INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Coarticulation of Sibe Vowels and Dorsal Fricatives in Spontaneous Speech: An Acoustic Study

Jared Sharp[1], *Matthew Faytak*[1], *Hasutai Fei Xiong Liu*[2]

[1]University at Buffalo, United States
[2]Independent Researcher, China

jsharp5@buffalo.edu, faytak@buffalo.edu, hasutai@outlook.com

Abstract

Previous phonological analyses of dorsal segments in Sibe (Tungusic; Xinjiang, China) have treated the velar and uvular series as allophonic, with velar segments adjacent to /i ə u/ and uvular segments adjacent to /a o/. In this paper, we use a spontaneous speech corpus to examine the acoustic correlates of coarticulation of dorsal fricatives and vowels in Sibe. The first two spectral moments and mid-frequency spectral peak of dorsal fricatives were measured over the fricative onset in VC sequences and offset in CV sequences. Differences in spectral measures suggest that dorsal fricatives coarticulate with both preceding and following vowels primarily in terms of tongue dorsum backness, but with some role of height possible. These findings reflect a more complex relationship between Sibe vowels and dorsals than previously described; the distribution of labels used for the dorsal fricatives also suggests a gradient assimilation process.

Index Terms: dorsal fricatives, spectral moments, Sibe, coarticulation, acoustics

1. Introduction

1.1. Vowel-dorsal coarticulation in the Altaic area

Coarticulation, assimilation, and harmony processes involving vowels and dorsal consonants are an areal feature of languages of Northeast and Central Asia, especially within the Turkic, Mongolic, and Tungusic language families which form the Altaic linguistic area [1, 2, 3, 4]. Entire words are described as having a specification for backness affecting both the vowels and consonants. In Kazakh, for example, dorsal consonants in native words are found in complementary distribution: velar segments occur adjacent to front vowels, and uvular segments adjacent to back vowels [3]. Likewise, front and back vowels in native Kazakh words do not co-occur, which means that, at least word-internally, native words cannot contain both velars and back vowels, nor uvulars and front vowels. Kyrgyz vowel harmony has likewise been described in phonological terms as the spreading of a backness feature [5]. Vowels within a Kyrgyz word must match in terms of this backness feature. By contrast, Turkish does exhibit vowel backness harmony, where all vowels in the root share the same backness specification, but does not exhibit the vowel-consonant interaction observed in Kazakh or Kyrgyz [6].

These descriptions have tended to claim that vowel harmony and consonant assimilation are implemented in terms of the same single articulatory feature (i.e. tongue body backness or tongue root retraction). Acoustic studies have seemed to bear this out: in Kazakh, velar fricatives exhibited higher values for spectral center of gravity compared to uvulars, indicating a fronter constriction, and vowels have lower F2 values adjacent to uvular consonants [7]. However, the actual phonetic basis of this consonant-vowel interaction is rather poorly understood, and may well be more complex than typically described: when tongue articulation is more closely investigated, the mechanism of vowel production is seen to be more nuanced. Tongue ultrasound imaging studies have revealed that speakers of Kazakh, Kyrgyz, and Turkish articulate front and back vowels with the expected differences in the backness of the tongue body, but Kazakh and Kyrgyz also recruit the tongue root for producing the front-back distinction, i.e., tongue root retraction and tongue dorsum backing are co-produced [8].

These findings are important for the description of other, similar systems of vowel-consonant coarticulation or harmony: description of the process in terms of a single feature may not account for the full range of articulations produced by speakers. Further complicating the description of these harmony systems, some are known to be gradient or variable in their implementation. Prior acoustic investigation of the Kazakh vowels and dorsal consonants suggest that speakers do not strictly adhere to the co-occurrence restriction with back vowels in the production of their dorsal consonants, and that this co-occurrence restriction was not productive in nonce words.

1.2. Vowel-dorsal coarticulation in Sibe

Sibe ([ɕivə], ISO 639-3 sjo) is an endangered Tungusic language spoken in northwestern China. The Tungusic languages exhibit systems of vowel harmony and vowel-consonant assimilation similar to those described above [1, 2]. The Sibe dorsal consonants [k g q ɢ x χ ɣ ʁ] are conventionally described as alternating based on the height of adjacent vowels: the uvulars [q ɢ χ ʁ] have been described as co-occurring with [− high] vowels /a o/, whereas velars [k g x ɣ] are often described as co-occurring with the [+ high] vowels /i ə u/ [9]. For example, the words [xəx] ‘woman’ and [χaχ] ‘man’ follow a similar static co-occurrence restriction to that observed in Kazakh and Kyrgyz.

