Interspeech 2018
2-6 September 2018, Hyderabad, India

An ultrasound study of gemination in coronal stops in Eastern Oromo

Maida Percival, Alexei Kochetov, Yoonjung Kang

University of Toronto

maida.percival@mail.utoronto.ca, al.kochetov@utoronto.ca, kang@utsc.utoronto.ca

Abstract

This study extends the use of ultrasound methodology to stops in Eastern Oromo (Cushitic; Ethiopia) to examine the link between gemination, laryngeal features, and tongue shape.

Ultrasound data were collected from 5 native speakers of Eastern Oromo. Tokens consisted of 12 repetitions per speaker of [t^h, t, d, ɗ] and six of [tt^h, tt, dd, ɗɗ] in the environment of a_a. Tongue images at the point of maximum constriction during the stop closure were traced following [1] and their coordinates submitted to linear mixed effects models.

Results indicated differences in tongue shape between singletons and geminates, especially for ejectives and implosives. Singleton ejectives displayed raised tongue bodies not found in geminate ejectives. Singleton implosives resembled voiceless stops, but geminate implosives were variably produced with tongue body raising.

I suggest that the results can be attributed to fortition in geminates. Tongue body raising in singleton ejectives may be an enhancement strategy to the ejective contrast that is not necessary in longer geminates. The singleton implosive resembling a voiceless aspirated stop is predicted by [15] while the geminate tongue body raising may be retraction, c.f. [2]. The results support a link between tongue, larynx, and gemination.

Index Terms: ultrasound imaging, Eastern Oromo, stops, gemination, laryngeal contrasts

1. Introduction

Geminate consonants are known to differ from singleton consonants primarily in their duration, although less commonly other characteristics, mainly acoustic, have been noted to be cues ([3], [4]). [2] for Italian and [5] for Japanese also found articulatory differences in geminates, using electromagnetic articulography to show that there was greater linguopalatal contact for geminate than non-geminate stops. [2] noted that the tongue was higher and flatter in the mouth for coronal geminate sonorants and stops, suggesting more laminal rather than apical articulation. Both studies liken gemination with fortition processes.

This study provides further evidence for articulatory differences in geminate consonants by using ultrasound imaging technology to examine the interaction between gemination and laryngeal features in Eastern Oromo, a Cushitic language of Ethiopia with length and laryngeal contrasts in the coronal region between voiced, aspirated, ejective and implosive stops. The study also introduces new method of quantifying ultrasound imaging of the tongue using linear mixed effects models on normalized tongue spline radii. The findings support a view of gemination as fortition, uncover evidence for different articulatory strategies in the realization of laryngeal contrasts based on consonant duration, and may also have implications for previous analyses (e.g. [15]) of Oromo's implosive.

2. Methodology

2.1. Participants

Five adult native speakers of Eastern Oromo (2 male, 3 female) were recruited. Their age ranged from 32-59 with a mean of 47. All also spoke English and some spoke Amharic (2), Arabic (3), Italian (1), and Somali (2), but reported having acquired these languages after age 5 and speaking Oromo at home with family. No hearing or vision impairments were reported.

2.2. Materials

Tokens consisted of 12 repetitions per speaker of [t^h, t, d, ɗ] and six of [tt^h, tt, dd, ɗɗ] in the environment of a_a. The list of stimuli is shown in Table 1.

Table 1: *Stimuli*

	Ejective	**Aspirated**	**Voiced**	**Implosive**
Singleton	foo**t'**aa *'scarf'*	ma**t**aa *'head'*	ma**d**aa *'wound'*	ɗa**ɗ**aa *'butter'*
				haa**ɗ**a *'mother'*
Geminate	ma**tt'**annee *'skinny'*	ba**tt**ala *'quick'*	ba**dd**anno *(city name)*	ha**ɗɗ**aawaa *'bitter'*
				ha**ɗɗ**a *'ash'*

2.3. Instrumentation and Procedure

Data were recorded with an Echo B Portable Ultrasound using Articulate Assistant Advanced (AAA) software. The schemata in Figure 1 shows the set-up and equipment. The probe was affixed to the participant's chin with a head set [6]. Simultaneous electroglottograph data were collected but are not the subject of this paper. The ultrasound frame rate was approximately 40fps.

