Robustness of F0 ratio as a diagnostic: comparing creaky voice in Danish and Seoul Korean

Watkins, M.; Puggaard-Rode, R.; Boersma, P.; Hamann, S.

DOI
10.21437/Interspeech.2025-1949

Publication date
2025

Document Version
Final published version

Published in
Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH

License
Other

Interspeech 2025
17-21 August 2025, Rotterdam, The Netherlands

Robustness of F0 Ratio as a Diagnostic: Comparing Creaky Voice in Danish and Seoul Korean

Michaela Watkins[1], *Rasmus Puggaard-Rode*[2], *Paul Boersma*[1], *Silke Hamann*[1]

[1]University of Amsterdam, The Netherlands
[2]Institute for Phonetics and Speech Processing, LMU Munich, Germany

m.m.watkins@uva.nl, r.puggaard@phonetik.uni-muenchen.de, paul.boersma@uva.nl, silke.hamann@uva.nl

Abstract

We present an exploratory analysis of F0 ratio, a proposed method to pick up octave jumps in the speech signal. Such jumps, often considered errors, are possibly indicative of the presence of creaky voice. This paper focuses on co-intrinsic voice quality in Seoul Korean fortis stops, building on previous data, and on the Danish contrastive voice quality stød. The results suggest that in Seoul Korean F0 ratio captures a clear jump upwards indicating a modality switch from creaky to modal voice, with a gender difference observed. This suggests that pitch jumps are not necessarily erroneous but may reflect systematic cues to phonological contrasts cued with creak. In Danish, F0 ratio captures a rising intonational contour for non-stød tokens and is able to categorise between stød and non-stød tokens with high accuracy, although this leaves open the question whether F0 ratio captures modality shifts in Danish, or rather a combination of modality shift and pitch contour differences.
Index Terms: Phonetics, pitch tracking, speech processing, voice quality

1. Introduction

Pitch tracking algorithms can be used for different goals [1], one of them being the detection of aperiodicity in the speech wave caused by irregular vocal fold activity. Such irregular or aperiodic waveforms naturally occur due to the nature of the human voice, and can have linguistic and paralinguistic function. Creaky voice, an umbrella term for different types of aperiodic and/or irregular voice qualities, is one cause of such irregularities in the speech signal, with prototypical creak being described as having an irregular F0 with low rates of vocal fold vibration [2].

Creak, but also other non-modal voice qualities, can result in pitch jumps in the tracking algorithm, i.e., sudden octave changes upwards or downwards. "Real" pitch jumps in algorithms (i.e., not errors) have previously been acknowledged in the literature (e.g. [3]), and often co-occur with aperiodic creak (although not all creak necessitates a pitch jump being present).

In recent years it has become common to consider pitch jumps as errors in the pitch tracking algorithm, with notable new discussions on how to "smooth" pitch contours to remove such jumps (e.g. [4]) or to fully remove tokens exhibiting jumps before data analysis begins, either through manual inspection of the speech signal or omission of any tokens that appear "audibly creaky" [5, 6]. Other studies, particularly those concerning Korean, have mentioned "unstable" pitch contours and either fully removed such tokens, sometimes leading to a large exclusion of data (e.g. [7, 8]), or have measured pitch further into the vowel to avoid any possibility of pitch jumps in the first part of the vowel [9]. Often such jumps upwards or downwards are explicitly labeled as errors, and this is provided as justification for removal (e.g. [6, 9, 8, 10]). Thus there seems to be a problem with data exclusion of any token that contains pitch jumps with this now being extended to anything that "sounds creaky", regardless of whether or not pitch jumps occur within that token. This is especially problematic given the findings in [11] that suggest that female speakers of Seoul Korean (henceforth: SK) exhibit pitch jumps more than 20% of the time in their fortis stops, possibly indicating a gender exclusion bias in studies up-to-date. [11] therefore propose that pitch jumps, rather than being errors of the pitch-tracking algorithm, could in fact be systematic features of the language in question that should not be removed before thorough investigation.