The articulatory and acoustic dimensions involved in vowel-dorsal consonant coarticulation in Sibe have yet to be explored acoustically. Furthermore, as previously noted for Kazakh [7], however, vowel-dorsal co-occurrence restrictions in Sibe appear to vary in spontaneous speech. Previous impressionistic analyses suggest that speakers do not consistently produce uvulars and velars in the expected places, even across different productions of the same lexical item [10, 11]. This apparent loss of predictability in the distribution of velars and uvulars with respect to adjacent vowels raises the possibility that the two series are not allophonic variants of each other as typically described. It remains unclear the extent to which tran-

10.21437/Interspeech.2023-44

scribers' perceptual biases may impact the transcription of this variability.

1.3. Research questions

In this paper, we use acoustic measures of fricative spectra to gain insight into the phonetics of coarticulation of the Sibe dorsal fricatives and their surrounding vowels: is the basis for Sibe dorsal consonant production a single articulatory dimension such as tongue dorsum height or backness, or a more complex combination of such factors? We assess several measures expected to inversely vary with the length of the oral cavity in front of the fricative constriction (i.e., a longer cavity yields a lower measure): the first two spectral moments, center of gravity and standard deviation; and mid-frequency spectral peak [12, 13]. Differences in these measures may reflect differences in the relative anteriority of articulation (velar or uvular) of the dorsal fricatives, allowing for patterns associated with particular neighboring vowels to emerge from the data.

We also carry out a supplementary analysis on the transcribed labels (uvular [χ] versus velar [x]) to assess the relationship between perceived fricative place and the various acoustic measures. This serves to assess the biases that may have played a role in past impressionistic analysis, and to assess whether the impressionistic labels are determined primarily by the identity of surrounding vowels or by spectral characteristics of the fricative consonants themselves.

2. Methods

2.1. Materials

A Sibe spontaneous speech corpus was collected in Xinjiang between 2018 and 2021 for an educational WeChat channel; materials were taken as the basis of this study with the permission of the collectors. All recordings were made using various lapel microphones connected to a Sony PCM-D50 recorder. Recordings were processed at a sampling rate of 44,100Hz/16-bit.

Using a working orthography devised by the first and third authors, roughly three hours of continuous speech from seven speakers in the corpus were transcribed in Praat [14]. The resulting transcripts were used to generate a pronunciation dictionary, train an acoustic model, and carry out forced alignment using the Montreal Forced Aligner (MFA) [15]. Initial output alignments were hand-corrected as necessary.

The hand-corrected segment start and end points were used to extract short portions of target velar and uvular fricatives for inclusion in two models described in more detail below. One model is concerned with anticipatory coarticulation with following vowels in CV sequences, and the other is concerned with carryover coarticulation with preceding vowels in VC sequences. For CV sequences, the final 20% of the fricative's duration (the offset) was analyzed, and for VC sequences, the initial 20% of the fricative's duration (the onset) was analyzed. Analysis of the offset and onset portions was chosen to maximize influence from the following and preceding vowel. Intervocalic fricatives are counted twice in Tables 1-2, as both the offset and onset of these tokens are examined for anticipatory and carryover coarticulation, respectively.

This procedure yielded 3,524 tokens, inclusive of both fricative offsets and onsets (Table 1). The number of offset and onset tokens varies according to the adjacent vowel, reflecting the natural probability of the vowels in Sibe in spontaneous speech. Speakers contributed a comparable amount of tokens (Table 2) with the exception of M4, who contributes fewer than 50 tokens due to taking few speaking turns in the corpus.

2.2. Analysis

The first two spectral moments, spectral center of gravity (COG) and spectral standard deviation (SD) [12], as well as the mid-frequency spectral peak (MFSP), were extracted for all target dorsal fricative onsets and offsets in R using a modified version of the *Multitaper Spectral Analysis on Sound Segments* script by Wilson & Chodroff [16]. Intervocalic fricatives were variably voiced by speakers throughout the corpus. To avoid any confound of voicing in the spectral moments analysis, a Hann stop-band filter was used to remove frequencies lower than 550 Hz in the recordings, effectively removing the voicing source following [17, 18, 13]. All measurements were averaged across eight tapers in the 20% onset or offset window. Outliers for each measure more than two standard deviations from the mean for a given CV and VC pair were excluded from analysis.

