Itelmen Ejectives

Xiaotian Wang and Jonathan David Bobaljik

Harvard University

Draft: July 2025

1. Introduction

This paper offers a preliminary study of ejective consonants in Itelmen (ISO 639-3 itl, previously also known as Kamchadal), a Chukotko-Kamchatkan language traditionally spoken on the Kamchatka peninsula in what is now the Russian Far East. Except as noted, all reference to Itelmen here is to the contemporary Western variety, spoken in the 20th century in 8 villages along the Okhotsk Coast. 18th and 19th century sources provide word lists for distinct Eastern and Southern varieties, but there are no audio recordings of these varieties, which are no longer spoken. Itelmen is the only Chukotko-Kamchatkan language–and the only language in the area– with an ejective consonant series. We have approached our description with an eye towards presenting sufficient information to situate Itelmen ejectives within a phonological and phonetic typology of ejectives. We offer a few inconclusive remarks at the end of the paper regarding the origin of Itelmen ejectives.

The articulation of ejectives involves mechanisms that are distinguishable from other pulmonic sounds. Ladefoged & Maddieson (1996: p.78) describe the production of ejectives as involving a closed glottis, an occluded oral cavity, a raised larynx, and a compressed vocal tract above the glottis. Ejectives occur in around 16% of the world languages, but languages that, like Itelmen, allow ejectives to occur at multiple positions in the word are held to be quite uncommon (Fallon 2002, Percival 2019).

Some previous works (Kingston 1985, Lindau 1984) have argued that ejectives fall into two types: strong/stiff and weak/slack ejectives. Several acoustic dimensions of these two types of ejectives vary cross-linguistically. According to Lindau (1984) and Kingston (1985), strong ejectives are easier to perceive with long voice onset time, a loud burst, modal or tense voice on the following vowel, and raised pitch at the onset of voicing; weak ejectives, on the other hand, do not have a silent period between the oral burst and glottal release and are associated with creaky voice and lowered pitch at the onset of voicing, making them much harder for perception. Table 1 adapted from Ham (2007) summarizes the acoustic correlates of strong and weak ejectives reported in previous works (Lindau 1984; Kingston 1985; Bird 2002; Wright et al. 2002; Hargus 2011).

Correlates	Strong Ejectives	Weak Ejectives
VOT	long	short
Burst	loud	quiet
Voice Quality	modal or tense	creaky
Pitch at Voicing Onset	raised	lowered

Amplitude at Voicing Onset	sudden increase	gradual increase
Ease of Perception	easy	hard

Table 1: The Acoustic Correlates of Strong and Weak Ejectives

We find that Itelmen ejectives pattern mostly as strong under this classification, although with some qualifications. Very long VOT and strong burst intensity (for stops but not affricates) are consistent properties of the ejectives, but we note that there is substantial inter-speaker variation regarding vowel quality, even in our sample of three speakers.

1.1. Background

The consonantal phoneme inventory of (Western) Itelmen, given in Table 2. We include the glottal stop as a distinct phoneme. There is variability in the realization of this element, and a lack of consensus among the few authors who have considered its status in the language. Volodin 1976:34-39 considers, but ultimately leaves open, whether the glottal stop is phonemic, and Asinovskij 2002:246-7 (primarily on the basis of a phonetic analysis of related Chukchi) considers the glottal stop to be suprasegmental across the Chukotko-Kamchatkan family, including in Itelmen: one (but not the only) manifestation of laryngealization. Resolving this issue is beyond the scope of this paper.

	Bilabial	(Post)-Alveolar	Palatal	Velar	Uvular	Glottal
Plain	p	t		k	q	ʔ
Ejective	p’	t’		k’	q’	
Nasal	m	n	ɲ	ŋ		
Trill		r				
Fricative	ɸ β	ɬ	s z	x (ɣ)	χ	
Affricate			tʃ,tʃ’			
Lateral		l	lʲ			
Approximant			j			

Table 2: Itelmen Consonant Inventory

Another point of uncertainty in the phonemic inventory is the voiced velar fricative (in parentheses in Table 2). For the labial and velar fricatives, both voiced and voiceless, a range of pronunciations is attested and there is inter- and intra-speaker variation in the pronunciation of (some) words: ɸ ~ x^{w} ~ x and correspondingly β ~ ɣ^{w} ~ ɣ (~ ɰ). Whether there is a two- or three- way phonemic contrast for the voiced and the voiceless fricatives remains unsettled. The official orthography recognizes only two voiceless fricatives (ɸ, x) and one voiced (β), but Volodin 1976 (and K.N. Khalojmova, personal communication) acknowledge ɣ^{w} or ɣ as a separate phoneme. As this is orthogonal to our study, we have simply followed the dictionary form in all transcriptions below, unless otherwise noted.

The primary point of interest for present purposes is the ejective series: stops: /p’,t’,k’,q’/, and affricate /tʃ’/, which occur alongside a corresponding series of voiceless oral stops /p,t,k,q/, and the affricate /tʃ/. We perceive the plain (non-ejective) voiceless stops as aspirated, which is consistent with VOT measurements reported below, where the non-ejective stops would be classified as long-lag, i.e., aspirated, under the schema in Chodroff et al. 2019 (average of average VOT across bilabial, alveolar, and dorsal stops ≥35ms). We note that this differs from Russian, for which voiceless stops are unaspirated (short-lag, average of average VOT < 35ms), a fact that is relevant due to concerns about Russian influence on the pronunciation of Itelmen by bilingual speakers (see below). That is, the primary contrast in Itelmen stops and affricates is voiceless aspirated (long-lag VOT) versus voiceless ejective (very long VOT).

The plain-ejective contrast is generally easy to perceive in careful speech, although it is sometimes obscured in connected speech. In less frequent words, sources sometimes differ in transcribing an ejective or plain stop, and there is some amount of speaker variation. Voiced stops do not occur in the (etymologically) native vocabulary, but centuries of contact with Russian have introduced a large number of loanwords with voiced stops, such that any sizeable corpus of Itelmen speech will also contain voiced stops.[1]

Ejectives may occur in any position in a word (initial, medial or final, though not all are readily found in all positions) and may occur internal to consonant clusters (Itelmen tolerates complex clusters, though to a first approximation disallows sonorants from cluster-medial position; see Volodin 1976, Bobaljik 1998), as well as in pre-vocalic position. A given word or morpheme may have more than one ejective. As far as we know, we find no evidence for the kinds of laryngeal cooccurrence restrictions investigated in Gallagher 2010 and much other work. A sample of ejectives in varying positions in words is given in (1).

(1) Itelmen ejectives

	p’	t’	k’	q’	tʃ’
Initial /_V					
	p’etʃ	t’iʔn	k’aβatʃ	q’otʃ’miŋsx	tʃ’inʲekeʔnʲtʃ
	‘child’	‘relatives’	‘back’	‘call me’	‘flower’ (DIM.PL)
Medial /_V					
	p’axp’aŋ	t’it’im	etʃk’elaχ	qeɬq’elekas	kitʃ’e
	‘boiled fish’	‘relative’	‘bad’	‘complain’	‘stag,bull’
Final					
	n/a	t'ot’ot’	k’ik’	q’iq’	itʃ’
		‘sandpiper’	‘throat’	‘ice hole’	‘birch’
In clusters					
	p’k’akas	t’sal	k’nəntx	q’lekas	tʃ’mamɬkas

[1] Voiced stops also occur in 18th and 19th century transcriptions of Eastern and Southern Itelmen, but their phonemic status is uncertain.

‘hide’	‘fox’	‘who (PL)’	‘get lost’	‘cry (as child)’
q’əmp’q’ax	met’sk’aj	ənk’sx	qt’k’nixtʃik	xiʔəmtʃ’laχ
‘cut it open’	‘bear’	‘birchbark’	‘press it’	‘very bitter’

It bears mention that an ejective and a stop-glottal stop sequence are distinct, although some authors write both as C’, leading to a potential ambiguity. Thus: *atʃinʲtʃ-ʔan* ‘yesterday’s’ and *k’-iɬtʃ-ʔin* ‘silent one’, both with the attributive suffix *-ʔan~-ʔin* following plain tʃ, are distinct from ejective tʃ’ in *kitʃ’e-ʔn* ‘stag-PL’ and likewise the [t] in *k’-ajtat-ʔan* ‘chased out’ is clearly not ejective, contrasting with ejective [t’] in *t’-aŋja-z-βin* ‘I praise you.’

Ejectives also alternate with plain consonants as allomorphs of certain prefixes. To a first approximation, inflectional prefixes consisting of a single stop, including 1SG *t-*, participial/nominalizer *k-*, and imperative/irrealis *q-* have an ejective allomorph when immediately preceding a (full) vowel or a prevocalic nasal and are non-ejective otherwise. Some examples are given in (2).