2.4. Annotation and Analysis

Annotations were done in AAA. Ultrasound video frames were selected at the point of maximum constriction of the stop, i.e. at the most stable, middle portion of the closure [1]. Tongue tracings were made as fan splines, where the tongue is conceptualized as the arc of a circle and its surface contour is quantified as the length of the radius from the centre of the probe to the surface of the tongue along 42 degrees of this circle. The tracings were rotated to the participant's bite plane [7], and the 42 points along the surface of each token's tongue tracing were exported from AAA and converted into polar coordinates so that the length of the radii of the splines could be submitted to statistical analysis.

As the 42 points were along the entire length of the probe's

10.21437/Interspeech.2018-2512

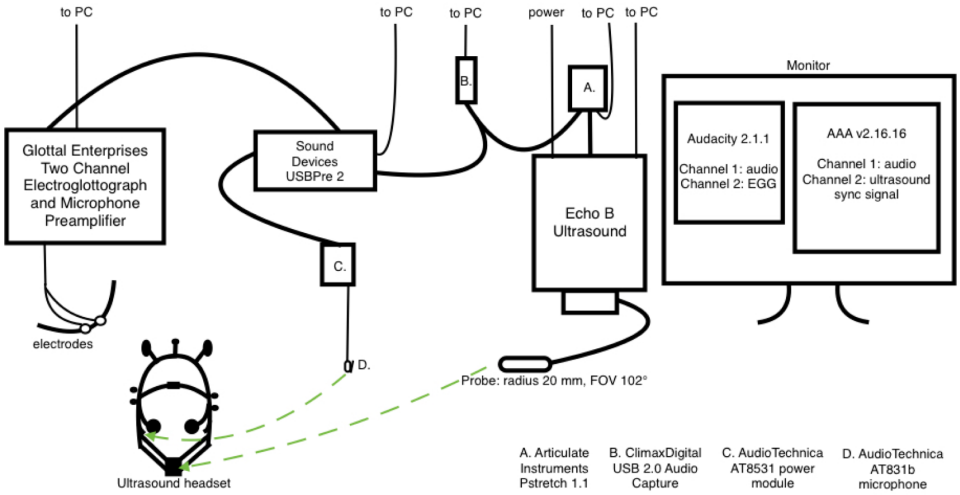

Figure 1: *Set up.*

field of vision, not all of them corresponded to the tongue surface–the tongue was generally smaller than the field of view. Therefore, the radius values of points not corresponding to parts of the tongue were removed from the analysis. The remaining radius values were grouped into three regions: coronal, velar, and pharyngeal, with points to either side of the closure of velar stops (recorded for this purpose) corresponding to the coronal-velar and velar-pharyngeal boundaries, as illustrated in Figure 2. For a similar segmentation of the tongue, see [8].

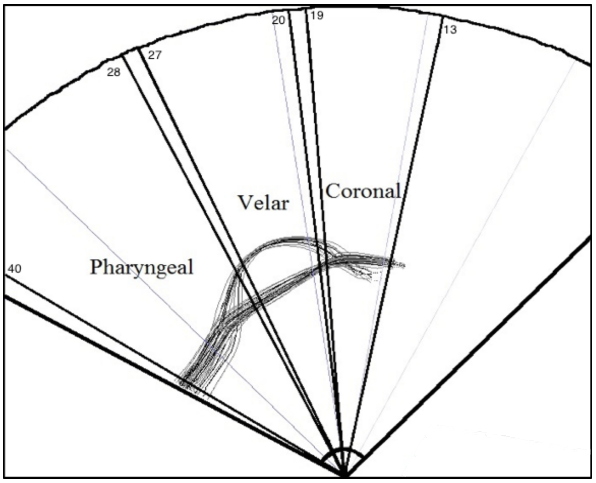

Figure 2: *Division of the tongue into regions for a sample speaker O5.*

Data preparation and statistical analysis were done in R [9]. Prior to the statistical analysis, the points along the tongue that made up each tongue region were normalized by scaling them so that there were the same number of points in a given region for all tokens across all speakers. The mean radius associated with each of the new points was also calculated. Previously different speakers had tongue regions with differing numbers of datapoints, as their tongues were of different sizes. Scaling the data so that there are the same number of datapoints for each token for each speaker allows for the use of linear mixed-effects models for statistical analysis rather than the traditional SS-ANOVA methodology of ultrasound, where, due to tongue differences, participants are more or less separate case studies.