Table 1: *Fricative portion tokens by preceding and following vowel context.*

Vowel	Offset (CV)	Onset (VC)
/i/	108	381
/ə/	766	259
/a/	461	497
/o/	147	218
/u/	485	202
Total	1,967	1,557

Table 2: *List of speakers (F = female, M = male) with number of targets contributed per speaker.*

Speaker	Age	Total tokens
F1	65	498
F2	85	626
M1	40-42	713
M2	66	338
M3	68	846
M4	75	24
M5	76	479
Total		3,524

Acoustic measures were submitted to mixed-effects linear regressions for each combination of measure and position (onset or offset). Each model included fixed effects of preceding or following vowel and speaker gender, with a random slope for consonant duration by speaker, and a random intercept for word. The random slope for duration is included to account for duration-related gestural undershoot in a speaker-specific way. The intercept reference was set to [a] so that comparisons could be made with respect to height and backness of other vowels.

Impressionistic fricative labels ([x] or [χ]) were provided for each dorsal fricative by the third author, a Sibe speaker who assisted with transcription. We visualize all acoustic data splitting by these impressionistic labels, but they were not included as a factor in the linear models. These impressionistic labels were, however, subject to an additional analysis in Section 3.3 to assess the relationship between perceived fricative place and the various acoustic measures (see Figure 3).

3. Results

The offset data plotted in Figure 1 suggest an association of lowered center of gravity (COG) and standard deviation (SD) with lower and backer following vowels. An association of *lower* mid-frequency spectral peak (MFSP) with *higher* vowels is also suggested by the distribution of the raw data, though as will be seen below, the linear mixed effects models offer a different characterization of MFSP more similar to the pattern for COG and SD. Figure 2 suggests a similar pattern for fricative onsets. We turn to the linear mixed-effects models to confirm these patterns.

3.1. Fricative offsets by following vowel

In the offset (CV) model for COG, the intercept for the reference level [a] (β = 836.103, t[5.023] = 19.394, $p < 0.001$) reached significance. The vowels [i] and [ə] significantly raised offset COG relative to baseline; [i] had a larger effect ([i]: β = 793.860, t[765.447] = 23.321, $p < 0.001$) compared to [ə] (β = 115.132, t[519.896] = 6.018, $p < 0.001$). The effects of the back vowels [o] and [u] failed to reach significance.

In the offset model for SD, the intercept for reference [a] reached significance (β = 309.266, t[5.971] = 4.877, $p < 0.01$). Relative to this baseline, the vowels [i], [ə], and [u] all significantly raised offset SD: the effect of following [i] is largest (β = 971.307, t[1185.995] = 20.954, $p < 0.001$), followed by [ə] (β = 348.843, t[739.104] = 13.051, $p < 0.001$), then [u] (β = 161.736, t[858.171] = 5.498, $p < 0.001$). The effect of [o] relative to baseline failed to reach significance.

Finally, in the offset model for MFSP, the intercept for reference [a] reached significance (β = 735.232, t[4.882] = 26.370, $p < 0.001$). Following [i] raised offset MFSP (β = 181.050, t[477.785] = 8.567, $p < 0.001$), and following [o] lowered offset MFSP (β = -42.480, t[392.140] = -2.272, $p < 0.05$); the effects for [ə] and [u] did not reach significance.

3.2. Fricative onsets by preceding vowel

In the onset model for COG, the [a] intercept reaches significance (β = 798.0099, t[4.5258] = 11.266, $p < 0.001$). The non-back vowels [i] and [ə] exhibit significantly raised COG relative to this baseline, with a larger effect for [i] (β = 616.3253, t[847.4821] = 15.527, $p < 0.001$) than [ə] (β = 247.9935, t[857.7469] = 6.114, $p < 0.001$). As in the onset model, the effects for the back vowels [o] and [u] failed to reach significance.

In the onset model for SD, as with the offset model, the [a] intercept was significant (β = 236.087, t[4.638] = 3.420, $p < 0.05$). The vowels [i], [ə], and [u] significantly raised SD, again with a larger effect for [i] (β = 783.396, t[974.190] = 20.876, $p < 0.001$) compared to [ə] (β = 379.811, t[929.146] = 9.767, $p < 0.001$) and [u](β = 206.487, t[857.296] = 4.939, $p < 0.001$). The effect of the vowel [o] relative to baseline did not reach significance.