(2) Ejective allomorphs of prefixes

a. 1SG t-/t’-

t’-i-s-kitʃen	1SGS-go-PRES-1SGS	‘I’m going’	cf. tinu-ʔn	‘relative-PL’
t’-nu-a-s-kitʃen	1SGS-eat-DESID-PRES-1SGS	‘I want to eat’	cf. tnum-ke	‘down-DAT’
t-k’oɬ-kitʃen	1SGS-come-1SGS	‘I came’		

b. participial k-/k’-

k’-aβa-knen	PRTC-rise-SG	‘it rose’	cf. kaβe-laχ	‘brittle-ADJ’
k’-omt-ʔin	PRTC-tie-3SG	‘tied’	cf.koma-knen	‘footwear-NOM’
k’-nita-ʔin	PRTC-soul-ATTR	‘wise’	cf. knin	‘your’
k-tʃanzo-qzo-ʔan	PRTC-follow-ASP-3SG	‘he followed him’		

While there is some variation among sources, and probably among speakers, regarding where the ejective prefixes occur, in particular for the pre-nasal and pre-schwa position, the examples we examined in preparing this paper are clearly ejectives in these positions. Note that plain stops are, as illustrated, otherwise licit in the environments that trigger the ejective allomorphs of the prefixes, when these are tautomorphemic sequences.

Table 3 presents our counts of the distribution of paired (plain:ejective) obstruents in word-initial, medial, and final position in the headwords in the *Comprehensive Itelmen Dictionary, 1st edition* (Volodin et al. 2021).

	Initial	Medial	Final	Total		Initial	Medial	Final	Total
p	254	408	32	694	p’	14	11	0	25
k	630	1155	135	1920	k’	118	277	16	411
tʃ	300	724	169	1193	tʃ’	125	168	5	298
q	230	381	136	747	q’	10	27	6	43
t	468	1559	109	2136	t’	47	123	1	171
	1882	4227	581	6690		314	606	28	948

Table 3: Distribution of selected segments

One notes immediately that ejectives are infrequent compared to the corresponding plain obstruents. Also, ejectives in this sample are relatively rare in word-final position (3% compared to 9% of the corresponding plain obstruents.[2] For the record, we note that Asinovskij 2002:223-4 reported a different distribution over a (smaller but overlapping) corpus. While his sample also had fewer ejectives (n=227) than plain stops overall (n=2114), and he also finds an initial:medial ratio of roughly 1:2 for plain stops, ejectives in his sample show a stronger preference for initial position than in our sample: approximately half of the ejectives in his sample are in initial position, and 48 of the 55 tokens of ejective affricates in Asinovskij’s corpus are word-initial. We have no explanation for this difference.[3]

2. Methodology

2.1 Speakers and materials

For the quantitative analysis reported here, we made use of existing recordings from three speakers: the language activist Klavdia Nikolaevna Khaloimova (1934-2018), Galina Afanasievna Zaporotskaya (1935-2019), and Georgi Dmitrevich Zaporotskij (1937-2014). K.N. Khaloimova was born in Sopochnoe, and G.A. Zaporotskaya and G.D. Zaporotskij were born in the neighbouring village of Moroshechnoe. Both villages are part of the southern (sub-)dialect group of Western Itelmen. The recordings were made over the period from 2011-2013, by Tatiana Degai and Chikako Ono, primarily under the auspices of the Itelmen Audio-Visual Dictionary Project (PIs: J.D. Bobaljik, D. Koester), for which Chikako Ono was the primary editor (see Volodin et al.

[2] The bilabial ejective does not occur at all in final position, but this is not a significant gap, since bilabial ejectives are relatively infrequent.

[3] It is conceivable that Asinovskij’s corpus included more forms with inflectional prefixes (which have ejective allomorphs), whereas our count includes more (uninflected) headwords from the dictionary. Note also that the dictionary was compiled across a variety of sources, by many transcribers, not all of which are consistent in transcribing ejectives. <If we have online supplemental material, we can present numbers for this>

2021). The recordings of Khaloimova and Zaporotskaya used in this study can be found in the online version of the dictionary at https://itelmenslovar.org.

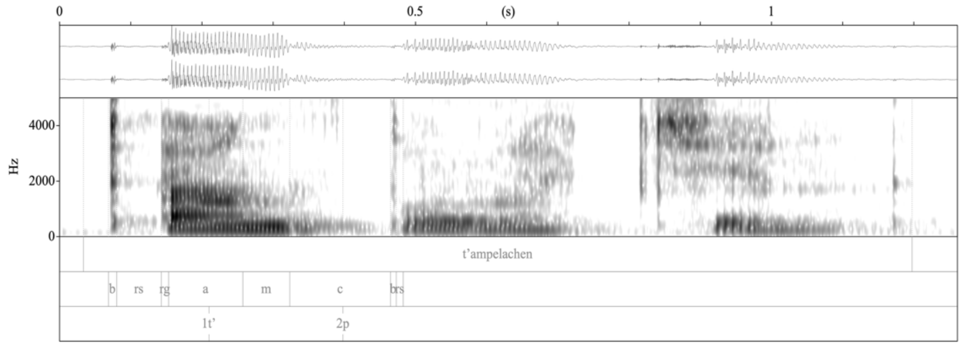
These speakers represent the last generation of fluent speakers of Itelmen exposed to the language as children. Since at least the forced resettlements and razing of native villages in the 1950s (and for many communities even earlier) all Itelmen speakers have been bilingual, and Russian has been their primary language of daily communication at least from school age onwards. In the later Soviet period, children were not permitted to speak Itelmen in schools. In some villages, including Moroshechnoe, Itelmen was still widely spoken until the 1950s, and speakers 2 and 3 reported using Itelmen in the home. Interviews with elders in the 1990s suggest that language shift (to Russian) began earlier in Sopochnoe, and use of the language was more restricted in village and home life even outside of the school. While K.N. Khaloimova (speaker 1) was the leading figure of the language revitalization (and documentation) efforts starting in the late 1980s, it appears from what we can glean from interviews that Russian may have been her first or dominant home language growing up, even though she was born in an Itelmen village.[4] Our recordings represent the Itelmen language of that generation, possibly with effects of bilingualism and language attrition. As a small control, we note below that there may be a difference between the mean VOT of voiceless stops between native Itelmen vocabulary and Russian loanwords even for these speakers, suggesting that there is a measurable difference between their Russian and Itelmen phonetic systems. As our focus is the ejectives, which are lacking entirely in Russian, we set these concerns aside but acknowledge that the variety we consider may differ from the speech of monolingual or dominant Itelmen speakers, sadly no longer accessible.

A further qualification is that the recordings we made use of were intended for an online dictionary. The speakers read words or short phrases aloud one by one specifically for the purpose of providing audio recordings for the multimedia (spoken) version of the dictionary. We acknowledge that this represents a genre that differs from connected/naturalistic speech, which we have not investigated.

2.2 Data

Figure 1 below displays a sample spectrogram with annotations following the instructions of the the Praat script used (Percival, 2024). Stop releases are divided into several components, including burst (labeled as b), silence (rs), and glottal release (rg) for ejective stops. Initial stops are labeled with points on the upcoming vowel, and medial stops are marked on the preceding vowels. The numbering indicates the position of the stop with 1 being initial.

[4] In addition to interviews conducted by the second author, we draw here, with thanks, on interview material and discussions with David Koester, Ingrid Summers, and Tatiana Degai.



.

Figure 1: Sample Spectrogram of the Target Word *t'ampelatʃen*

A total of 290 stops and 59 affricates were selected for data analysis. We tried to balance the number of stops across places of articulation. However, as shown above in Table 3, the distribution of ejective and non-ejective stops is in general quite unbalanced. Other than the alveolar and velar stop series, the ejective stops at the other places of articulation are relatively less common compared to their aspirated counterparts. Also, bilabial stops are rare, especially bilabial ejective stops. These observations are reflected here in our target word/segment selection.

	Bilabial	Alveolar	Velar	Uvular	Affricate	Total
Non-ejective	p	t	k	q	tʃ	226
	21	45	69	48	43	
Ejective	p'	t'	k'	q'	tʃ'	123
	10	35	44	18	16	
Total	31	80	113	66	59	349

Table 4: Consonants Extracted for Data Analysis

2.3 Measurements and Data Analysis

Annotations and measurements were made in Praat (Boersma & Weenink, 2024) using a script developed by Percival (2024). The script measures a wide range of parameters related to stops and affricates, such as burst, VOT, vowel formants at the 10%, 50%, and 90% points, H1-H2, and so on. All measurements are taken directly from the output of the script. Each stop and the following vowel (if available) were segmented and annotated following the script instructions (see Figure 1 above). Both F0 and voice quality are measured with perturbation following Wright et al (2002) and Percival (2019). The formulas used are shown in Table 5 below.