Some pitch tracking algorithms disallow general octave jumps by default, such as Praat's raw autocorrelation [12, 13] which punishes all jumps equally. Others explicitly allow for octave jumps, while at the same time punishing other forms of jumps [3]. In this paper we focus on the aforementioned raw autocorrelation algorithm, and we can account for octave jumps that are present due to voice quality changes [14] (language-specific examples include jumps found in Korean as in [8], or period doubling in Mandarin, as in [10]) by changing the base settings. In short, pitch jumps can be retained in Praat's pitch tracking algorithms by disabling the octave jump penalty (i.e., by setting "octave jump cost" to zero, for further details c.f. [11]). By doing so, [11] could observe that in a selection of SK fortis tokens there was a clear jump upwards roughly 1/3 of the way into the vowel from aperiodic to modal voice that appeared systematic (Fig. 1). These pitch jumps in SK were then quantified as the ratio between F0 in the first third and the final third of the vowel (henceforth: F0 ratio). By using F0 ratio, the creaky tokens that demonstrated pitch jumps could be included in their analysis, and a clear difference in F0 ratio was observed between aspirated and fortis tokens. [11] propose that F0 ratio could be used as a general diagnostic for differences in voice quality, namely the switch from modal to creaky voice and vice versa. Their preliminary research, however, was limited to a small group of female speakers in a single language (SK).

The aim of the current paper is to assess the claims made in [11] by evaluating a wider dataset of SK as well as a language with a different phonological contrast where creak plays a more prominent role, namely stød in Danish. In addition, the current research tests how well F0 ratio can distinguish between stød and non-stød tokens in Danish, and between fortis versus aspirated tokens in SK. To do so, the paper employs linear discriminant analysis (LDA) to obtain prediction accuracy measures for each group.

Our hypotheses can be summarised as follows:

- The larger set of SK data (including male speakers) is ex-

10.21437/Interspeech.2025-1949

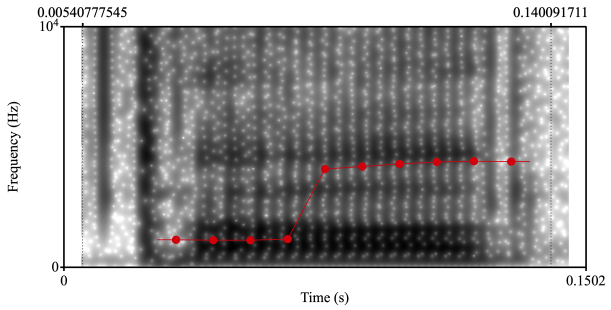

Figure 1: *Token* /t*a/ *of a female SK speaker showing an upwards pitch jump in the post-fortis vowel.*

pected to continue to show a clear switch from creaky to modal voice, and F0 ratio is expected to remain a consistent measure for this.

- For Danish, F0 ratio is expected to capture the contrast between stød and non-stød tokens, though a change in how to calculate F0 ratio might be necessary due to a different domain (the syllable) of the pitch jump than in SK (the vowel).
- The LDA is expected to have a higher prediction accuracy for Danish compared to SK because Danish stød is phonologised, while in SK creak is considered a secondary acoustic cue as it only occurs with a subset of the fortis stops.

2. Language Backgrounds

2.1. Seoul Korean

SK is a member of the largely extinct Koreanic family, and bears a three-way laryngeal contrast in stops and affricates between lenis, fortis, and aspirated, all of which are voiceless in phrase-initial position. How the three are contrasted is still in discussion, with voice quality being a lesser-researched area. Here we focus only on the stops. F0 and VOT appear to play a strong role for two of the three categories (lenis vs. aspirated), but the third category (fortis) has largely been excluded from previous research, as it is thought to be exclusively distinguished by VOT. However, fortis stops appear to be accompanied by creaky voice on the following vowel in certain contexts [15]: how frequently this creaky voice occurs is unclear. As mentioned in the introduction, creaky voice in SK (often also called "glottalization") has been actively avoided, and it has largely become standard to exclude or adjust creaky tokens where measures are taken before analysis begins with statements such as "target attainment F0 values were taken from the vowel duration whose glottalized portions were excluded in order to avoid acquiring low F0 values driven by glottalization" [9, p. 91]. This would exclude tokens that show the shift found in [11], where a low creaky F0 contour jumps upwards to modal roughly 1/3 of the way into the vowel (Fig. 1).