Statistical analysis was done in R using the lme4 package [10]. Linear mixed-effects models as in (1) were performed on the radii for the scaled points for each region. The fixed effects were consonant type, which had four levels (voiced, aspirated, ejective, implosive), consonant length, which had two levels (singleton and geminate), and *normalized spline*, which were the normalized points along the tongue at which the radii were measured. The random effects were participant (O1 through O5) and word (Table 1). The data were contrast coded with [dd] as reference level. Pairwise comparison using least-squares means served as post-hoc tests.

$$r \sim C * length * spl.normed + (C|participant) + (1|word) \quad (1)$$

26/90 tokens for O1 due to an ultrasound syncing issue and 10/90 for O4 due to a head stabilization issue were excluded from analysis. Nonetheless, there were at least 3 tokens for each sound for O1 and at least 5 for each sound for O4.

3. Results

The results are graphically displayed in Figure 3, which shows the mean radius normalized by spline number of coronal, velar, pharyngeal regions of tongue for geminate and singleton consonants for each of the 4 consonant types for the 5 participants.

In general, results of the linear mixed effects analysis show for each of the regions (coronal, velar, pharyngeal) an effect of the normalized splines, which is reflected as the slope as seen in Figure 3. This effect is always expected to be significant, since different points along the surface of the tongue correspond to radii of different lengths.

For the coronal region the effect of the normalized splines was t=-65.981, $p<0.001$, df = 2413.8 showing that the radii decrease in size by 3.71 mm +- 0.056 s.e. as the points along the tongue increase (i.e., going from the tongue tip backwards). For the velar region, for each point of measure going backwards along the tongue, the radii decreased by a mean of 5.346 mm $\pm$ 0.06 s.e. (t = -82.844, $p<0.001$, df=2491). For the pharyn-

