ACOUSTIC PROPERTIES OF SINGLETON AND GEMINATE EJECTIVE STOPS IN TSOVA-TUSH

Bryn Hauk and Jacob Hakim

Department of Linguistics, University of Hawai‘i at Mānoa
bhauk@hawaii.edu, hakimj@hawaii.edu

ABSTRACT

This paper presents an analysis of singleton and geminate ejectives in Tsova-Tush, an understudied Northeast Caucasian language spoken in Zemo Alvani, Georgia. The Tsova-Tush phoneme inventory contrasts ejectives with aspirated stops at four places of articulation: /p’ t’ k’ q’/. Additionally, geminate stops contrast with singletons at two places of articulation: /t^h: t’: q^h: q’:/. Geminate ejective stops are cross-linguistically rare phonemes. The goal of this study is therefore to describe the acoustic properties of these stop contrasts, with special attention to the geminate ejectives.

Analyzed data included 500 segmented stops from high-quality recordings of six native speakers of Tsova-Tush. The results showed that geminates were significantly longer than singletons in closure duration, but did not differ in VOT. Ejectives had significantly shorter VOT than aspirated stops, as well a difference in f_0 and H_1*-H_2* in the following vowel, but with marked interspeaker variation in the latter measures.

Keywords: ejectives, geminates, acoustics, endangered languages, Tsova-Tush

1. INTRODUCTION

Tsova-Tush, also known as Batsbi or Bats [ISO 639-3: bbl] is a critically endangered Northeast Caucasian language spoken by a few hundred people in Zemo Alvani, Georgia. Tsova-Tush exhibits a phonemic contrast between ejectives and aspirated stops at four places of articulation, with geminate stops at two places of articulation, as shown in Table 1. Of particular interest are the two geminate ejectives, /t’:/ and /q’:/. Phonemic geminate ejectives are cross-linguistically rare, reported in only 13 of the 2,155 phoneme inventories listed in PHOIBLE [14].

The present study provides the first acoustic description of the contrast between aspirated and ejective stops in Tsova-Tush. We explored the following acoustic measures as potential correlates of the aspirated vs. ejective stop distinction: total duration, closure duration, VOT, length of the preceding vowel, and voice quality during the following vowel. We find that Tsova-Tush ejectives are characterized by shorter total duration, shorter VOT, and creakier voice quality at vowel onset.

This study included only voiceless stops. However, as shown in Table 1, Tsova-Tush has phonemic voiced stops as well, which will be discussed briefly in section 4.

Table 1: Tsova-Tush stops that contrast by airstream

	bilabial	coronal	velar	uvular
aspirated singleton geminate	p^h	t^h t^h:	k^h	q^h q^h:
ejective singleton geminate	p’	t’ t’:	k’	q’ q’:
voiced (singleton)	b	d	g	

Our study builds on a previous study comparing singleton and geminate stops in Tsova-Tush by Hauk [9]. Prior to that study, some descriptions of these Tsova-Tush stops had referred to the geminates as “intensive,” claiming that intensives were characterized by some other acoustic properties of intensiveness rather than duration alone. The study in [9] found that the so-called intensives (henceforth: geminates) were characterized only by longer total duration and longer closure duration. The other acoustic properties measured in [9] (VOT, length of the preceding vowel, intensity of the burst, intensity during the post-burst interval, and voice quality of the following vowel) did not differ for geminates, contra claims that these segments were somehow intensive.

Hauk’s study did not specifically target the contrast between aspirated and ejective stops; however, some effects were found. The ejectives at coronal (varying between dental and alveolar) and uvular places of articulation included in the study had a slightly shorter total duration, a shorter VOT, a shorter preceding vowel, and different voice quality in the following vowel, where the direction of the latter effect varied by speaker. However, Hauk’s study compared only those segments that contrasted in terms of intensiveness, excluding bilabial and velar ejectives, and was based on data from only three speakers. Therefore, these observations about ejectives can only be taken as preliminary.

Our study extends Hauk’s study by including the bilabial and velar stops to provide a complete picture of the contrast between aspirated and ejective stops at all places of articulation. Further, we include data from three additional speakers, which is crucial for investigating ejectives, given that the previous study indicated some interspeaker variation in the production of ejectives.