Here we compare fortis and aspirated stops, both thought to be phonologically marked for laryngeal features, but while fortis is marked for constricted glottis correlated with a "tense", "pressed" or creaky voice quality, aspirated are marked for spread glottis, and are produced with full aspiration and often a breathy voice in the following vowel [16, 17]. Creaky voice quality is seemingly not (yet) phonologised in SK, as not all fortis stop tokens contain creaky voice and/or pitch jumps.

2.2. Danish

Danish is a West-Germanic language which contrasts two different voice qualities in stressed syllables which have either long vowels or sonorant codas. The levels of the contrast are typically referred to as "stød" and "non-stød". There are several cues to the contrast and these cues differ in different varieties of Danish [18]. In the standard variety as spoken around Copenhagen, the perhaps most salient cue is that non-stød syllables are realized with modal voice whereas stød syllables are realized with creaky voice [19]. Stød syllables are not creaky throughout, but are generally recognized as having two distinct phases: in the first phase, they are modal, typically accompanied by local pitch raising immediately at the syllabic onset, and in the second phase, they are creaky [20, 21]. As a result, the creaky phase is typically quite short. Pitch jumps also occur within stød tokens, although the location is not as consistent as with SK.

There is an extensive body of literature on the phonetics of stød (see [22] for an overview). While creaky voice is integral to the contrast and there is a body of work that aims to quantify it using various measures of voice quality [23, 24], the corresponding pitch jumps are usually described as pitch tracking errors.

3. Analysis

As discussed above, both languages have some kind of creaky voice, although the presence of it differs from contrastive (Danish) to non-contrastive (SK). In both languages, pitch jumps have typically been labelled as errors.

For both languages, a linear discriminant analysis (LDA) was used to test how well category prediction could be made on the basis of F0 ratio. LDA was chosen since we are examining auditory cues for a phonation-type contrast, it makes sense to look at what percentage of tokens can be classified correctly from those cues, given a simple classification algorithm.

All acoustic analyses described below were conducted in Praat [12] using raw autocorrelation [13].

3.1. Seoul Korean Analysis

All SK data were taken from the *Seoul Corpus* [25]. All 40 speakers were included, divided evenly by gender. 933 fortis and 643 aspirated stops were extracted. The procedure described in [11] was maintained for consistency: all word-initial fortis plosives followed by an /a/ vowel were extracted, and F0 ratio (dividing the F0 of the first by that of the last third) calculated. Pitch floor and ceiling were adjusted for male speakers at 50 and 300 Hz respectively, and for female speakers at 75 and 500 Hz. Manual inspection confirmed the original claim that speakers produced a jump (i.e. an F0 ratio of over 1.7 for doubling and under 0.6 for halving, consistent with [11]) roughly 1/3 of the way into the vowel (cf. Fig. 1). No noticeable intonation rise was observed in the target vowel at any point during the investigation. Aspirated stops were also extracted in the same way, albeit with a higher pitch floor and ceiling to accommodate for a higher F0 (75–400 Hz for males, 100–500 Hz for female speakers). Summary statistics (Left panel, Fig. 2) indicated that for fortis stops female speakers produced jumps more often than male speakers: 36.5% of fortis tokens for female compared to 13.9% fortis tokens for male speakers showed F0 jumps. Pitch doubling was far more common than pitch halving, at 33% for female and 11.2% for male tokens. For aspirated stops, only 4.2% of all tokens showed a jump, with 7% of female tokens

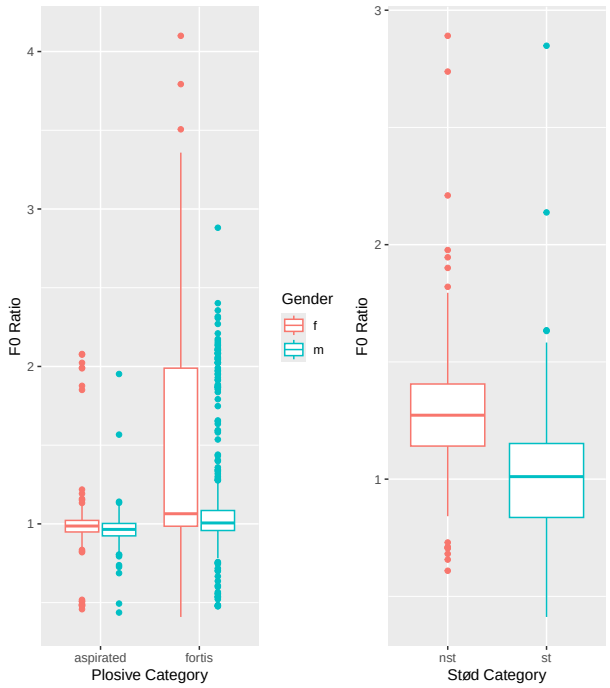

Figure 2: *Summary statistics of SK (left) and Danish (right) in terms of category distribution.*

