the middle of the speaker's range (M), a rise (R), and a short rise (sR) (a rise followed by a dip in the latter part of the syllable) (Fig. 1). This manual approach was feasible given the size of the current dataset ($n = 185$). The rises and short rises were generally classified into clusters represented with a final rise (Fig. 4). However, the falls and flat-mid contours were spread across clusters, probably due to their low frequency. The apparent discrepancy between contour clustering outputs (Fig. 2), which categorized most contours in cluster 2 (low-flat, $n = 110$), and the manual analysis, which classified the majority of IP-final f_o patterns as rises ($n = 133$), require clarification. That is, users must note that the averaged contour in clustering outputs was statistically derived, with each cluster encompassing diverse contour shapes as indicated by standard deviation measures. The clustering outputs underscore the significance of speakers' f_o level in the numerical classification of contours over the multisyllabic domain; on the other hand, the movement direction (i.e., rise, fall, and flat) may have occurred in small magnitudes with variability in slope parameters.

For the speaker and question type predictors (Fig. 5), while the results did not provide strong evidence for GS speakers' tendency towards using IP-final f_o falls reported in the literature, distributional differences between the two varieties were nonetheless observed. A rise was common for both dialects, but GS speakers showed a high level of dispersion in f_o patterns compared to CC speakers, as further discussed in Sec. IV B. Further, a dialect-specific f_o shape was noted; the short rise, an f_o rise followed by a dip in the latter part of the syllable, was observed only for GS. The short rise could be interpreted as a multitonal boundary tone, HLH%, in K-ToBI (Jun, 2000). However, further research is required to clarify its function and potential phonetic difference from canonical realizations of the HLH% tone in Seoul Korean.

B. The interaction between variety and question type

Most questions in the present data were polar questions ($n = 105$, 58%) and wh-questions ($n = 75$, 41%). There were only a few tokens constituting alternative questions ($n = 5$, 3%). The frequent use of polar questions was also shown in Yoon's (2010) Korean conversational data collected from family members or friends over tea or dinner (polar, $n = 229$, 70%; wh, $n = 95$, 29%; alternative, $n = 2$, 1%).

The random forest analysis (Fig. 3) showed that the interaction between variety and question was an important predictor for the six-cluster output (Fig. 2). While the clustering output seemed to indicate a stronger preference for falls for GS compared to CC, the source of the interaction was not clear. That is, out of the six clusters, the averaged f_o contour for cluster 4 showed a fall, which was expected for GS. While more GS speakers indeed contributed to cluster 4 (three CC speakers, six GS speakers), only 16 f_o contours were in cluster 4. As the clustering was carried out for the multisyllabic domain, the interaction here could be interpreted such that speakers of the two varieties differed in their use of f_o level and its pattern across the question types.

When the frequency distribution of the variety and question type was examined (see III A and the OSF), both polar and wh-questions were observed across clusters 1, 2, 3, and 4 for both varieties. Cluster 5 ($n = 5$) included only polar questions from both varieties (CC, $n = 2$; GS, $n = 3$), but its f_o contour was perceptually a rise with its acoustic form influenced by consonantal perturbation. Cluster 6 ($n = 6$), showing a high-rise, included both a polar ($n = 1$) and a wh-question ($n = 1$) for CC and only polar ($n = 4$) questions for GS; but again, the number of data points is not sufficient for a meaningful interpretation.

The manual analysis results (Sec. III B) revealed that the IP-final rises were the most common for both varieties, followed by falls. Flat f_o stretches around the middle of the speaker's range were also observed. Short rises (i.e., a rise followed by a dip in the latter part of the syllable) were observed only for GS.

Dialectal differences were found in terms of distributional properties, GS with high internal dispersion (cf. Henriksen, 2013). CC speakers mostly used a rise, whereas GS speakers' productions included all four patterns. While IP-final rises were frequent for GS, half of the GS speakers also produced falls. The distributional difference seems to be the source of the interaction involving the variety and question type predictors. Notably, the findings indicate that dialectal differences reported in naturalistic settings may be driven by a subset of speakers. In the current analysis, three speakers from each dialectal group (representing 25% of each group) contributed a substantial proportion of the data analyzed (62% for CC, 46% for GS).