Finally, in the onset model for MFSP, the intercept [a] again reached significance (β = 745.307, t[7.985] = 23.673, $p < 0.001$). The high front vowel [i] was the only vowel with a significant effect on MFSP relative to baseline, somewhat raising it (β = 221.603, t[299.235] = 7.925, $p < 0.001$). The effects of preceding [ə], [o], and [u] failed to reach significance.

3.3. Acoustic basis of transcriptional labels

As seen in Figures 1-2, in most vowel contexts, only a small difference can be observed in spectral measures between the onsets and offsets of fricatives labeled velar [x] and uvular [χ]. In order to assess the relationship between the acoustic measures and labels typically chosen for (non-contrastive) dorsal fricative place, we provide count data for each transcribed label in each vowel context in Figure 3. The uvular [χ] transcription is primarily associated with the retracted vowels [a] and [o], whereas dorsal fricatives associated with [i] are mostly transcribed as velar [x].

We additionally compared the two labeled groups ([x] and [χ]) on each acoustic measurement (COG, SD, MFSP) at onset and offset, using Welch's paired t-tests (two-tailed). T-tests for each measure (COG, SD, MFSP) showed significant differences between the tokens assigned each label. Fricative offsets labeled as uvular had significantly lower COG (M = 827.437) than those labeled as velar (M = 1015.041), t[841.55] = -10.027, $p < 0.001$. Likewise, fricatives labeled as uvular had significantly lower SD (M = 465.010) than those labeled as velar (M = 784.717), t[940.10] = -13.913, $p < 0.001$. Fricatives labeled uvular also had lower average mid-frequency spectral peaks (M = 729.0451) than those labeled as velar (M = 746.2047), but the difference was not significant (t[808.66] = -1.5881, $p > 0.05$).

T-tests also revealed significant differences in spectral moments measurements between fricative onsets labeled as uvular or velar. Both offsets and onsets of fricatives (in CV and VC sequences, respectively) labeled as uvular had lower center of gravity (M = 859.024) than those labeled as velar (M = 1315.330), t[708.93] = -14.498, $p < 0.001$; spectral standard deviations were also lower for those labeled uvular (uvular M = 482.791, velar M = 1033.881, t[833.99] = -18.016, $p < 0.001$). Mid-frequency spectral peaks were likewise lower for fricatives labeled uvular (M = 751.572) than those labeled velar (M = 900.818), t[670.80] = -6.5419, $p < 0.001$.

4. Discussion

In this study, we examined Sibe dorsal fricatives at offset and onset to determine the relevant dimensions (height, backness) of Sibe vowel-dorsal coarticulation through acoustic measures. Linear mixed effects models revealed differences in spectral moments (COG, SD) and mid-frequency spectral peak dependent on the vowel preceding onset or the vowel following offset. Variation in COG and SD at fricative onset and offset was primarily driven by the frontness of the flanking vowel, counter to the analysis in [9], which described height as the relevant factor. This suggests a tendency for a more anterior (velar) articulation with more anterior vowels, and more posterior (uvular) articulation with the back vowels: essentially, the closer to [i] the articulation of a vowel, the higher the COG and SD. Modeled MFSP also appears to vary mainly according to the frontness of the adjacent vowel, with /i/ conditioning a higher MFSP than all other vowels. From prior studies of sibilant fricatives, MFSP is known to be tightly related to the anteriority of constriction [13, 17, 18]; this finding again suggests that an articulatory back-front dimension drives most variation in dorsal place.

As with neighboring languages, previous phonological descriptions of Sibe describe the vowel-dorsal consonant coarticulation system as being one-dimensional, limited to one axis of movement, with height favored as the basis [9]. This study's findings suggest that the uvular variants of the Sibe dorsal consonants are gradiently conditioned by the backness of the surrounding vowels. We cannot rule out, however, a role of vowel height, as evidenced by occasional significant main effects of the high back [u] on acoustic measures. Alternately, a more