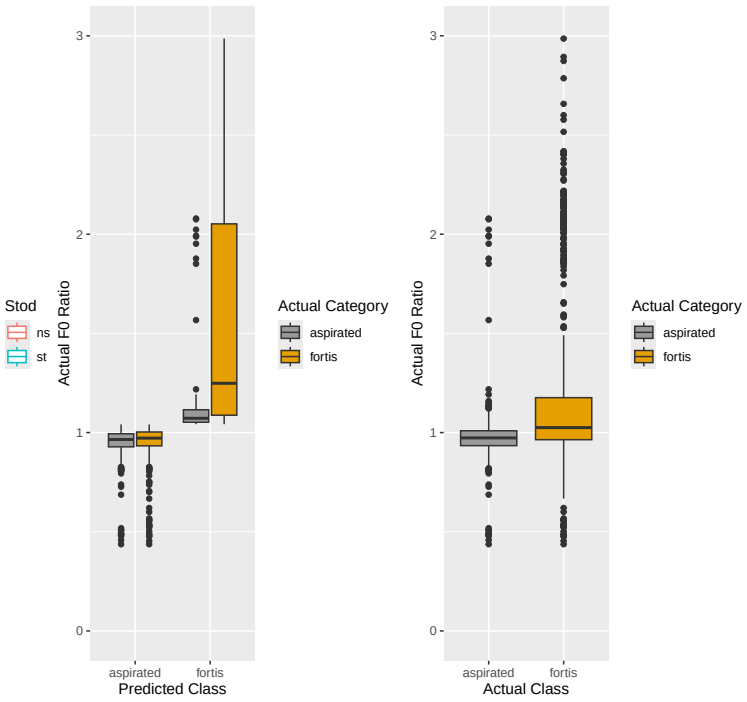

Figure 3: *Korean LDA results with predicted category vs actual ratio (left) compared to actual category and actual F0 ratio (right).*

compared to 1.7% of male tokens showing jumps. To assess further, the MASS package [26] in RStudio [27] was used to conduct an LDA. A random sample of fortis tokens was taken to match the number of aspirated tokens exactly (643 each). Based on F0 ratio alone, a prediction accuracy of 64% was produced (Fig. 3). As can be seen in Fig. 3, the category prediction switch was made at an F0 ratio of just over 1.

A linear regression model was fitted using the lme4 package [28] in RStudio [27]. The model assessed F0 ratio based on category (fortis or aspirated), and modelled the interaction effect between gender and category. By-speaker random intercepts were included as well as by-speaker random slopes for category, speaker, and their interaction:

$$f0Ratio \sim \ category * gender + (category \mid speaker)$$

Significant main effects were found for category, with fortis stops produced on average with a 0.27 higher F0 ratio than aspirated stops (CI = 0.21 .. 0.33, t = 8.886, $p = 7.6 \cdot 10^{-11}$), as well as for gender, with female speakers producing on average a 0.17 higher F0 ratio than male speakers (CI = 0.11 .. 0.24, t = 5.438, $p = 3.9 \cdot 10^{-6}$). These main effects are qualified by a significant interaction between gender and category: the female-minus-male difference in F0 ratio is greater for fortis than aspirated stops, by an estimated 0.27 (CI = 0.15 .. 0.40, t = 4.333, p = 0.00010).