2. METHODS

2.1. Data collection

This study combines data from Hauk's study [9] with new data from three additional speakers. In [9], three Tsova-Tush speakers (one female, two male), were recorded pronouncing 62 target words in a carrier sentence /ɑs _ ɛɬnɑs/ 'I said _.' The recordings were made in identical recording conditions on August 9, 2017, inside a consultant's home in Zemo Alvani.

For the present study, three additional speakers (two female, one male) were recorded pronouncing 102 target words (i.e., the previous word list extended to contain more bilabial and velar stops) in the same carrier sentence. These recordings were made in July of 2018 at the consultants' homes in Zemo Alvani. All recordings were made using a Zoom H2 solid state recorder with an external lapel microphone recording at 48 kHz/24 bit.

The speakers whose productions were analyzed in this study are listed in Table 2.

Table 2: Speakers analyzed in this study

initials	sex	approx. age at recording	year recorded
NB	F	60	2017
RO	M	60	2017
RS	M	60	2017
KD	F	60	2018
PQ	M	93	2018
TQ	F	93	2018

2.2. Data preparation

The recordings were segmented in Praat [4] on the phoneme level and on a sub-phonemic phonetic level, identifying the components of each target phoneme (Table 1), in accordance with segmentation recommendations by [1, 8]. Measurements of duration were extracted via Praat scripts. Measures of voice quality (H_1*-H_2* and H_1*-A_3*) were collected via the VoiceSauce application [11] in MATLAB®.

2.3. Data analysis

The following measures were selected for analysis of the distinction in airstream and consonant length: total duration, closure duration, VOT, duration of the preceding vowel, and spectral tilt (H_1*-H_2*, H_1*-A_3*) of the following vowel. Measures of duration were predicted to be associated with gemination, while spectral tilt was expected to correlate with the distinction in airstream [13, 8]. However, each of these measures have the potential to be at least a secondary effect for either geminates or ejectives. For instance, in some languages (Italian [15], Japanese [10], Lebanese Arabic [2]), geminates can be produced with tighter adduction of the vocal folds, resulting in a creaky voice quality (i.e. lower spectral tilt) in the surrounding vowels.

Because not all places of articulation contrast in terms of consonant length, a subset of the available data comprising only singleton stops was used initially to explore whether airstream (aspirated vs. ejective) predicted the aforementioned measures. Additionally, a subset of the available data comprising only coronal and uvular stops was used to compare geminate ejectives (/t'ː q'ː/) with singletons and aspirated stops. The number of data points analyzed, therefore, varied across statistical models.

The data were analyzed using linear mixed effects regression models with the relevant acoustic measure as the dependent variable and either airstream (aspirated, ejective) or consonant length (singleton, geminate) as the independent variable, using the package `lme4` [3] in RStudio [17]. The fixed effects included place of articulation (bilabial, coronal, velar, uvular) for all models, as well as vowel height (low, mid) for analyses of spectral tilt. Owing to the small size of the datasets, no interactions were included. The random intercepts were by speaker and by word, with random slopes for airstream by speaker.

Deviation coding (via the function `contr.sum`) was used, comparing each level of the independent variables to the grand mean. P-values were calculated using the Kenward-Roger approximation in the `lmerTest` package [12]. RDI (pirate) plots were made using the package `yarrr` [16].

3. RESULTS

3.1 Total duration

In this study, total duration of stops was the sum of closure duration and VOT. Following a pause, stops lack a well-defined closure duration; preceding a pause, stops lack a well-defined VOT. Total duration was therefore undefined for these stops. Therefore, those segments were excluded from the relevant duration measures.

Among singleton stops contrasting by airstream, there were 335 tokens for which total duration could be defined. Linear regressions showed that the total duration of ejective stops was shorter than the grand mean ($p < 0.001$, $\beta = 42$ ms). Additionally, bilabial stops were found to be slightly longer than the grand mean (*p* = .035, *β* = 22 ms). Other places of articulation did not differ in total duration.