Although only a few alternative questions were observed in the present data (i.e., three CC speakers with a rise, $n = 4$, and only one token with a fall for GS), the distribution of the contour pattern indicates the dominance of a rise for CC. For polar questions, the IP-final rise was most common for both varieties. For CC ($n = 49$), only one speaker produced a few IPs with a final fall. For GS ($n = 56$), a rise was dominant, but a fall was also common, produced by five speakers. Only two GS speakers produced a flat f_o stretch at the mid-level, and only five GS speakers produced a short rise. For wh-questions (CC, $n = 42$; GS, $n = 56$), although a rise was most common for both varieties, it was produced by eight CC speakers and by only four GS speakers. Only one CC speaker but two GS speakers produced falls. The flat f_o stretch at the mid-level was produced by one CC speaker and two GS speakers. More GS speakers produced IP-final syllables with a short rise compared to a rise.

To recapitulate, the present findings do not strongly support the traditional account of GS intonation that the IP-final fall is dominant regardless of sentence modality. The discrepancy may be partly attributed to the observation for read speech in previous work and the use of conversational speech in the present corpus analysis. In laboratory-based investigations, speakers who have a strong regional identity are recorded reading materials designed to have dialectal features, and the experimenter may encourage speakers to use their dialect and may even discard unsatisfactory tokens. Furthermore, read speech is stylistically different from conversational speech. It is well known that speech style, such as the level of formality in the situation of language use (Labov, 1994), affects speakers' use of intonation (Henriksen, 2013). For instance, Bari Italian speakers tend to use rising-falling intonation in unscripted speech, but when reading a transcript of a dialogue they tend to append a rise, resulting in a rising-falling-rising intonation (Grice *et al.*, 1997). Similarly, Cruttenden (2007) reports that a Glaswegian English speaker produced a rising or rising-slumping intonation pattern, which is a characteristic of the

https://doi.org/10.1121/10.0037191

regional variety, as default for conversational speech, while she used a falling tune as default in read speech. For the present data, interlocutors knew each other well and their task was unscripted as well as more informal than the tasks that are employed in typical experimental settings.

The present findings have implications for cross-linguistic tendencies. The frequent use of a final rise to a high pitch in questions has been linked to the biological "frequency code" (Ohala, 1983; Gussenhoven, 2002). The assumption is that smaller larynxes produce a higher pitch compared to larger larynxes, and that this association may have played a role in the evolution of the distributional bias, expressing "uncertainty" with a high pitch in questions and "certainty" with a low pitch in asserting statements. However, this distributional bias interacts with linguistic structures; it has been argued that, cross-linguistically, polar questions are often marked with a rise, while wh-questions are commonly produced with a fall. The logic is that once the lexical cue, such as a wh-word is present, there is less need for exploiting the intonational cue (Frota, 2002; Kohler, 2004). The present findings are in line with this description. The present-day GS variety does not seem to form a clear exception (as in some African languages with falling "lax" endings) (Rialland, 2009) to the trend frequently associating polar questions with a rise. Speakers of CC, in which morphological cues for questions are often ambiguous, constantly preferred a rise across question types. On the other hand, speakers of GS, in which statements, wh-questions, and polar questions can be differentiated by sentence-final particles, produced more varied f_o contours, particularly for wh-questions. This may suggest that the availability of morphosyntactic cues to question marking increases the range of intonational forms from which speakers can select.

Finally, the role of lexical pitch accent requires further investigation using both acoustic and perceptual methods. The varying pitch accent location in GS utterances may have contributed to the wider variation in f_o patterns and the use of the high level found in GS. Furthermore, it is possible that the frequent falls in GS reported in the literature were also susceptible to the effect of lexical pitch accent. In some studies, any f_o pattern not showing a clear rise could have been classified as a fall. For instance, Lee (2007) reported that questions were characterized by a relatively high f_o and a rise, while statements tended to have a lower final f_o for GS. This may be relevant to the present finding that only two out of six GS speakers' f_o contours were grouped into the largest cluster 2 near the middle of the speakers' range. Most GS speakers in the present study produced high-level f_o for questions. However, some f_o tracks in Lee (2007) seemed to be subject to different interpretations. For instance, while the question in Fig. 6 ends in a high f_o, the rise is not salient. This boundary tone may be labelled as H% and interpreted as a rise, but listeners may also perceive a fall from the penultimate syllable, which carries the peak of the lexical pitch accent, interpreting the final tone as L%. However, if the question and the statement in Fig. 6 are paired, then probably the statement is more likely to be perceived as having a fall (L%) due to the steeper slope of the falling portion, although the f_o rises towards the end. Even for the statement, f_o does not fall to near the bottom of the speaker's f_o range, probably due to the presence of the lexical pitch accent.