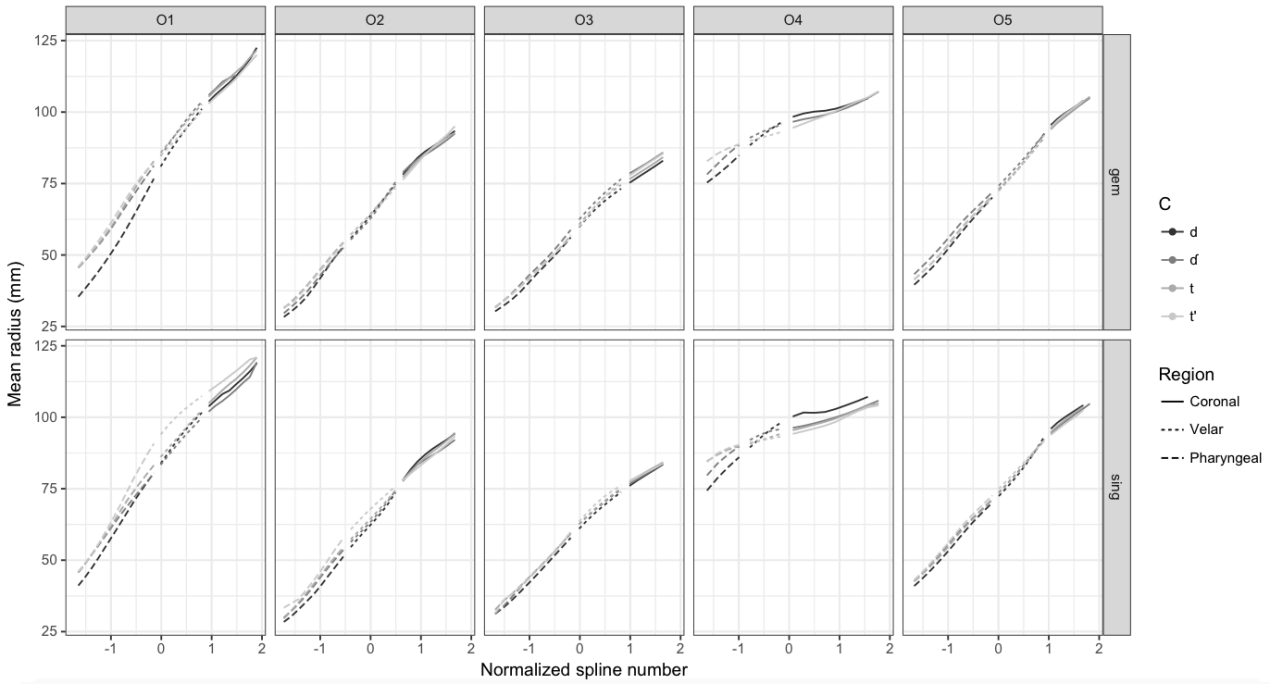

Figure 3: *Mean radius normalized by spline number of coronal, velar, pharyngeal regions of tongue for geminate (top) and singleton (bottom) consonants for each of the 4 consonant types for the 5 participants. Note that O1 and O5 are male, the rest female.*

geal region, it decreased by 8.71 mm ± 0.08 s.e. (t = -111.402, p<0.001, df=3726).

3.1. Coronal region

In the coronal region, no significant effects for consonant type or consonant length were found. This suggests there were no significant differences in the shapes of the tongue for geminate and singleton stops or for stops of different laryngeal categories. This can be seen in Figure 3–there is a great deal of overlap in the slopes in the coronal region and where there isn't overlap, there is no discernible pattern that holds across all speakers.

3.2. Velar region

In the velar region, there was an effect of consonant length (t=4.008, p=0.005, df = 7.2) whereby the radii of singletons is on average 0.8 mm ± 0.2 s.e. longer than those of geminates. A significant interaction between [t'] and consonant length found that the difference between singletons and geminates is even greater for ejectives (Est. = 2.769 ± 0.587 s.e.; t=4.715, p = 0.002, df=8) and post-hoc tests revealed that it in fact only holds for ejectives (p<0.001). These results are evident in Figure 3 where the ejectives (palest grey line) in the singleton panel (bottom) have the highest radii for all speakers (except O4, who may be somewhat of an outlier). The implosives also displayed a marginal trend in the opposite direction, whereby radii were greater for geminates (p=0.097).

The radii of [t'] being longer than [tt'] in the velar region suggests that the tongue body is higher for [t']. This is evident in images of AAA tracings of the tongue, such as Figures 4 and 5, which show the mean tongue splines of singletons and geminates for a representative speaker. The tongue body for [t'] (topmost line) is raised compared to other consonants in 4, while [tt'] is not raised in 5, where it is nearly indistinguishable from [tt^h] (middle lines).

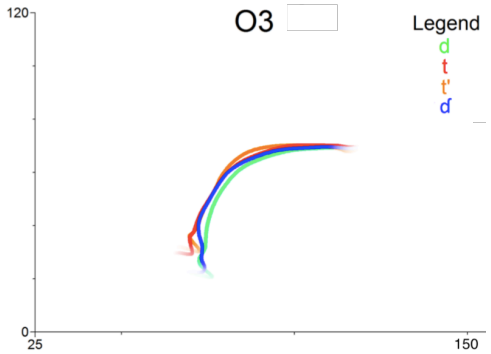

Figure 4: *Image of mean tongue tracings for singleton consonants for a representative speaker (O3) (in mm)*

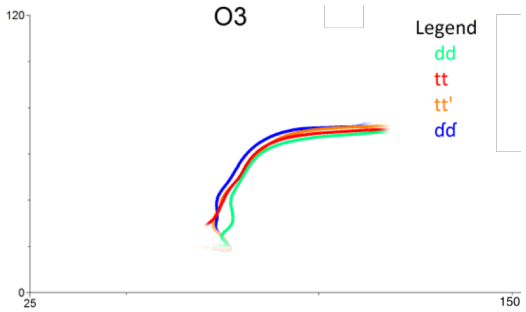

Figure 5: *Image of mean tongue tracings for geminate consonants for a representative speaker (O3) (in mm)*

Finally, there was a significant interaction between [t'] and the normalized splines (Est.=1.05692, s.e.=0.19282; t=5.481, p<0.001, df=2491), expressing that longer radii for singleton ejectives result in a shallower slope of the tongue.