F0 Perturbation	mean F0 at voicing onset – mean F0 at vowel midpoint
H1*-H2* Perturbation	mean H1*-H2* at voicing onset – mean H1*-H2* at vowel midpoint

Table 5: Perturbation Formulas

With the limitations of the data discussed above, we present graphs derived from both raw data extracted from each individual speaker and z-normalized data (the underlying data is available at: link-to-be-provided). The transformed z-score describes each individual value relative to the entire set of measurements (Glass & Hopkins, 1995). This method, adopted from Ham (2007), is intended to help minimize the inter-speaker variation for a phonetic study with a relatively small number of tokens like ours.

While all instances of ejectives are drawn from the etymologically native Itelmen vocabulary, in the first pass, we selected plain stops from the dictionary without careful attention to the etymological source of the words. Thus, in addition to etymologically Itelmen examples, our tokens of plain stops originally included Russian loans with varying degrees of integration. Examples include: *delegat* 'delegate' (no evidence of integration, note voiced [d,g]), *avgust-ank* 'in August' (root violates Itelmen vowel harmony and contains Russian phonemes [v,g], but combines with productive Itelmen locative suffix), *tʃaska-ʔn* 'cups' (etymologically Russian root *tʃaška*, but with accommodation to Itelmen phoneme inventory: *š* → *s*, as well as Itelmen plural suffix), and *tʃasit* 'now' (clearly from Russian, cf. *tʃas* 'hour', *sejtʃas* 'now', but not a contemporary Russian word). We report in section 3.1 only measurements from stops in etymologically Itelmen words. We include measurements for ejective affricates relative to non-ejective affricates due to the general rarity cross-linguistically of ejective affricates (i.e., only 6% of the 2186 languages included in the UCLA Phonological Segment Inventory Database have /tʃʼ/ in the phoneme inventory (Moran & McCloy, 2019)), but treat the stops and affricates in separate sections since some measurements, such as VOT, cannot be directly compared across the two.

3. Results

3.1 Stops

3.1.1 Release Duration

Following previous studies (Percival 2019, Ham 2007), the release duration or VOT (voice onset time) is measured from the release of the burst to the onset voicing of the following vowel for word-initial and word-medial pre-vocalic stops. For word-final stops, the release duration was measured from the release of the burst to the end point at which the noise of the release was no longer visible in Praat.

Excluding all loanwords and affricates, the release duration of ejective and non-ejective (labeled as VA) stops is shown in Figure 1 below. The mean release duration of ejective and non-ejective stops produced by each speaker is summarized in Table 6. Speaker 3 produced stops with the shortest release duration while the first speaker produced the longest. The voiceless aspirated stops from the first speaker have an especially long release duration. Across all speakers, the ejective stops have consistently have a significantly longer VOT than the non-ejectives.

	Speaker 1	Speaker 2	Speaker 3
Ejective	117.1 ms	114.1 ms	92.0 ms
Voiceless Aspirated	87.6 ms	56.4 ms	55.2 ms

Table 6: Mean VOT of Stops

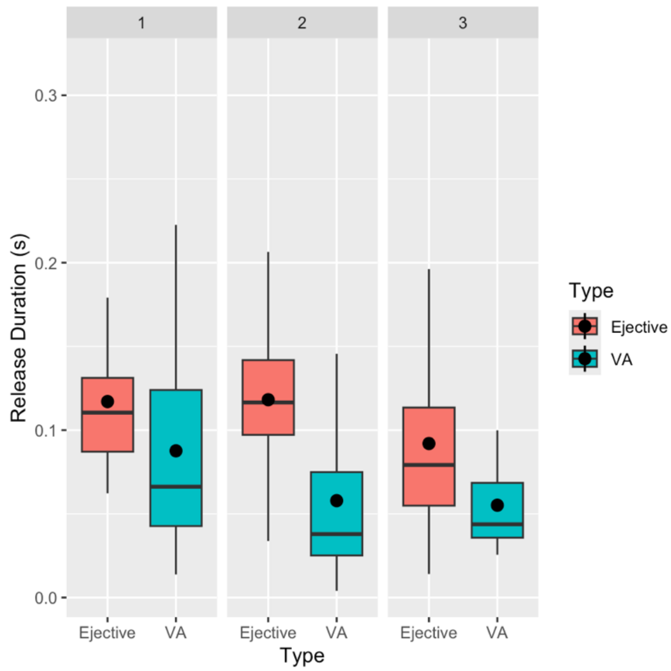


Figure 2: Release Duration of Stops

The boxplot below visualizes the z-normalized VOT across the two stop series. Outliers are removed from the plot. The line represents the median, and the dots correspond to the mean. The mean z-score for ejective stops is positive while that for voiceless aspirated stops is negative, suggesting that the patterns are different. The result from the one-way ANOVA test indicates a significant difference in release duration according to laryngeal type (F(1)=63.95, p<0.001).

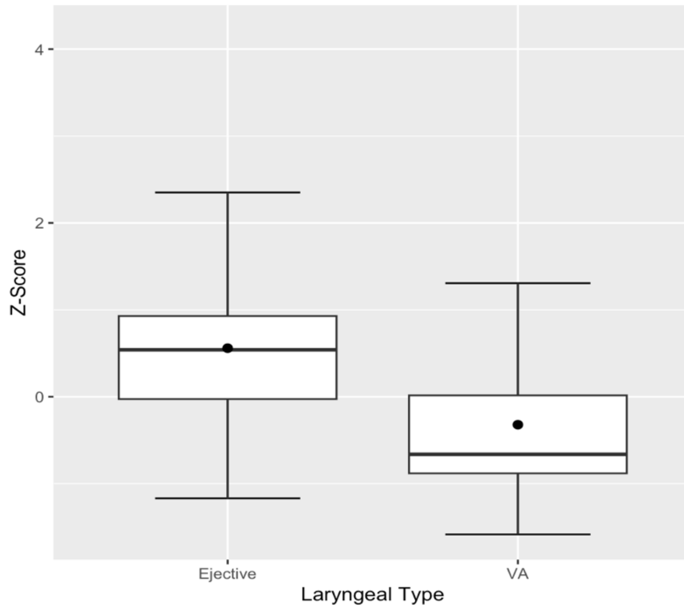


Figure 3: Z-normalized VOT

We noted above that all recorded Itelmen speakers are Russian dominant bilinguals. We excluded Russian loanwords in the measurements above. Interestingly, though we have examined far less data, it does appear that while Itelmen speakers pronounce voiceless non-ejective stops as aspirated (mean VOT across positions > 35ms) in Itelmen words, corresponding stops are (mostly) non-aspirated in Russian loanwords. We did not have representative stops in loanwords from all places of articulation for all speakers, but where we did have Itelmen-loanword pairs, we report the mean VOT for non-ejective stops in Table 7: for bilabial and alveolar stops, and for the mean across position (excluding uvulars, which Russian lacks), the VOT values for stops in Russian loanwords are similar to those reported elsewhere for Russian, namely short-lag (unaspirated) (Chodroff et al. 2019, supplemental data), whereas those for etymologically Itelmen words are substantially longer.

	Bilabial (P2)	Alveolar (P1)	Velar (P1&P2)	Mean of means across positions
Native Itelmen	61.1 ms	49.8 ms	61.8 ms	57.6 ms
Russian loanword	28.6 ms	18.6 ms	56.9 ms	34.7 ms

Table 7: VOT in loanwords vs Itelmen plain stops

3.1.2 Burst Intensity

Burst intensity was measured as the mean burst intensity of the target stop from the onset to the offset. The values were taken directly from the output of the Praat script. The mean burst intensity values of ejective stops and voiceless aspirated stops produced by all three speakers are demonstrated below. The stops collected from the third speaker are generally louder than those of the first two speakers. However, the pattern is consistent – the ejective and voiceless aspirated stops are similar in terms of mean burst intensity, but the mean burst intensity of the ejectives is for each speaker consistently slightly higher than for their aspirated stops.