Among stops contrasting by length (i.e., only coronal and uvular stops), there were 335 tokens for which total duration could be defined. The total duration of geminate stops was found to be significantly longer than the grand mean (*p* < .001, *β* = 113 ms), while ejectives remained significantly shorter (*p* = .001, *β* = 45 ms).

3.2 Closure duration

Among singleton stops contrasting by airstream, there were 372 tokens for which closure duration could be defined. Linear regressions showed that the closure duration of ejective stops did not differ from the grand mean (*p* = .172, *β* = 7 ms). Only velar stops differed from the grand mean in this measure, with a slightly shorter closure duration (*p* = .031, *β* = 18 ms).

Among stops contrasting by length, there were 396 tokens for which closure duration could be defined. Linear regressions showed that the closure duration of geminate stops was significantly longer than the grand mean (*p* < .001, *β* = 114 ms), while ejectives again showed no difference.

3.3 VOT

Among singleton stops contrasting by airstream, there were 334 tokens for which VOT could be defined. Linear regressions showed that the VOT of ejective stops was significantly shorter than the grand mean (*p* < .001, *β* = 32 ms). The VOT of uvular stops was found to be slightly shorter than the grand mean (*p* = .041, *β* = 14 ms), while no significant difference in VOT was found for other places of articulation.

Among stops contrasting by length, there were 333 tokens for which VOT could be defined. Linear regressions found that the VOT of geminate stops did not differ from the grand mean (*p* = .970, *β* = 0 ms), while the VOT of ejectives was again confirmed to be shorter (*p* < .001, *β* = 26 ms).

3.4 Duration of preceding vowel

There were 238 phonemically short (or non-long) vowels in the data set that preceded aspirated or ejective singleton stops. Linear regressions showed that these vowels preceding an ejective stop did not differ in length from the grand mean (*p* = .323, *β* = 6 ms). Vowels were found to be slightly longer before coronal stops (*p* = .010, *β* = 28 ms), with no significant differences before stops at other places of articulation.

There were 371 phonemically short vowels that preceded singleton and geminate, coronal or uvular stops. Linear regressions showed that the duration of vowels preceding geminate stops did not differ from the grand mean (*p* = .665, *β* = 2 ms). However, the duration of vowels preceding ejectives was found to be shorter (*p* = .007, *β* = 12 ms). Although these results are consistent with the findings of the previous study in [9], it is curious that the difference in preceding vowel duration was found to be significant in this data subset (with singleton and geminate coronal and uvular stops), but not in the other subset (with only singleton stops at all four places of articulation).

3.5 Spectral tilt in following vowel

Spectral tilt was only measured in low or mid vowels (/ɑ, o, e/) immediately following a target consonant in the second syllable of a target word. We excluded high vowels because, even after corrections to amplitude measurements to account for different vowel qualities, the spectral tilt of high vowels differed too much to be meaningfully compared. Because the exclusion of high vowels left only 204 data points to compare, geminates were left in this dataset, and consonant length (singleton, geminate) was treated as an additional fixed effect.

Figure 1: RDI plot of H_1*-A_3* in following vowel by airstream (2017 data)

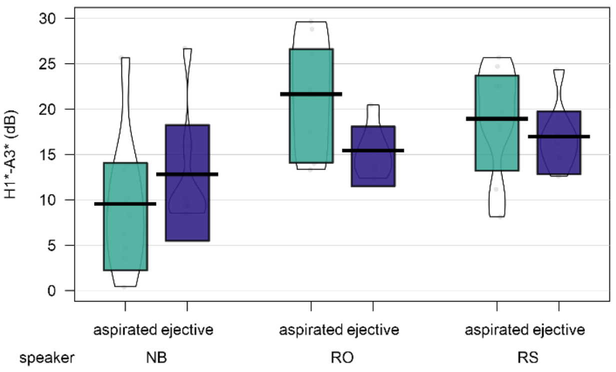
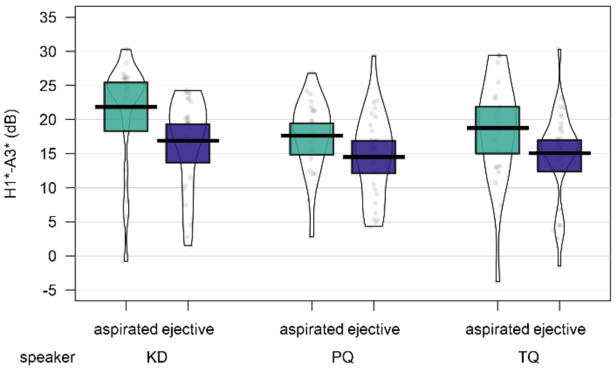