3.2. Danish Analysis

The Danish data come from 12 speakers of the Copenhagen variety of Modern Standard Danish (10 female), who were recorded in late 2023 for a forthcoming study on co-intrinsic pitch. These materials are also suitable for our purposes here since they systematically vary stød and non-stød. Recordings were made in sound-attenuated recording booths at the universities of Copenhagen and Aarhus using respectively a Sennheiser MKH40 microphone and an AKG P420 microphone through a Zoom H5. The speakers each read 137 sentences aloud that were prompted on a computer screen using SpeechRecorder [29]. The sentences were alternative questions of the type *Er det **dine** eller er det mine?* 'Are they yours or are they mine?', where the first part of the sentence was designed to elicit a high intonational pitch target, and the second part a low target. Recordings were force-aligned using the DanFA module of the Autophon tool [30] and manually corrected. For this study, we use a subset of 60 suitable sentences and focus only on syllables with high pitch targets.

F0 ratio was calculated in a similar way as for the SK data, using the first third divided by final third of the syllable (instead of the vowel). Tokens containing stød produced a lower F0 ratio than non-stød tokens (Right panel in Fig. 2). A mixed-effects model again in RStudio using the lme4 package [28] was created, in the same style as SK with by-speaker random intercepts and by-speaker random slopes for stød:

$$f0Ratio \sim \ stød + (stød \mid speaker)$$

No gender effect was included as there were not enough male speakers to model this effect. A significant effect of stød type on F0 ratio was observed, with non-stød tokens producing a ratio 0.35 higher on average compared to stød tokens (CI = 0.28 .. 0.43, t = 9.438, p = 0.00000141). The LDA analysis was performed using the stød classes as the classifying variable and produced an accuracy of 79.3% for predicting stød vs non-stød, better than the SK data. As can be seen in Fig. 5, the model predicted non-stød values above roughly 1.2, although there is

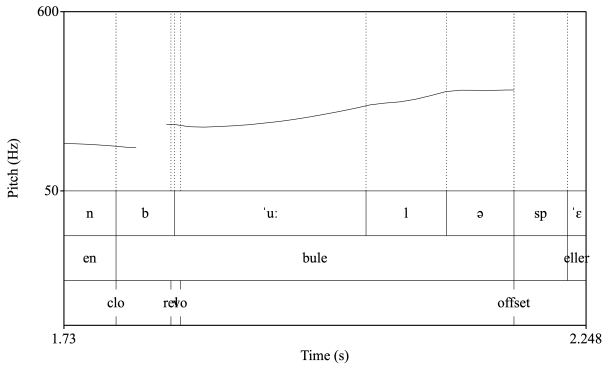

Figure 4: *Non-stød token of the word "bule" demonstrating a rising pitch contour.*

still overlap between the two. Therefore the ratio provided a difference between the two categories, but there is not an absolute distinction using F0 ratio alone.

4. Discussion

As seen in [11], SK fortis stops appear to show a systematic pattern of pitch jumps, that we here find to be stronger for female speakers. This supports the point made in the introduction that removing such jumps could be detrimental to the study of SK. The measure of F0 ratio was shown to be a working diagnostic for pitch jumps, however, it does not demonstrate where the jumps occur, and for SK it does not imply that fortis stops *not* containing jumps are modal; rather it appears good at detecting modality switches from creaky to modal voice. Although this implies that creak is present (somewhere) in that token, it does not entail that all creaky tokens have jumps. Though not explicitly stated, it is therefore likely that SK female speakers are more at risk than male speakers of being excluded from studies involving fortis stops.

In terms of categorisation results, F0 ratio appeared to struggle to distinguish between aspirated and fortis tokens for SK, although it did perform above chance level. However, the reader should be aware that little was done to "prepare" the data – that is to say, no account was taken of sentence structure, speech rate, word type, etc. The corpus used is conversational speech, which makes this even more striking. With a more structured dataset and more controlled variables, we expect F0 ratio to perform even better at identifying the difference between the two stop categories. To fully understand the performance of the present model, introduction of the third stop type, lenis, which was not included here, would be required.

Danish stød appeared to perform better, which is most likely due to the fact that this is a phonologised category in the language, but could also be influenced by the fact that the dataset is more controlled (i.e. matched for word, length, and sentence position). However, it seems possible that the LDA categorisation model focused on an intonation rise in non-stød tokens (c.f. Fig. 4) rather than pitch jumps: future research could provide a systematic overview of where these jumps occur, and then a more complex ratio could be employed to account for jump locations.