In sum, for IP-final f_o patterns, while a rise was most common for both varieties across the question types, GS showed more dispersion compared to CC. The interaction between the variety and question type shown in the clustering output indicates potential cross-dialectal differences in the use of f_o level and shape over several syllables. Furthermore, descriptions of GS intonation in the literature may confound the local f_o movement and variation in the f_o level related to the placement of lexical pitch accent or other factors. Further investigation is required to clarify dialectal differences in the use of overall f_o level and how far speakers make use of global f_o, trends such as declination and final lowering (see Ladd, 2008) as a way of differentiating utterance modalities.

V. CONCLUSION

This study used unscripted conversational speech data to investigate differences in question intonation between two Korean varieties. The data represented a high-level of ecological validity with dialectal and between-speaker variation in an unevenly distributed dataset. Semi-automatic analysis, using agglomerative hierarchical clustering and a random forest analysis, was carried out for f_o contours over IP-final two or three syllables. These techniques were useful for data processing and for examining f_o variation. Subsequently, a manual analysis was carried out for IP-final syllables, classifying f_o patterns into a rise, a fall, a flat stretch, and a short rise (Fig. 1). The present approach can be extended to further examine the relationship between intonation and different utterance modalities, morpho-lexical properties, and speakers' stance (Kohler, 2004). This novel way of analyzing intonation may enable us to address the long-standing dilemma in the field, whereby researchers are forced to choose either the linguistic analysis overlooking phonetic details or the purely phonetic analysis neglecting the communicative and linguistic roles of intonation ('t Hart *et al.*, 1990, Sec. 1.1).

Overall, the results here support distributional approaches, showing cross-dialectal differences in *preferences* for IP-final

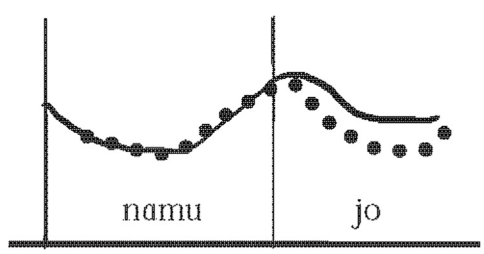

FIG. 6. Schematic representation of question (thick line) and statement (dotted line) intonation in North GS for /nɑmujo/ ["(it is) a tree." or "(is it) a tree?" *tree* + POL] based on Lee's (2007) data.

f_o contours, which might be driven by a subset of speakers (Savino, 2012; Henriksen, 2013) rather than a one-to-one mapping between a particular variety and its default pattern. The variety and question type predictors interacted; the interaction was an important predictor for the clustering outputs, which classified f_o contours differing in their level and movement patterns (rise, flat, and fall). For IP-final syllables, the most frequent f_o contour shape was a rise; this was followed in frequency by a fall for both varieties. However, the GS speakers' choice of f_o pattern showed more dispersion. While a rise was commonly used across the two question types for CC, the question type seemed to affect the contour selected by GS speakers: both a rise and a fall were commonly used in polar questions, and a short rise was more common than a regular rise in wh-questions.

ACKNOWLEDGMENTS

This work was supported by the Academy of Korean Studies grant (Grant No. AKS-2023-R040) and the German Research Foundation (DFG), Project-ID 281511265, SFB-1252 "Prominence in Language." We would like to thank Jaehyung Park, Sunran Shin, and Hyunjung So for their assistance with data annotation and anonymous reviewers for their constructive feedback on this work.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts of interest to disclose.

Ethics Approval

The present study used speech materials available as open data for public use. Therefore, obtaining informed consent from participants or ethics approval was not required.

DATA AVAILABILITY

The data that support the findings of this study are openly available in OSF at http://doi.org/10.17605/OSF.IO/5XNF2.

[1]The particle type affects speakers' choice of intonation (Yun, 2023). The particle type effect is not reported in the present analysis because the data were not sufficient for a meaningful analysis. Analysis results incorporating the particles are available on the OSF repository (supplementary materials B).