Figure 2: RDI plot of H_1*-A_3* in following vowel by airstream (2018 data)



For these 204 tokens, spectral tilt as measured by H1*-H2* in vowels following an ejective was found to be significantly lower than the grand mean (p = .038, β = 1.5 dB). This tendency was confirmed with a second measure for spectral tilt, H1*-A3*, which was again significantly lower for vowels following an ejective (p = .017, β = 3.8 dB). The latter model was also a better fit for the data (pseudo-R2 = .609, vs. .440 for H1*-H2*). Additionally, mid vowels had a lower H1*-A3* than the grand mean (p < .001, β = 5.4 dB), an effect that was not found of H1*-H2* for mid vowels. No significant effect of consonant length was found for either measure (for H1*-H2*, p = .999, β = 0 dB; for H1*-A3*, p = .140, β = 2.4).

Hauk's study in [9] had also found a difference in spectral tilt following an ejective, although the consistency and the direction of the effect differed for the three speakers analyzed, as shown in Figure 1; speaker NB had breathier vowels following an ejective. However, the new speakers added for the current study showed much less variation: for all three speakers, vowels were significantly creakier following an ejective, as shown in Figure 2. The wide boxes (representing the 95% Highest Density Interval) and wide spread of data points shown in these figures probably indicate that there is an additional factor conditioning a change in voice quality beyond the factors considered here.

This tendency for creaky voice in vowels following an ejective is similar to what is observed for ejectives cross-linguistically [8, 18].

4. DISCUSSION AND CONCLUSIONS

We found that Tsova-Tush ejectives were characterized by a shorter total duration, shorter VOT, and creakier voice quality in the following vowel, but not by closure duration or the duration of a preceding vowel, as summarized in Table 3. Further, this study confirmed the findings of a previous study [9] comparing singleton and geminate stops in Tsova-Tush with three additional speakers; as before, geminates differed only in total duration and closure duration, and not by any additional measure.

Table 3: Results for acoustic measures of ejective and geminate stops

	effect for ejectives	effect for geminates
Total dur.	**shorter**	**longer**
Closure dur.	no difference	**longer**
VOT	**shorter**	no difference
Preceding V	no difference	no difference
H_1*-H_2*	**lower**	no difference
H_1*-A_3*	**lower**	no difference

As noted above, in addition to aspirated and ejective stops, Tsova-Tush has voiced stops (/b d g/, which were not analyzed in this study. Some of the parameters measured here (especially VOT in the negative dimension [6] and duration of the preceding vowel) are certainly relevant to the contrast in voicing. An additional study including voiced stops would be revealing, especially if designed such that the target consonants were to follow both phonemically long and short vowels.

The findings of our study further highlight why the previous study [9] failed to find any acoustic correlates of the "intensiveness" of Tsova-Tush stops. Many of the acoustic properties associated with "fortis," "tense," or "strong" consonants in other languages (a difference in VOT, preceding vowel duration, burst intensity, and voice quality in the following vowel) are reserved in Tsova-Tush for the contrast between aspirated and ejective stops.

Table 4 summarizes the means of the measures examined in this study for each stop type. In this dataset, the closure duration of ejective geminates was roughly 2.2 times longer than that of ejective singletons, while the closure duration of aspirated geminates was roughly 2.1 times longer than that of aspirated singletons. The VOT of all aspirated stops was roughly 1.7 times longer than that of all ejectives. The Tsova-Tush ejectives observed here had a shorter VOT for all places of articulation than has been found cross-linguistically [5]. Based on this study, we suggest that Tsova-Tush ejectives are typologically most similar to the "slack" type of ejectives, characterized by shorter total duration, shorter VOT, and creakier voice quality at vowel onset, rather than "stiff" ejectives [13].