The results therefore indicate that F0 ratio appears to be useful as a general measure for inferring changes in the pitch contour, be that an intonational rise or pitch jumps. It functions

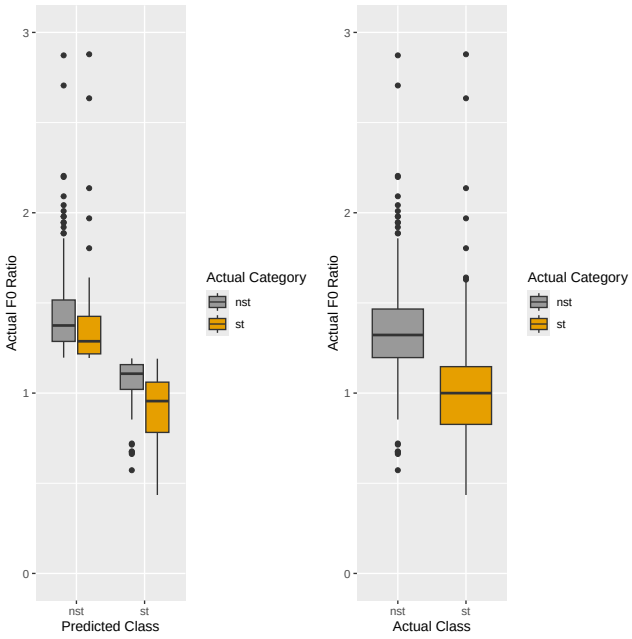

Figure 5: *Danish LDA results with predicted category vs actual ratio (left) compared to actual category and actual F0 ratio (right).*

well for both conversational as well as structured datasets, and with further adjustments reliant on language-specific knowledge it is possible that the accuracy of pitch jump identification and the subsequent creaky voice will increase.

4.1. Considerations and Future Research

This paper has looked at F0-based cues for the phonation-type contrasts at hand. Future research will have to include other types of cues, perhaps spectral ones such as H1–H2 (the strength of the first harmonic with respect to the second harmonic [31, 32]) and combine the results to compare with the output of F0 ratio. This is particularly relevant given the fact that not all tokens that are aperiodic contain jumps; this needs to be confirmed with spectral analysis. The first consideration would be to assess datasets that are equally structured: future research could evaluate if SK improves in accuracy with a more systematic (non-conversational) database that controls for word and word-frequency. This is in combination with the suggestion to include the third category, lenis.

Additionally, we did not adjust F0 ratio beyond a simple division between the first third and the final third of the vowel or the syllable. It is possible that accuracy could be improved for languages where multiple jumps occur if the ratio is adjusted; a suggestion is to consider this possibility for Danish stød which can contain multiple pitch jumps, and see if changing the ratio to include language-specific patterns assists in category classification.

5. References

[1] P. Boersma, "Acoustic analysis," in *Research methods in linguistics*, R. Podesva and D. Sharma, Eds. Cambridge University Press, 2013, pp. 375–396.

[2] P. Keating, M. Garellek, and J. Kreiman, "Acoustic properties of

different kinds of creaky voice," in *18th International Congress of Phonetic Sciences*, 2015.

[3] D. Talkin, "A robust algorithm for pitch tracking (RAPT)," in *Speech Coding and Synthesis*, W. Kleijn and K. Paliwal, Eds. Elsevier Science, 1995, ch. 14, pp. 495–518.

[4] T. Sostarics and J. Cole, "Pitchmendr: A tool for the diagnosis and treatment of F0 irregularities," in *Proc. Speech Prosody*, Leiden, the Netherlands, 2024, pp. 289–293.

[5] X. Ren and P. Mok, "Tonogenesis in Seoul Korean and L3 production of Korean stops by Cantonese-English bilinguals," in *International Speech Communication Association*, 2018, pp. 192–196.

[6] M. Broersma, "Korean lenis, fortis, and aspirated stops: Effect of place of articulation on acoustic realization," in *Proc. Interspeech 2010*, Makuhari, Japan, 2010, pp. 941–944.

[7] H.-G. Byun, "F0 as a primary cue for signaling word-initial stops of Seoul Korean," *Phonetics and Speech Sciences*, vol. 8, pp. 25–36, 2016.