	Speaker 1	Speaker 2	Speaker 3
Ejective	54.8 dB	55.9 dB	73.0 dB
Voiceless Aspirated	52.2 dB	54.9 dB	69.9 dB

Table 8: Mean Burst Intensity of Stops

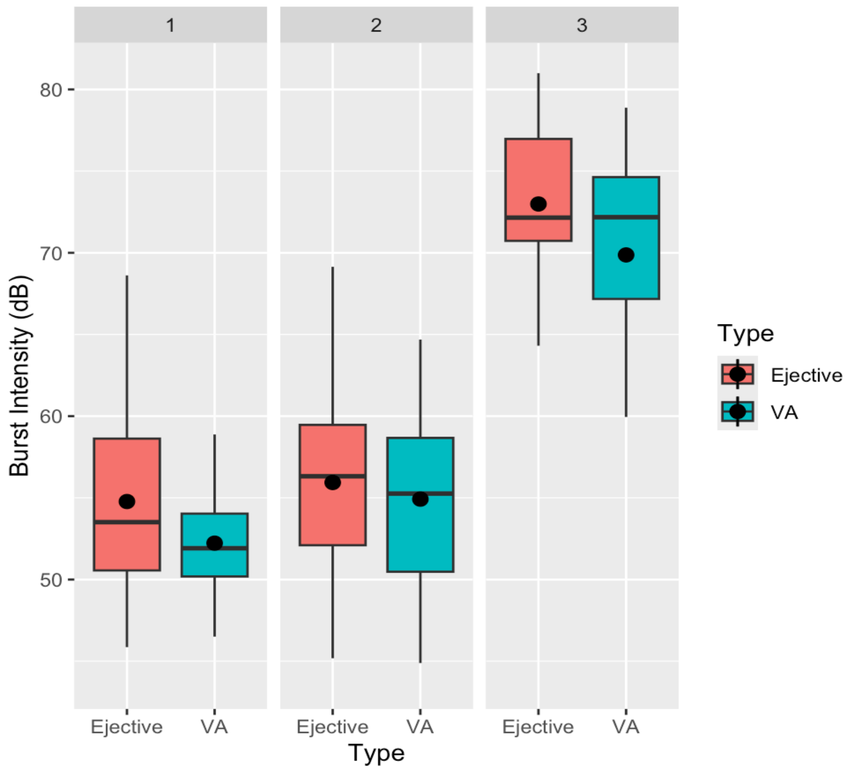


Figure 4: Mean Burst Intensity of Stops

The same pattern can be seen in the z-normalized mean burst intensity. The range of the box is similar across stops, with the ejectives having higher mean burst intensity. The mean z-score around 0 indicates that the values are close to the average. Nevertheless, we found a statistically significant difference between the mean burst intensity values of ejective and voiceless aspirated stop through a one-way ANOVA test ($F(1)=7.475$, $p<0.01$).

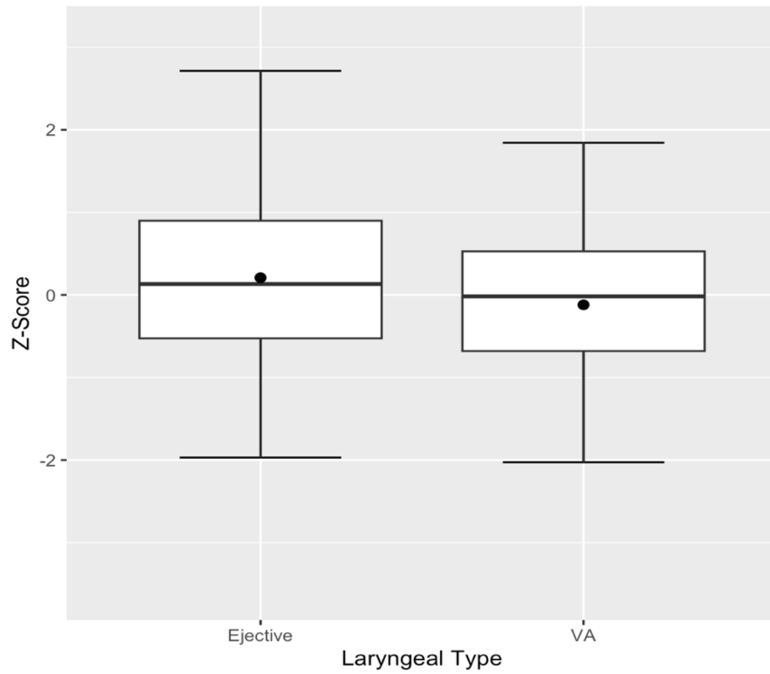


Figure 5: Z-normalized Mean Burst Intensity

3.1.3 F0

F0 was measured as perturbation to determine whether there is any difference in pitch between the voicing onset (10% of the vowel) to the 50% point of the vowel. The perturbation was calculated by subtracting the mean F0 at the onset from that at the midpoint (onset F0 – midpoint F0). Figure 5 shows the data extracted from all three speakers and the mean F0 perturbation (including the outliers). For two of the three speakers, both ejectives and voiced aspirates have a profile where the F0 value at voicing onset is generally higher than that at vowel midpoint, but across all speakers, the ejective stops have, on average, a higher F0 perturbation than voiceless aspirated stops.

	Speaker 1	Speaker 2	Speaker 3
Ejective	10.0 Hz	8.2 Hz	5.2 Hz
Voiceless Aspirated	2.6 Hz	4.8 Hz	-0.8 Hz

Table 9: Mean F0 Perturbation of Vowels following Stops

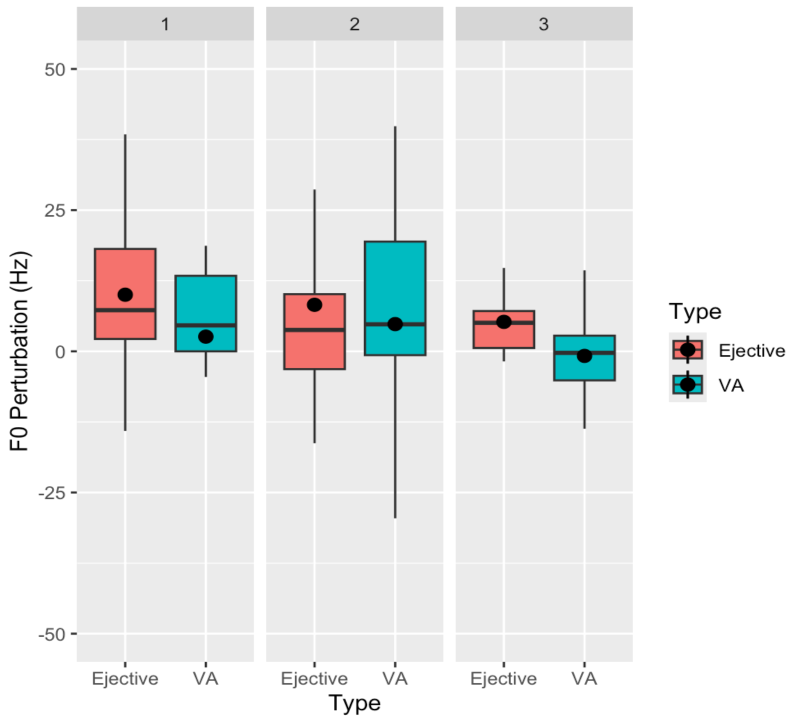


Figure 5: F0 Perturbation of Vowels following Stops

The z-normalized F0 perturbation is displayed in Figure 6. The z-score values across the two stops are close to each other. Both values are around 0. The z-normalized F0 perturbation according to laryngeal type is also statistically significant ($F(1)=5.663$, $p<0.05$).

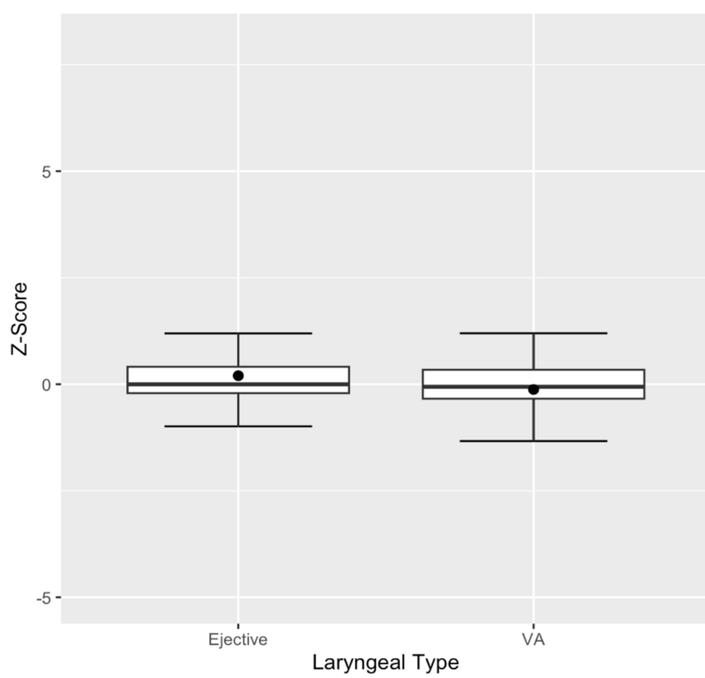


Figure 6: Z-normalized F0 Perturbation

The picture is less clear when we look at each speaker individually. Figure 7 below plots the mean F0 at the 10%, 50%, and 90% points of the vowel following ejective and voiceless aspirated stops, visualizing the change of F0 over time. For Speaker 1 and 2, the contour shapes across the two stops are similar. The F0 always starts at a higher point and gradually decreases until it reaches the lowest point at the 90% point of the vowel. Also, the vowels following ejective stops have a higher F0 than non-ejective stops for Speaker 1. No noticeable difference is seen in the contour of the two stops produced by Speaker 2. With Speaker 3, the reverse pattern is observed where the vowels when preceded by voiceless aspirated stops have a slightly higher F0. The lowering trend is also only seen in the ejective stops.