[2]GAMM can be used in a complementary manner, making it possible to visualize averaged f_o contours with standard deviations and differences between contours (see Kaland *et al.*, 2023; Steffman *et al.*, 2024). However, reducing f_o contours to one representative contour for each level of predictors was deemed inappropriate for the present dataset, with a high degree of between- and within-speaker variation in both the frequency and shape of f_o contours. Variation of f_o contour shapes could lead to misleading representations of intonation by, for instance, showing a flat f_o contour averaged across rises and falls.

[3]Abbreviations used for gloss: INT, intimate; POL, polite; Q, question.

Beckman, M. E. (**1997**). "A typology of spontaneous speech," in *Computing Prosody*, edited by Y. Sagisaka, N. Campbell, and N. Higuchi (Springer, New York), pp. 7–26.

Boersma, P., and Weenink, D. (**2023**). "Praat: Doing phonetics by computer (version 6.1.16) [computer program]," http://www.praat.org/ (Last viewed September 1, 2023).

Breiman, L. (**2001**). "Random forests," Mach. Learn. **45**, 5–32.

Brown, L., and Yeon, J. (**2015**). "Varieties of contemporary Korean," in *The Handbook of Korean Linguistics*, edited by L. Brown and J. Yeon (John Wiley & Sons, Inc., New York), pp. 459–476.

Cangemi, F., Grice, M., Jeon, H.-S., and Setter, J. (**2023**). "Contrast or context, that is the question," in *Proceedings of the 20th International Congress of Phonetic Sciences*, August 7–11, Prague, Czech Republic, pp. 1360–1364.

Cole, J., Steffman, J., Shattuck-Hufnagel, S., and Tilsen, S. (**2023**). "Hierarchical distinctions in the production and perception of nuclear tunes in American English," Lab. Phon **14**(1), 1–51.

Cruttenden, A. (**1997**). *Intonation*, 2nd ed. (Cambridge University Press, Cambridge, UK).

Cruttenden, A. (**2007**). "Intonational diglossia: A case study of Glasgow," J. Int. Phonetic Assoc. **37**, 257–274.

Frota, S. (**2002**). "Nuclear falls and rises in European Portuguese: A phonological analysis of declarative and question intonation," Probus **14**, 113–146.

Grice, M., Ladd, D. R., and Arvaniti, A. (**2000**). "On the place of phrase accents in intonational phonology," Phonology **17**, 143–185.

Grice, M., Savino, M., and Refice, M. (**1997**). "The intonation of questions in Bari Italian: Do speakers replicate their spontaneous speech when reading?," Phonus **3**, 1–7.

Gussenhoven, C. (**2002**). "Intonation and interpretation: Phonetics and phonology," in *Proceedings of Speech Prosody 2002*, April 11–13, Aix-en-Provence, France, pp. 47–57.

Han, S.-H., and Oh, M. (**1999**). "Hangugeo eokyangguui gyeonggyeton (The boundary tones in Korean Intonational Phrases)," Speech Sci. **5**(2), 109–129.

Henriksen, N. (**2013**). "Style, prosodic variation, and the social meaning of intonation," J. Int. Phonetic Assoc **43**(2), 153–193.

Hirschberg, J., Beňuš, Š., Gravano, A., and Levitan, R. (**2020**). "Prosody in discourse and speaker state," in *The Oxford Handbook of Language Prosody*, edited by C. Gussenhoven and A. Chen (Oxford University Press, Oxford, UK), pp. 467–476.

Hwang, H. K. (**2006**). "Intonation Patterns of Wh-interrogatives in South Kyungsang Korean and Fukuoka Japanese," Eoneohak **45**, 39–59.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (**2013**). *An Introduction to Statistical Learning with Applications in R* (Springer-Verlag, New York).

Jeon, H.-S. (**2015**). "Prosody," in *Handbook of Korean Linguistics*, edited by L. Brown and J. Yeon (Wiley-Blackwell, New York), pp. 41–58.

Jeon, H.-S., and Heinrich, A. (**2022**). "Perception of pitch height and prominence by old and young listeners," in *Proceedings of Speech Prosody 2022*, May 23–26, Lisbon, Portugal, pp. 654–658.

Jeon, H.-S., Kaland, C., and Grice, M. (**2024**). "Cluster analysis of Korean IP-final intonation," in *Proceedings of Speech Prosody 2024*, July 2–5, Leiden, the Netherlands, pp. 1025–1029.

Jeon, H.-S., and Nolan, F. (**2013**). "The role of pitch and timing cues in the perception of phrasal grouping in Seoul Korean," J. Acoust. Soc. Am. **133**, 3039–3049.