Table 4: Means and standard deviations of stops

	singleton		geminate	
	aspirated	ejective	aspirated	ejective
Total dur.	184 ms	141 ms	301 ms	256 ms
SD	42 ms	40 ms	73 ms	80 ms
Closure dur.	113 ms	99 ms	232 ms	215 ms
SD	31 ms	30 ms	57 ms	66 ms
VOT	72 ms	43 ms	70 ms	42 ms
SD	25 ms	27 ms	30 ms	34 ms
Preceding V	106 ms	124 ms	126 ms	115 ms
SD	23 ms	29 ms	28 ms	25 ms
H_1*-H_2*	6.5 dB	5.7 dB	4.4 dB	4.8 dB
SD	3.2 dB	2.8 dB	2.7 dB	3.5 dB
H_1*-A_3*	20.4 dB	16.9 dB	15.4 dB	15.4 dB
SD	6.5 dB	8.3 dB	6.3 dB	6.0 dB

This study therefore contributes the first acoustic description of ejectives in Tsova-Tush, as well as one of the first acoustic descriptions of geminate ejective stops. These cross-linguistically rare segments only exist in the phoneme inventories of understudied languages, such as Tsova-Tush, further highlighting the value of phonetic documentation of such languages.

5. REFERENCES

[1] Abramson, A., Whalen, D.S. 2017. Voice Onset Time at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics* 63. 75–86. doi:10.1016/j.wocn.2017.05.002.

[2] Al-Tamimi, Jalal & Ghada Khattab. 2018. Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops. *Journal of Phonetics* 71. 306–325. doi:10.1016/j.wocn.2018.09.010.

[3] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.

[4] Boersma, P. & Weenink, D. 2017. Praat: doing phonetics by computer. http://www.praat.org.

[5] Cho, Taehong & Peter Ladefoged. 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27(2). 207–229. doi:10.1006/jpho.1999.0094.

[6] Cho, Taehong, D.H. Whalen & Gerard Docherty. 2019. Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics* 72. 52–65. doi:10.1016/j.wocn.2018.11.002.

[7] Gordon, Matthew & Peter Ladefoged. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29(4). 383–406. doi:10.1006/jpho.2001.0147.

[8] Grawunder, S., Simpson, A., Khalilov, M. 2010. Phonetic characteristics of ejectives: Samples from Caucasian languages. In Fuchs, S., Zygis, M. Toda, M. (eds.), *Turbulent sounds: An interdisciplinary guide*, 209–244. Berlin/Boston: De Gruyter, Inc.

[9] Hauk, B. 2018. Tsova-Tush 'intensive' consonants. Presentation at *16th LabPhon*. Lisbon, Portugal.

[10] Idemaru, Kaori & Susan G. Guion. 2008. Acoustic covariants of length contrast in Japanese stops. *Journal of the International Phonetic Association* 38(2). 167–186. doi:10.1017/S0025100308003459.

[11] Iseli, M., Shue, Y.-L., Alwan, A. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *J. Acoust. Soc. Am* 121(4). 2283–2295. doi:10.1121/1.2697522.

[12] Kuznetsova, A., Brockhoff, P.B, Christensen, R.H.B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13). 1–26. doi:10.18637/jss.v082.i13.

[13] Lindau, M. 1984. Phonetic differences in glottalic consonants. *Journal of Phonetics* 12. 147–155.

[14] Moran, S, McCloy. D., Wright, R. (eds.). 2014. PHOIBLE Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://phoible.org/.

[15] Payne, Elinor M. 2006. Non-durational indices in Italian geminate consonants. *Journal of the International Phonetic Association* 36(1). 83–95. doi:10.1017/S0025100306002398.

[16] Phillips, N. 2017. yarrr: A Companion to the e-Book "YaRrr!: The Pirate's Guide to R." https://CRAN.R-project.org/package=yarrr.

[17] RStudio Team. 2015. RStudio: Integrated Development Environment for R. Boston, MA: RStudio, Inc. http://www.rstudio.com/.

[18] Warner, N. 1996. Acoustic characteristics of ejectives in Ingush. *Proc. of the 1996 Intl. Conf. on Spoken Language Processing*, 3, 1525–1528. Philadelphia, PA: IEEE.