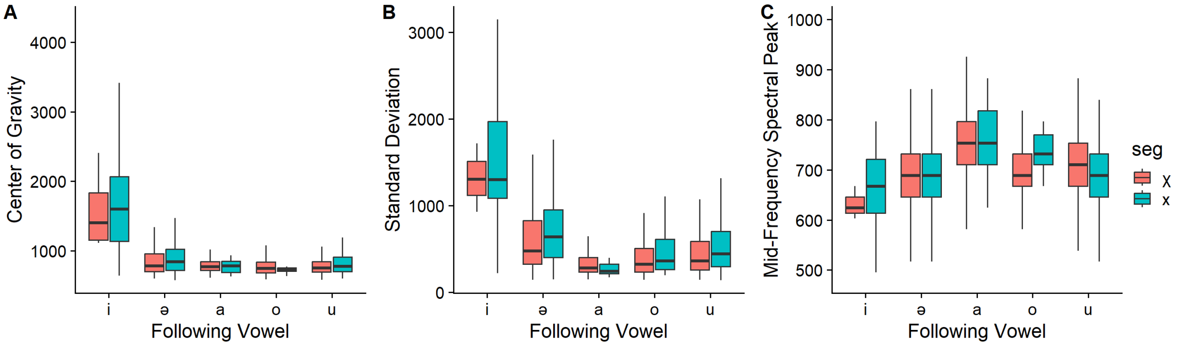

Figure 1: *A) Center of gravity, B) spectral standard deviation, and C) mid-frequency spectral peak of dorsal fricative offset by following vowel.*

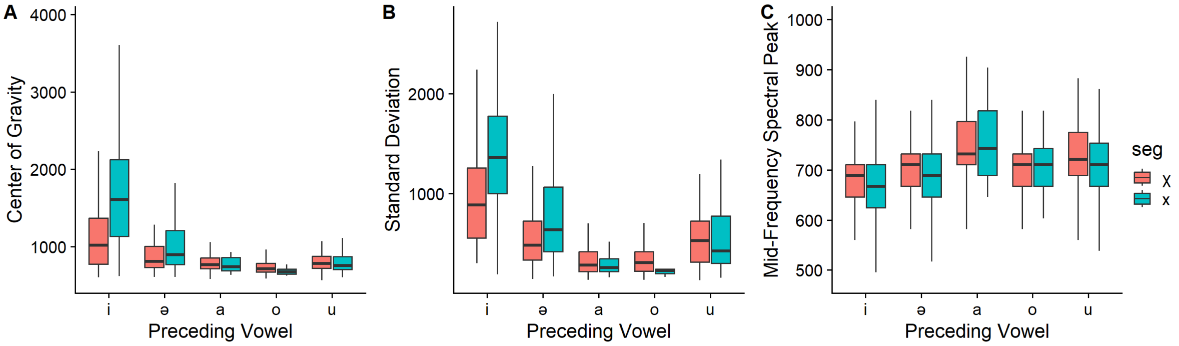

Figure 2: *A) Center of gravity, B) standard deviation, and C) mid-frequency spectral peak of dorsal fricative onset by preceding vowel.*

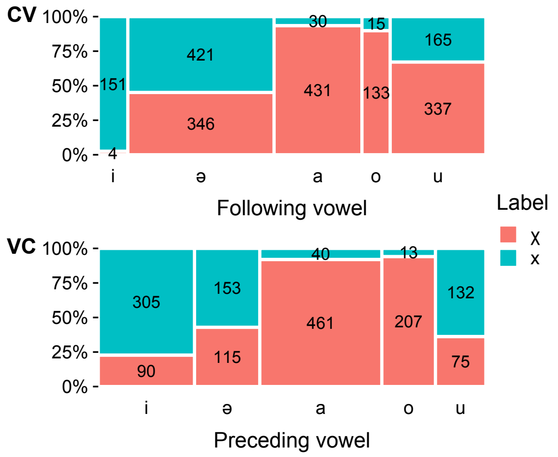

Figure 3: *Mosaic plot of label distribution by following vowel in CV contexts (top) and by preceding vowel in VC contexts (bottom).*

biomechanically realistic dimension such as lingual retraction or constriction against the epilarynx [19] may provide a simple basis for this harmony, as evidenced by [a] and [o] frequently grouping together in their effects on fricative acoustics.