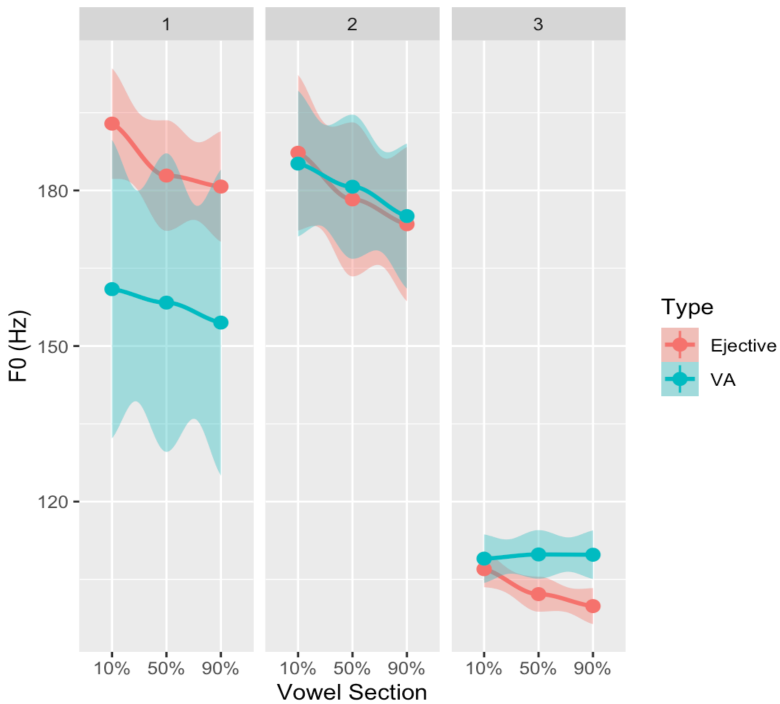


Figure 7: Mean F0 at the 10%, 50%, and 90% Points of the Vowels following Stops

3.1.4 H1*-H2*

H1-H2, one of the spectral tilt measures referring to the difference in the amplitude between the first and second harmonics, is used widely to distinguish contrastive phonation types (Chai & Garellek, 2022). The measurement reflects glottal spreading/constriction from a voice articulation perspective. H1-H2 values lower continuously from breathier, to modal, to creaky phonation types (Garellek, 2022); a negative H1-H2 is considered one of the correlates of creaky phonation (Keating et al., 2015; Percival, 2019). H1-H2 after formant correction is used here to control for the effects of vowel quality (Garellek, 2022).

The H1*-H2* perturbation was measured by subtracting the initial corrected H1-H2 from the midpoint mean. The mean H1*-H2* values are presented in the table below. Aside from Speaker 1, the other two speakers both had a higher H1*-H2* perturbation for voiceless aspirated stops than ejective stops. However, even with the reversed pattern seen in Speaker 1, the positive values across the board (except for Speaker 3's ejective stops) indicate that the vowels are generally more modal at voicing onset than at midpoint.

	Speaker 1	Speaker 2	Speaker 3
Ejective	4.6 dB	1.9 dB	-0.3 dB
Voiceless Aspirated	2.1 dB	2.9 dB	1.2 dB

Table 10: Mean H1*-H2* of Vowels following Stops

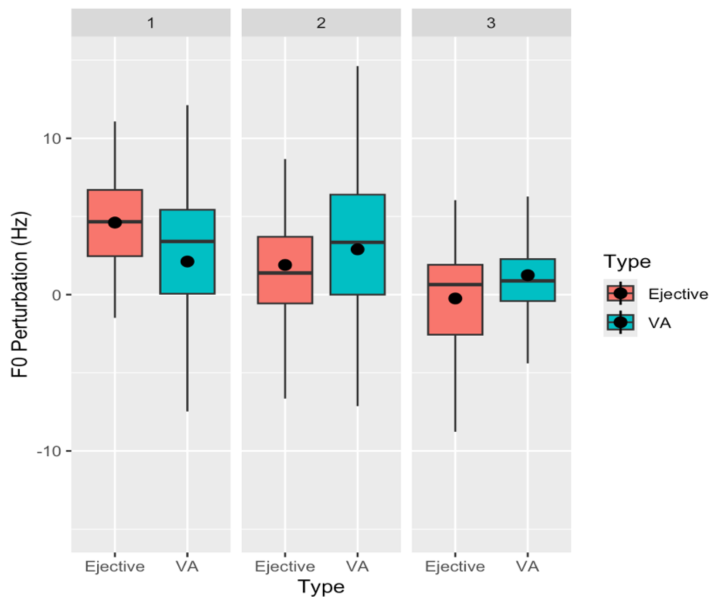


Figure 8: H1*-H2* Perturbation of Vowels Following Stops

The z-score close to 0 reflects the distribution of the values. The values across all three speakers are close to the average value. The ANOVA test result suggests that the difference is not statistically significant (F(1)=0.472, p=0.493).

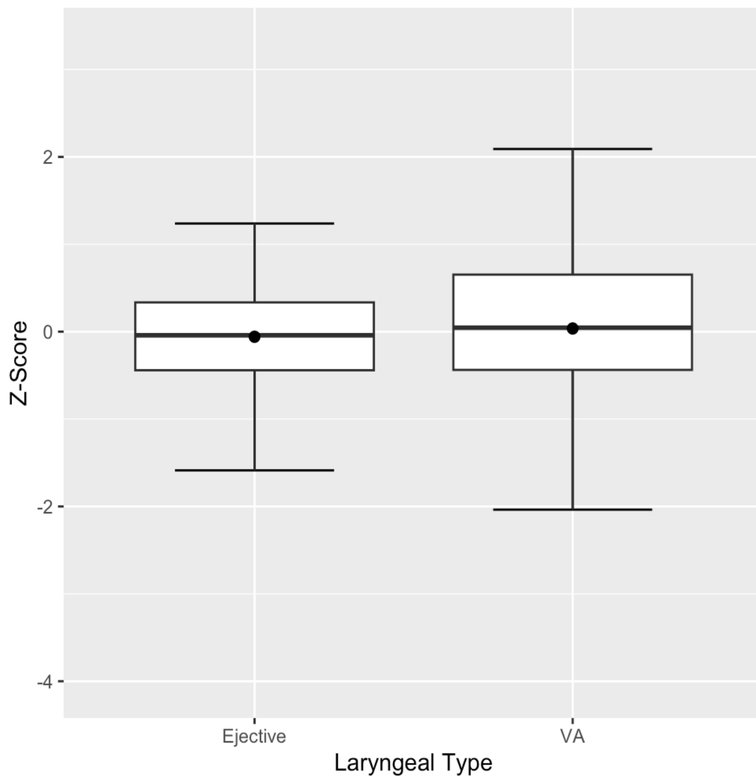


Figure 9: Z-normalized H1*-H2* Perturbation

The figure below presents the change of H1*-H2* across different sections of the vowels. At vowel onset, the H1*-H2* of vowels following voiceless aspirated stops is higher than ejectives for Speaker2 and 3; however, a different pattern is found in Speaker 1, where the onset of the vowel is much more modal/breathier when preceded by ejective stops. The overlapping area between the two stops is substantial for two out of the three speakers. The vowels following ejectives produced by the first speaker have a higher H1*-H2* before 90% of the vowel.

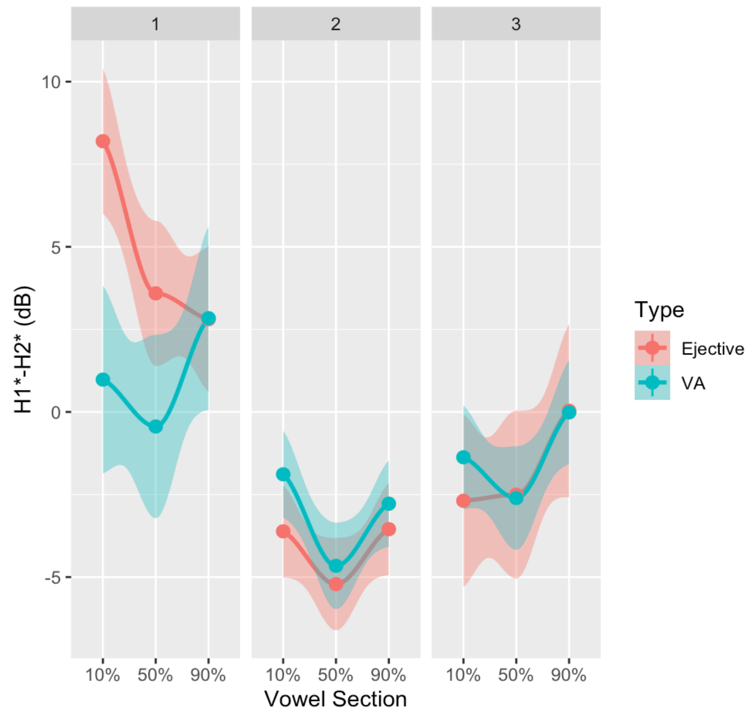


Figure 10: Mean Corrected H1-H2 at the 10%, 50%, and 90% Points of the Vowels following Stops

3.2 Affricates

We also examined the acoustics of both ejective and non-ejective affricates. Similar to stops, release duration, intensity, and the following vowel quality were considered.