Jeon, L. (**2013**). "Drawing boundaries and revealing language attitudes: Mapping perceptions of dialects in Korea," Master's thesis, University of North Texas, Denton, TX.

Jun, S.-A. (**1998**). "The accentual phrase in the Korean prosodic hierarchy," Phonology **15**(2), 189–226.

Jun, S.-A. (**2000**). "K-ToBI (Korean ToBI) labelling conventions (ver. 3.1)," UCLA Work. Pap. Phonetics **99**, 149–173.

Kaland, C. (**2023a**). "Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours," J. Int. Phonetic Assoc. **53**(1), 159–188.

Kaland, C. (**2023b**). "Intonation contour similarity: f0 representations and distance measures compared to human perception in two languages," J. Acoust. Soc. Am. **154**, 95–107.

Kaland, C., and Ellison, T. (**2023**). "Evaluating cluster analysis on f0 contours: An information theoretic approach on three languages," in *Proceedings of the 20th International Congress of Phonetic Sciences*, August 7–11, Prague, Czech Republic, pp. 1360–1364.

Kaland, C., and Grice, M. (**2024**). "Exploring and explaining variation in phrase-final f0 movements in spontaneous Papuan Malay," Phonetica **81** (3), 321–349.

Kaland, C., Kluge, A., and Van Heuven, V. J. (**2021**). "Lexical analyses of the function and phonology of Papuan Malay word stress," Phonetica **78** (2), 141–168.

Kaland, C., Steffman, J., and Cole, J. (**2024**). "K-means and hierarchical clustering of f0 contours," in *Proceedings of Interspeech 2024*, September 1–5, Kos, Greece, pp. 1520–1524.

Kaland, C., Swerts, M., and Himmelmann, N. P. (**2023**). "Red and blue bananas: Time-series f0 analysis of contrastively focused noun phrases in Papuan Malay and Dutch," J. Phonet. **96**, 101200.

Kaufman, L., and Rousseeuw, P. J. (**1990**). *Finding Groups in Data: An Introduction to Cluster Analysis* (Wiley, Hoboken, New Jersey).

Kim, H. R. S. (**2010**). "A high boundary tone as a resource for a social action: The Korean sentence-ender -ta," J. Pragmat. **42**, 3055–3077.

Kim, M.-S. (**2003**). "A study of the correlations between intonation and word-contraction in Gyungsang-do dialects," Eomunhak **87**, 111–130.

Kim, M.-S. (**2015**). "A distinct declarative question design in Korean conversation: An examination of turn-final ko questions," J. Pragmat. **79**, 60–78.

Kohler, K. (**2004**). "Pragmatic and attitudinal meanings of pitch patterns in German syntactically marked questions," in *Working Papers of the Kiel Institute of Phonetics and Digital Speech Processing (AIPUK)*, Vol. 35a, pp. 125–142.

Koutsombogera, M., and Vogel, C. (**2018**). "Modeling collaborative multimodal behavior in group dialogues: The MULTISIMO corpus," In *Proceedings of the 11th Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, May 7–12, Miyazaki, Japan, pp. 2945–2951.

Labov, W. (**1994**). *Principles of Linguistic Change: Internal Factors* (Blackwell, Oxford, UK).

Ladd, D. R. (**2008**). *Intonational Phonology*, 2nd ed. (Cambridge University Press, Cambridge, UK).

Ladd, D. R. (**2022**). "The trouble with ToBI," in *Prosodic Theory and Practice*, edited by J. Barnes and S. Shattuck-Hufnagel (The MIT Press, Cambridge, MA), pp. 247–258.

Lee, B.-W. (**1988**). "A comparative study of the central, Gyeongnam, and Jeonnam dialects," Urimal Yeonku **8**, 1–62.

Lee, H. (**2020**). "Cross-generational acoustic comparisons of tonal Kyungsang Korean stops," J. Acoust. Soc. Am **148**(2), EL172–EL178.

Lee, H., and Jongman, A. (**2015**). "Acoustic evidence for diachronic sound change in Korean prosody: A comparative study of the Seoul and South Kyungsang dialects," J. Phonet. **50**, 15–33.

Lee, H., and Zhang, J. (**2014**). "The nominal pitch accent system of South Kyungsang Korean," J. East Asian Linguist. **23**(1), 71–111.