3.3. Pharyngeal region

In the pharyngeal region, there was an effect of consonant length (t=6.23, p<0.001, df=7) which indicated that the radii of singletons are on average 1.3 mm ± 0.2 s.e. longer than those of geminates. A significant interaction between [ɗ] and consonant length (Est.=-1.38, s.e.=0.559; t=-2.477,p,0.040, df=8) found that singleton implosives are less different than other stop types. Post-hoc tests confirmed that singletons have significantly longer radii than geminates only for voiced (p=0.045), aspirated (0.05), and ejective stops (p=0.001). AAA tracings for a representative speaker in Figure 6, where mean singletons (dotted line) and geminates (solid line) are superimposed, show that the tongue is produced more advanced in the mouth for geminates than singletons but barely so for implosives.

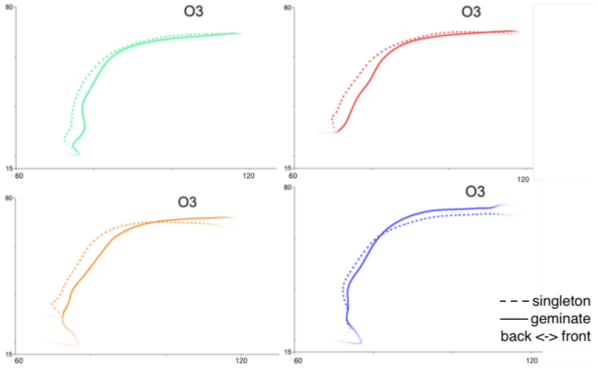

Figure 6: *Image of mean tongue tracings for singleton vs. geminate consonants for a representative speaker (O3). Top-left = voiced, Top-right = aspirated, Bottom-left = ejective, Bottom-right = implosive. Dashed = singleton, Solid = geminate. (mm)*

There was also significant effect of consonant. The length of the radii for [ɗ] (t=5.384, p<0.001, df=9), [t^h] (t=3.628, p<0.011, df=6), and [t'] (t=3.180, p<0.021, df=5) were longer than that of [d], regardless of length. They were on average 3.74 mm ± 0.70 s.e. longer for [ɗ], 4.05 mm ± 1.11 s.e. for [t^h], and 4.14 mm ± 1.30 s.e. for [t']. This can be seen in Figure 3. The shorter radii of voiced stops in the pharyngeal region may be attributed to advanced tongue root, as suggested by the position of the bottom portion of the tongue in Figures 4 and 5 when comparing voiced to other stop types.

Finally, there were interactions with the normalized splines. Singletons had steeper slopes than geminates (Est.=-0.694, s.e.=0.096; t=-7.231, p<0.001, df=3726). This was especially the case for ejectives (Est.=-2.326, s.e.=0.289; t=-8.047, p<0.001, df=3727). An interaction with implosives indicated a less steep slope for [ɗɗ] (Est.=0.802, s.e.=0.208; t=3.862, p<0.001, df=3726).

4. Discussion and conclusions

The findings point towards differences in tongue shape based on both consonant length and laryngeal contrast. In general, geminate consonants differed from singletons in that they were produced with a more advanced tongue. This is different than the retracted laminal closure and raised tongue [2] found in Italian. As Oromo stops are reported to be dental ([11]) rather than alveolar like the Italian consonants, the findings may point towards the geminates involving a more forceful closure, the difference in place of articulation simply resulting in tongue advancement rather than tongue raising to achieve this effect.

Voiced stops were found to be produced with advanced tongue root. This is expected given their articulation and the nature of voicing. Vocal fold vibration is initiated via a build-up of air pressure below the vocal folds so that when the pressure is great enough compared to the supraglottal cavity, air will force its way through and vibration will begin [12]. The pressure difference between the sub- and supra-glottal cavities is difficult to maintain for stops because the oral closure creates an intraoral cavity which also increases in air pressure, meaning that pressure in the subglottal cavity must increase even more [12], [13]. [12] and [14] suggest that this build-up of air pressure may be achieved by expanding the oral cavity or compressing the subglottal cavity and they note that strategies to help achieve this include advancing the tongue root.

Oromo implosives did not display advanced tongue root as expected of voiced sounds. [15] suggests that due to its phonological patterning and historical origin, the implosive in Oromo should be considered voiceless. The lack of advanced tongue root supports such an analysis. However, impressionistically, implosives in Oromo seem to be phonetically voiced–a forthcoming EGG study will have more to say on this. It may be that, as in English where voiced stops that are phonetically voiceless are still produced with advanced tongue root [16], phonological voicing is more crucial to determining tongue root.