3.2.1 Release Duration

Following Abramson & Whalen (2017), the VOT of affricates is measured as the duration from the end of the frication to the onset of the voiced part. Additionally, the frication duration is also included in the analysis to enhance the compatibility between our results and those of Kye (2024). The mean values of VOT are shown below in Table 11. The ejective affricates have a longer VOT than the non-ejectives for both speakers. Note that the non-ejective voiceless affricates have a much lower VOT than those reported in Kye (2024).

	Speaker 1	Speaker 2
Ejective	81.4 ms	61.5 ms
Non-ejective	1.3 ms	8.2 ms

Table 11: Mean VOT of Affricates

The following figure visualizes the raw VOT data of both speakers (we had no contrasting affricate pairs for Speaker 3), confirming the observation that ejective affricates have a longer VOT than non-ejectives. The VOT of non-ejective affricates is close to zero.

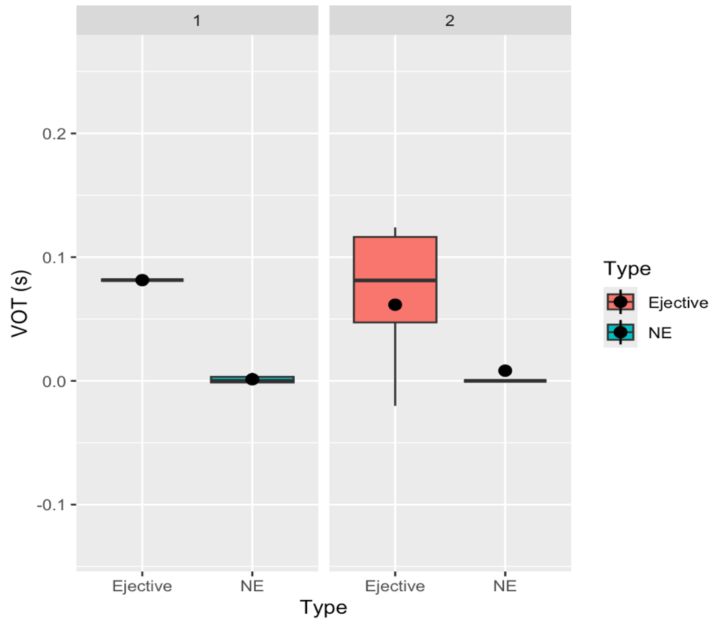


Figure 11: VOT of Affricates

The z-scored VOT below indicates that while the VOT of ejective affricates is likely to be higher than the mean value, that of non-ejective affricates tends to be lower. The difference between ejective and non-ejective affricates is statistically significant ($F(1)=13.77$, $p<0.001$)

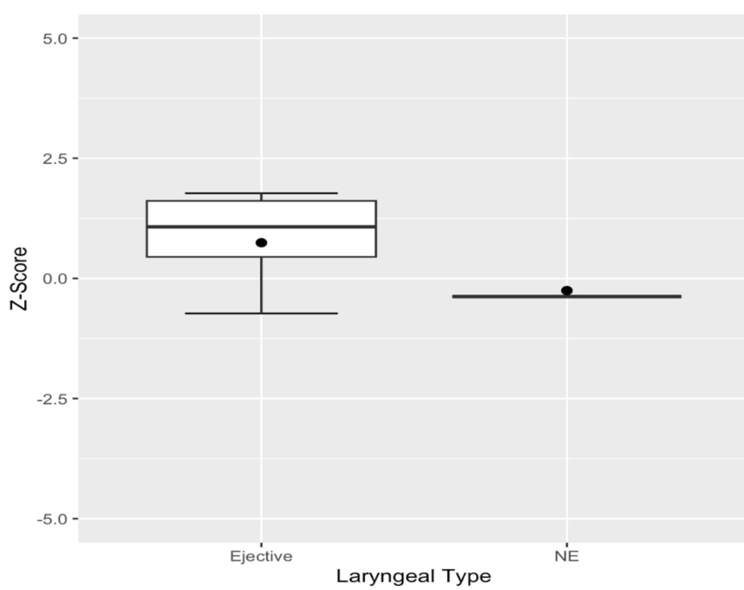


Figure 12: Z-normalized VOT

The mean frication duration values of affricates are displayed in the table below.

	Speaker 1	Speaker 2
Ejective	57.9 ms	70.7 ms
Non-ejective	65.9 ms	99.0 ms

Table 12: Mean Frication Duration of Affricates

The opposite pattern is seen regarding frication release duration, where the non-ejective (voiceless) affricates have a higher frication duration than ejectives.

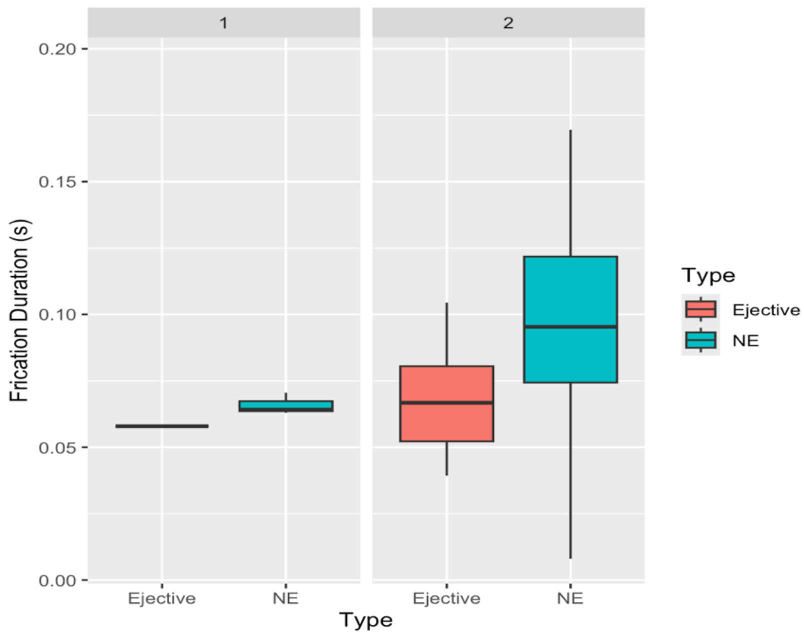


Figure 13: Frication Duration of Affricates

The negative z-score for ejectives suggests that the data point is below zero, and the frication duration of non-ejective affricates lies above the mean. Ejectives with a longer VOT and a shorter frication duration compared to the voiceless affricates are also reported in Kye (2024). Our results are consistent with Kye's (2024) description of /tʃ/ and /tʃ'/.

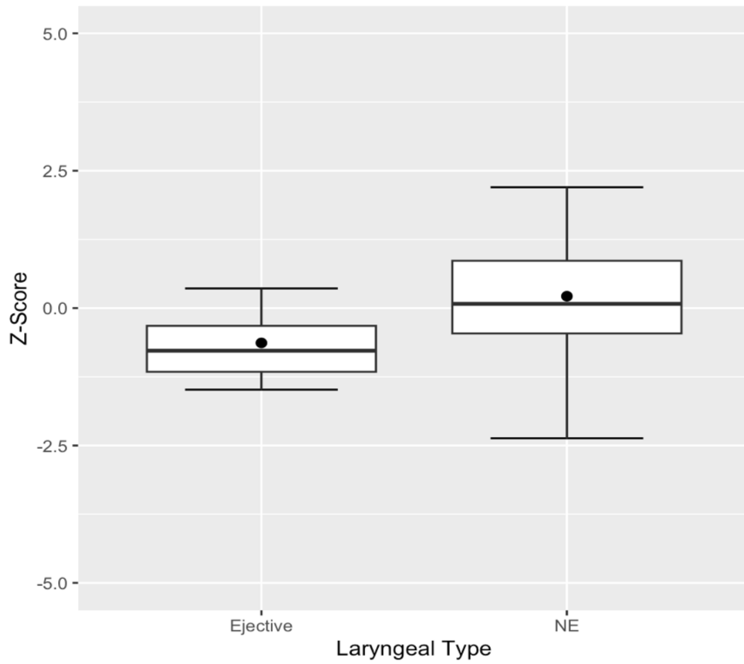


Figure 14: Z-scored Frication Duration

3.2.2 Burst Intensity

The mean values of burst intensity across speakers and types are listed in Table 13.

	Speaker 1	Speaker 2
Ejective	62.41 dB	56.0 dB
Non-ejective	53.8 dB	54.4 dB

Table 13: Mean Burst Intensity of Affricates

As reflected in Figure 15, the ejective affricates have a higher mean burst intensity than the voiceless affricates. However, note that the difference is much more trivial for Speaker 2 since the mean values across stop types are very similar.