Lee, H.-S. (**2007**). "Interrogative intonation in North Kyungsang Korean: Language-specificity and universality of acoustic and perceptual cues," Work. Papers Cornell Phonetics Lab. **16**, 57–100.

Lee, H.-Y. (**1999**). "An acoustic phonetic study of Korean nuclear tones," Malsori **37**, 25–39.

Ohala, J. J. (**1983**). "Cross-language use of pitch: An ethological view," Phonetica **40**, 1–18.

Park, M.-J. (**2012**). *The Meaning of Korean Prosodic Boundary Tones* (Brill, Leiden, Netherlands), pp. 1–277.

Pierrehumbert, J. (**1980**). "The phonetics and phonology of English intonation," Ph.D. thesis, MIT, Cambridge, MA.

Puggaard-Rode, R. (**2024**). "praatpicture: "Praat Picture" Style Plots of Acoustic Data," In CRAN: Contributed Packages. The R Foundation.

R Core Team (**2024**). "R: A language and environment for statistical computing (version 4. 4. 0) [computer program]," http://www.R-project.org/ (Last viewed June 1, 2024).

Rialland, A. (**2009**). "The African lax question prosody: Its realisation and geographical distribution," Lingua **119**, 928–949.

Rissanen, J. (**1978**). "Modeling by shortest data description," Automatica **14**(5), 465–471.

Rose, P. (**1987**). "Considerations in the normalisation of the fundamental frequency of linguistic tone," Speech Commun. **6**, 343–352.

R Studio Team (**2023**). "RStudio: Integrated Development for R (Build 561) [computer program]," http://www.rstudio.com/ (Last viewed June 1, 2024).

Saltlux (**2021**). "Korean dialect data," https://aihub.or.kr/ (Last viewed May 31, 2023).

Savino, M. (**2012**). "The intonation of polar questions in Italian: Where is the rise?," J. Int. Phonetic Assoc **42**(1), 23–48.

Scitovski, R., Sabo, K., Martínez-Álvarez, F., and Ungar, Š. (**2021**). *Cluster Analysis and Applications* (Springer International Publishing, Cham, Switzerland).

Silverman, B. W. (**1986**). *Density Estimation for Statistics and Data Analysis* (Chapman & Hall, London, UK), pp. 1–186.

Sohn, H.-M. (**1999**). *The Korean Language* (Cambridge University Press, Cambridge, UK).

Speiser, J. L., Miller, M. E., Tooze, J., and Ip, E. (**2019**). "A comparison of random forest variable selection methods for classification prediction modeling," Expert Syst. Appl. **134**, 93–101.

Steffman, J., Cole, J., and Shattuck-Hufnagel, S. (**2024**). "Intonational categories and continua in American English rising nuclear tunes," J. Phonet. **104**, 101310.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. (**2008**). "Conditional variable importance for random forests," BMC Bioinf. **9**, 307.

Strobl, C., Malley, J., and Tutz, G. (**2009**). "An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests," Psychol. Methods **14**, 323–348.

't Hart, J., Collier, R., and Cohen, A. (**1990**). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody* (Cambridge University Press, Cambridge, UK).

Turk, A., Nakai, S., and Sugahara, M. (**2006**). "Acoustic segment durations in prosodic research: A practical guide," in *Methods in Empirical Prosody Research*, edited by S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schließer (De Gruyter, Berlin, New York), pp. 1–28.

Wagner, P., Trouvain, J., and Zimmerer, F. (**2015**). "In defense of stylistic diversity in speech research," J. Phonet. **48**, 1–12.

Wright, M. N., and Ziegler, A. (**2017**). "ranger: A fast implementation of random forests for high dimensional data in C++ and R," J. Stat. Soft. **77**, 1–17.

Xu, Y. (**2011**). "Speech prosody: A methodological review," J. Speech Sci. **1**, 85–115.

Yoon, K.-E. (**2010**). "Questions and responses in Korean conversation," J. Pragmat **42**(10), 2782–2798.

Yoon, T.-J. (**2023**). "Korean forced aligner [online application]," https://tutorial.tyoon.net/ (Last viewed September 13, 2023).

Yun, J. (**2023**). "Wh-intonation in Korean," in *Japanese/Korean Linguistics 30*, edited by S. Williamson, A. A. Babayode-Lawal, L. Bosman, N. Chan, S. Cho, I. Fong, and K. Holubowsky (CSLI Publications, Stanford), pp. 45–60.