T-tests revealed significantly higher spectral center of gravity and standard deviation for fricatives labeled velar compared to those labeled uvular, as well as lower MFSP, for both fricative onset and offset measures, across all vowel contexts. These results suggest that transcriptional labels do reflect acoustic variation in dorsal fricative production and are not simply a product of perceptual integration with the surrounding segmental environment. Nonetheless, phonological descriptions of these languages may not capture the full range of articulations performed by speakers, giving acoustic and articulatory investigations of these languages extra urgency: these languages are at risk of disappearing before acoustic or articulatory investigations can be carried out. Our results have implications for the phonological literature on vowel-vowel and vowel-dorsal consonant coarticulation in Sibe specifically and in the Altaic area in general; they also highlight the importance of phonetic investigation when describing systems of assimilation or coarticulation to produce a more accurate account of a language's phonology.

5. Acknowledgements

The authors thank Haibo Wang (王海波) and Christian DiCanio for helpful feedback. We also express our gratitude to Sirdan (孟荣路) for his continued work with Sibe language documentation and preservation, without which the present research could not have been accomplished.

6. References

[1] J. Ard, "A sketch of vowel harmony in the Tungus languages," *Paper in Linguistics*, vol. 16, no. 3-4, pp. 23–43, 1983. [Online]. Available: https://doi.org/10.1080/08351818309370594

[2] G. Hansson, *Consonant Harmony: Long-Distance Interaction in Phonology. Number 145 in University of California Publications*

in Linguistics. University of California Press, Berkeley, CA, 2010.

[3] A. G. McCollum and S. Chen, "Kazakh," *Journal of the International Phonetic Association*, vol. 51, no. 2, p. 276–298, 2021.

[4] J.-O. Svantesson, A. D. Tsendina, A. Karlsson, and V. Franzén, *The Phonology of Mongolian*, ser. The Phonology of the World's Languages. Oxford University Press, 2005.

[5] A. G. McCollum, "Vowel harmony and positional variation in kyrgyz," *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 2020.

[6] G. N. Clements and E. Sezer, "Vowel and consonant disharmony in Turkish," in *The Structure of Phonological Representations, pt. 2*, H. van der Hulst and N. Smith, Eds. Dordrecht: Foris Publications, 1982.

[7] H. L. Yawney, "Acoustic properties for the Kazakh velar and uvular distribution," *Proceedings of the Workshop on Turkic and Languages in Contact with Turkic*, vol. 6, 2021.

[8] J. N. Washington, "An investigation of the articulatory correlates of vowel anteriority in Kazakh, Kyrgyz, and Turkish using ultrasound tongue imaging," 2019.

[9] T. Kubo, "A sketch of Sibe phonology," *Gogaku Kenkyuu Fooramu*, vol. 16, pp. 127–142, 2008.

[10] T. Kubo, N. Kogura, and S. Zhuang, シベ語の基礎. Research Institute for Languages and Cultures of Asia and Africa. Tokyo University of Foreign Studies, 2011, the Fundamentals of the Sibe Language.

[11] H. Wang, "満洲・シベ語現代方言音韻論 [The Phonology of the Modern Dialects of the Manchu-Sibe Languages]," Ph.D. dissertation, University of Tokyo, 2018.

[12] K. Forrest, G. Weismer, P. Milenkovic, and R. N. Dougall, "Statistical analysis of word-initial voiceless obstruents: preliminary data," *The Journal of the Acoustical Society of America*, vol. 84, no. 1, pp. 115–123, 1988.

[13] L. Koenig, C. Shadle, J. Preston, and C. Mooshammer, "Toward improved spectral measures of /s/: results from adolescents," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 4, pp. 1175–1189, 2013.

[14] P. Boersma, "Praat, a system for doing phonetics by computer," 2001.

[15] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal forced aligner: Trainable text-speech alignment using Kaldi," in *INTERSPEECH*, 2017.

[16] C. Wilson and E. Chodroff, "Multitaper spectral analysis on sound segments," 2014. [Online]. Available: https://github.com/echodroff/praat_scripts/blob/main/MultitaperSpectralMomentsPeak_trains.R

[17] E. Chodroff, "Structured variation in obstruent production and perception," Ph.D. dissertation, 2017.

[18] E. Chodroff and C. Wilson, "Uniformity in phonetic realization: Evidence from sibilant place of articulation in american english," *Language*, vol. 98, no. 2, 2022.

[19] J. H. Esling, "There are no back vowels: The larygeal articulator model," *Canadian Journal of Linguistics/Revue canadienne de linguistique*, vol. 50, no. 1-4, pp. 13–44, 2005.