[8] J. Schertz, T. Cho, A. Lotto, and N. Warner, "Individual differences in phonetic cue use in production and perception of a non-native sound contrast," *Journal of Phonetics*, vol. 52, pp. 183–204, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.wocn.2015.07.003

[9] T. Cho, D. J. Kim, and S. Kim, "Prosodic strengthening in reference to the lexical pitch accent system in South Kyungsang Korean," *Linguistic Review*, vol. 36, pp. 85–115, 2019.

[10] J. Kuang, "Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice," *The Journal of the Acoustical Society of America*, vol. 142, pp. 1693–1706, 2017.

[11] M. Watkins, P. Boersma, and S. Hamann, "Revisiting pitch jumps: F0 ratio in Seoul Korean," in *Proc. INTERSPEECH 2024 – 23rd Annual Conference of the International Speech Communication Association*, Kos, Greece, Sep. 2024, pp. 3135–3139.

[12] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer," 2025, [computer program]. Version 6.4.27, retrieved 18 February 2025 from http://www.praat.org.

[13] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Report of the Institute of Phonetic Sciences Amsterdam*, vol. 17, 1993, p. 97–110.

[14] B. Gerratt and J. Kreiman, "Toward a taxonomy of nonmodal phonation," *Journal of Phonetics*, vol. 29, pp. 365 – 381, 2001.

[15] T. Cho, S. A. Jun, and P. Ladefoged, "Acoustic and aerodynamic correlates of Korean stops and fricatives," *Journal of Phonetics*, vol. 30, pp. 193–228, 2002.

[16] L. Lombardi, "Laryngeal features and laryngeal neutralization," Ph.D. dissertation, University of Massachusetts, Amherst, 1994.

[17] Y. Kang, J. Schertz, and S. Han, *The Phonology and Phonetics of Korean Stop Laryngeal Contrasts*. Cambridge University Press, 2022, pp. 215–247.

[18] I. Ejskjær, "Stød and pitch accents in the Danish dialects," *Acta Linguistica Hafniensia*, vol. 22, no. 1, pp. 49–75, 1990.

[19] N. Grønnum, M. Vazquez-Larruscaín, and H. Basbøll, "Danish stød. Laryngealization or tone," *Phonetica*, vol. 70, no. 1/2, pp. 66–92, 2013.

[20] E. Fischer-Jørgensen, "Phonetic analysis of the stød in Standard Danish," *Phonetica*, vol. 46, no. 1/2/3, pp. 1–59, 1989.

[21] J. M. Peña, "Stød timing and domain in Danish," *Languages*, vol. 7, pp. 50–88, 2022.

[22] N. Grønnum, "Three quarters of a century of phonetic research on common Danish stød," *Nordic Journal of Linguistics*, vol. 46, no. 3, pp. 299–330, 2023.

[23] G. F. Hansen, "Stød og stemmekvalitet," PhD dissertation, University of Copenhagen, 2015.

[24] J. M. Peña, "The production and perception of stød in Danish," PhD dissertation, New York University, 2024.

[25] W. Yun, K. Yoon, S. Park, J. Lee, S. Cho, D. Kang, K. Byun, H. Hahn, and J. Kim, "The Korean corpus of spontaneous speech," *Journal of the Korean Society of Speech Sciences*, vol. 7, no. 2, pp. 103–109, 2015.

[26] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed. New York: Springer, 2002, iSBN 0-387-95457-0. [Online]. Available: https://www.stats.ox.ac.uk/pub/MASS4/

[27] Posit team, *RStudio: Integrated Development Environment for R*, Posit Software, PBC, Boston, MA, 2025. [Online]. Available: http://www.posit.co/

[28] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, pp. 1–48, 2015.

[29] C. Draxler and K. Jänsch, "SpeechRecorder. A universal platform independent multi-channel audio recording software," in *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, and R. Silva, Eds. Lisbon: European Language Resources Association, 2004, pp. 559–562.

[30] N. J. Young and K. Anikwe, "Autophon.org. An online tool for automatic phonetic annotation," in *Proceedings from FONETIK 2024*. Stockholm University, 2024, pp. 89–94.

[31] H. M. Hanson, "Glottal characteristics of female speakers: Acoustic correlates," *The Journal of the Acoustical Society of America*, vol. 101, pp. 466–481, 1997.

[32] E. Fischer-Jørgensen, "Phonetic analysis of breathy (murmured) vowels in Gujarati," *Indian Linguistics*, vol. 28, pp. 71–139, 1967.