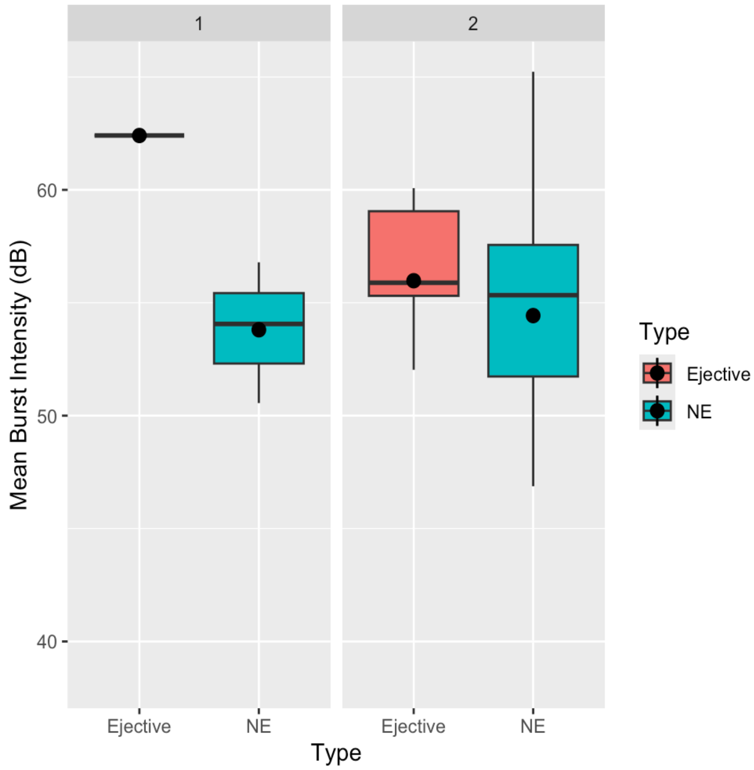


Figure 15: Mean Burst Intensity of Affricates

The normalized data show identical patterns. The ejective data is generally higher than the mean value while the non-ejective data is more likely to lie below the mean. Even though the pattern seems different, we did not find a statistically significant difference ($F(1)= 1.921$, $p=0.171$).

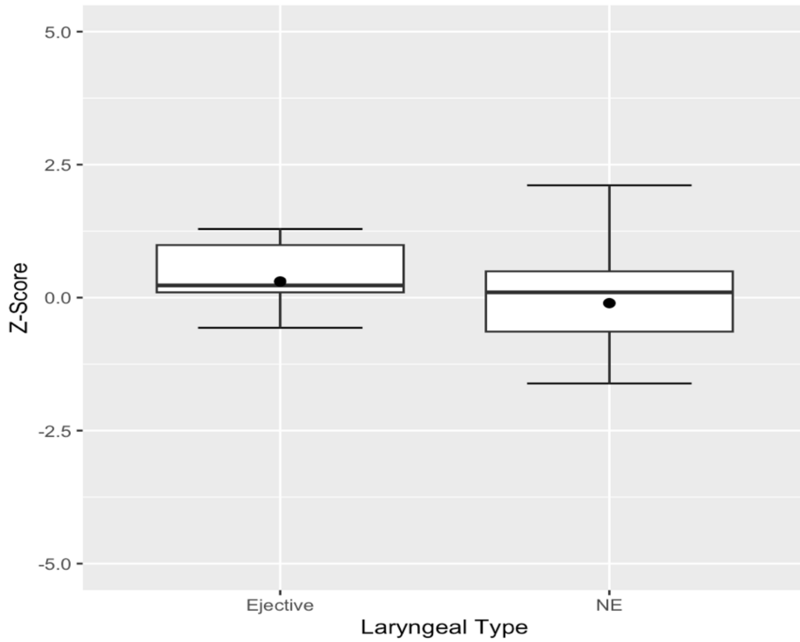


Figure 16: Z-normalized Mean Burst Intensity

3.2.3 F0 Perturbation

To examine whether F0 is raised or depressed at voicing onset when following ejective affricates, F0 perturbation was also evaluated.

	Speaker 1	Speaker 2
Ejective	-9.5 Hz	9.9 Hz
Non-ejective	9.0 Hz	17.1 Hz

Table 14: Mean F0 Perturbation of Vowels following Affricates

The mean values and Figure 17 demonstrate that F0 of the vowel is depressed at voicing onset when following ejective affricates for Speaker 1. However, the data from Speaker 2 shows a raised F0 at vowel onset, though to a smaller extent than vowels following non-ejective affricates. Vowels preceded by non-ejective affricates always have their F0 raised.

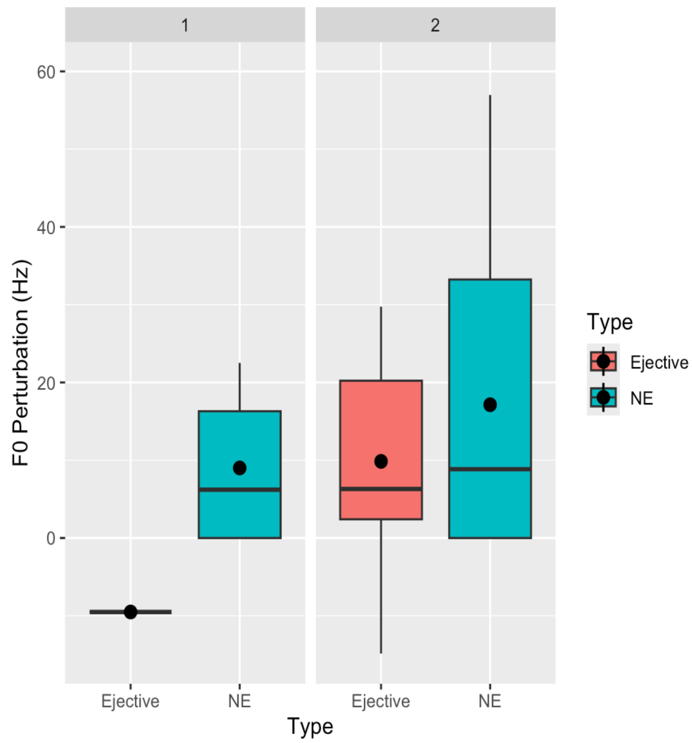


Figure 17: F0 Perturbation of Vowels following Affricates

The z-score indicates that F0 perturbation of vowels following ejective affricates is generally below the mean values, but the positive z-score of non-ejective affricates represents an overall distribution higher than the mean. The difference is not statistically significant (F(1)=1.407, p=0.242).

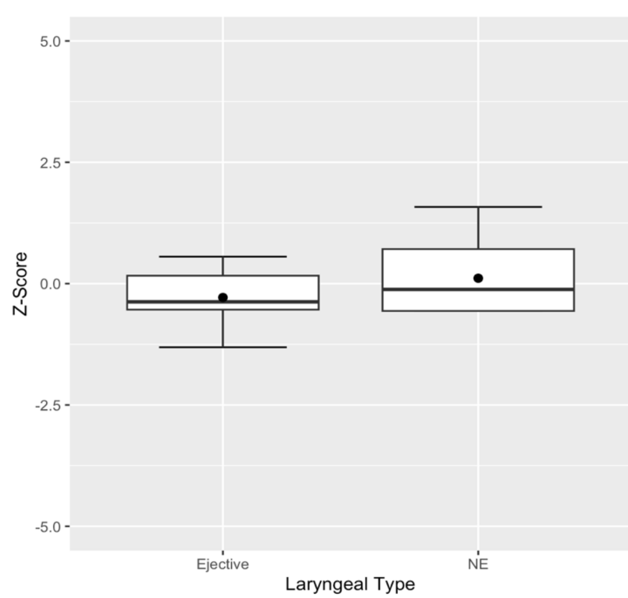


Figure 18: Z-normalized F0 Perturbation

The change in F0 over time further visualizes F0 perturbation. For Speaker 1, the F0 at 10% of the vowel is lower relative to the vowel midpoint; for the second speaker the vowel has a slightly lower F0 at 50% point than the onset. Consistently higher F0 is found at 10% point when the vowel follows non-ejective affricates, aligning with the results of Kye (2024).