The geminate implosive did not display tongue advancement like other geminates, but did display a marginally higher tongue body in the velar region. This is suggestive of what [2] found for geminates in Italian, where they were more laminal or palatalized. [11] describes the implosive in Oromo as (unlike other coronals) alveolar or slightly retroflexed. It may be that the implosive is articulated further back in the geminate since that is its target place of articulation.

Ejective stops displayed tongue body raising in singletons but not geminates. This is the opposite pattern of what [2] predicts. I propose that the raising of the tongue body for [t'] may be an enhancement strategy to help maintain the ejective quality. Ejectives involve a closure at the glottis and in the oral cavity accompanied by an increase in air pressure in the intraoral cavity. Raising the tongue body would increase the air pressure in the intraoral cavity (similar to how larynx raising compresses it from the other direction). [t'], being short, may use tongue body raising as an enhancement strategy, whereas [tt'], being long, has more time during the stop closure in which air pressure can increase, and does not appear to need this strategy. This finding suggests that multiple strategies may be used by a speaker to maintain laryngeal stop contrasts, and that consonant duration can affect which one is used.

To conclude, this paper has presented an ultrasound study of stops in Eastern Oromo and found that the shape and position of the tongue relate to both the gemination and laryngeal contrasts. Gemination can be viewed as fortition which influences the production of implosives and ejectives in particular.

5. Acknowledgements

Thank you: Dr. Phil Monahan, Kiranpreet Nara, Abdulhamid Mohammed, Radu Craioveanu, Luke Zhou, Emily Clare, all of the Oromo participants & SSHRC.

6. References

[1] A. Kochetov, N. Sreedevi, M. Kasim, and R. Manjula, "Spatial and dynamic aspects of retroflex production: An ultrasound and ema study of kannada geminate stops," *Journal of Phonetics*, vol. 46, pp. 168–184, 2014.

[2] E. M. Payne, "Non-durational indices in italian geminate consonants," *Journal of the International Phonetic Association*, vol. 3, no. 1, pp. 83–95, 2006.

[3] H. Kubozono, "Introduction to the phonetics and phonology of geminate consonants," in *The Phonetics and Phonology of Geminate Consonants*, H. Kubozono, Ed. Oxford: Oxford University Press, 2017.

[4] S. Kawahara, "The phonetics of obstruent geminates, sokuon," in *The Mouton Handbook of Japanese Language and Linguistics*, H. Kubozono, Ed. Berlin: Mouton Gruyter, 2015.

[5] A. Kochetov and Y. Kang, "Supralaryngeal implementation of length and laryngeal contrasts in japanese and korean," *The Canadian Journal of Linguistics*, vol. 62, no. 1, pp. 18–55, 2017.

[6] J. M. Scobbie, A. A. Wrench, and M. van der Linden, "Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement," *Proceedings of the 8th International Seminar on Speech Production*, pp. 373–376, 2008.

[7] J. M. Scobbie, E. Lawson, S. Cowen, J. Cleland, and A. Wrench, "A common co-ordinate system for mid-sagittal articulatory measurement," *CASL Working Papers, Queen Margaret University, WP-20*, 2011.

[8] D. Recasens and C. Rodríguez, "A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound," *Journal of Phonetics*, vol. 59, pp. 58–75, 2016.

[9] R. C. Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2017. [Online]. Available: http://www.R-project.org/

[10] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[11] J. Owens, *A Grammar of Harar Oromo (Northeastern Ethiopia)*. Hamburg: Helmut Buske Verlag, 1985.

[12] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages*. Oxford: Blackwell, 1996.

[13] I. R. Titze, *The Myoelastic Aerodynamic Theory of Phonation*. Iowa City: National Center for Voice and Speech, 2006.

[14] S. Hamann and S. Fuchs, "Retroflexion of voiced stops: Data from dhao, thulung, afar, and german," *Language and Speech*, vol. 52, no. 2, pp. 181–216, 2010.

[15] M.-R. Lloret, "Implosive consonants: Their representation and sound change effects," *Belgian Journal of Linguistics*, vol. 9, pp. 59–72, 1994.

[16] S. Ahn, "Tongue root contributions to voicing in utterance-initial stops in american english," *Proceedings of Meetings on Acoustics*, vol. 25, 2017.