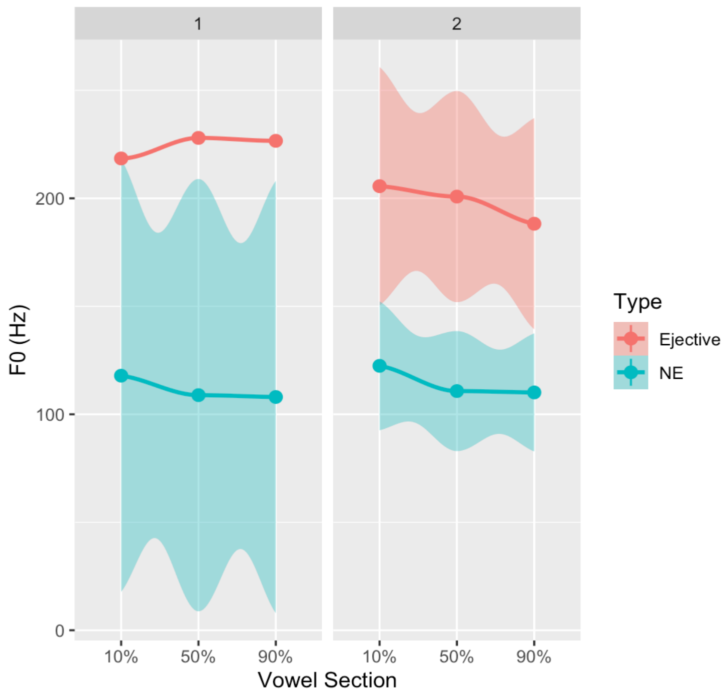


Figure 19: Mean F0 at the 10%, 50%, and 90% Points of the Vowels following Affricates

3.2.4 H1*-H2*

H1*-H2* values were also measured to determine the voice quality of the vowels preceded by different affricate series. The two speakers again demonstrated opposite patterns, where the vowels following ejective affricates are creakier for Speaker 1 but more modal for Speaker2.

	Participant 1	Participant 2
Ejective	-0.4 dB	1.0 dB
Non-ejective	2.3 dB	0.3 dB

Table 15: Mean Values of H1*-H2* of Vowels following Affricates

We see in the boxplot that for Speaker 1, ejective affricates are followed by vowels with lower H1*-H2* than non-ejectives. The overall H1*-H2* values of vowels produced by Speaker 2 are similar, and the ejective series has a slightly higher mean.

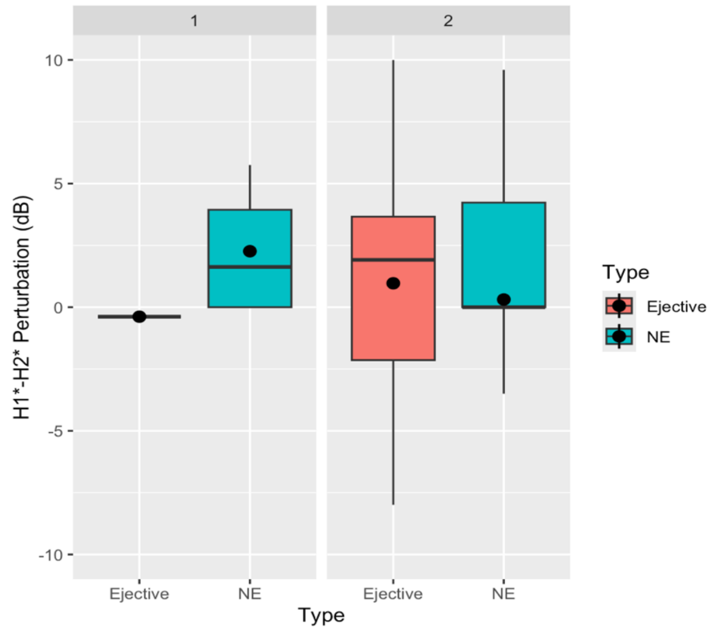


Figure 20: H1*-H2* Perturbation of Vowels following Affricates

The z-scores of the two categories are very similar, both close to zero. This observation is further confirmed by a one-way ANOVA test ($F(1)=0$, $p=0.993$), implying that the difference is not statistically significant.

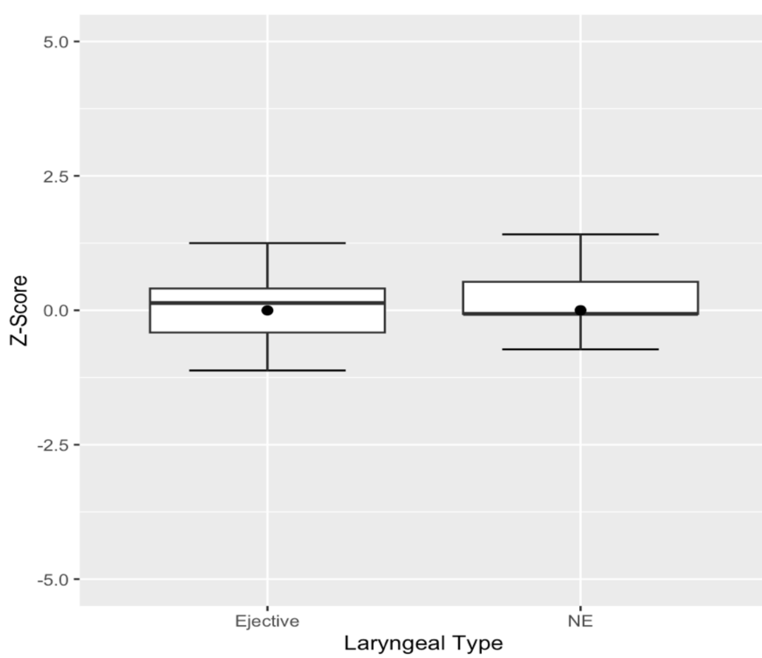


Figure 21: Z-normalized H1*-H2* Perturbation

The trajectories of H1*-H2* over time reflect the same pattern. The left panel is the visualized H1*-H2* contour of Speaker 1 compared to the vowel onset, H1*-H2* of vowels following ejective affricates is higher at vowel 50% point, and it is lower at vowel midpoint when the vowels are preceded by non-ejectives. Aside from different shapes, non-ejective categories have a higher H1*-H2* at vowel onset. For Speaker2, the contour shapes of both affricate series are identical with a higher H1*-H2* at vowel onset than vowel midpoint. Ejective affricates have an overall higher H1*-H2* than non-ejectives.

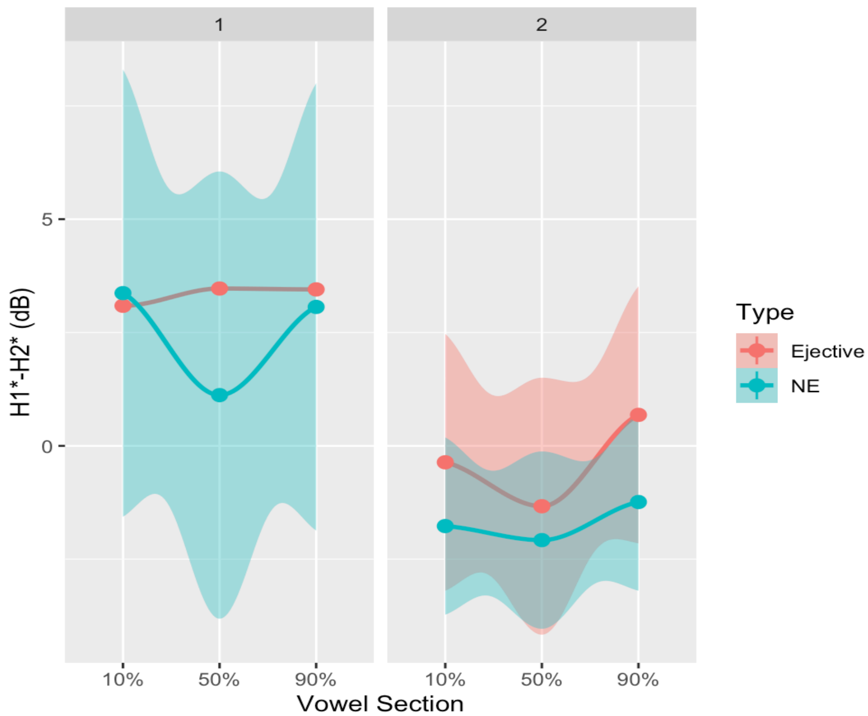


Figure 22: Mean H1*-H2* at the 10%, 50%, and 90% Points of the Vowels following Affricates

3.3 Summary

Table 16 below summarizes the combined results from all speakers, including the mean values and binary category supported by the data.

	Release Duration	Burst Intensity	F0 Perturbation	H1*-H2* Perturbation
Voiceless Aspirated Stops	64ms	52dB	7.38Hz (Raised F0)	4.83dB (Modal)
Ejective Stops	115ms	54dB	4.7Hz (Raised F0)	0.03dB (Modal)
Significance	Yes	Yes	Yes	No
Category	Strong	Strong	Strong	-

Table 16: Summary of Results from Stops

In general, we observed different patterns between ejectives and non-ejectives in most of the acoustic dimensions measured for stops, including VOT, mean burst intensity, and F0 perturbation. Ejective stops have longer release duration, higher burst intensity, and raised F0 at vowel onset. However, note that the voiceless aspirated stops also have raised rather than depressed F0 at vowel onset.

Table 17 presents the corresponding measure for affricates.

	VOT	Frication Duration	Burst Intensity	F0 Perturbation	H1*-H2* Perturbation
Voiceless Affricates	7.8 ms	69.8 ms	54.4 dB	8.2 Hz (Raised F0)	0.57 dB (Modal)
Ejective Affricates	62.9 ms	96.7 ms	56.4 dB	15.8 Hz (Raised F0)	0.88 dB (Modal)
Significance	Yes	N/A	No	No	No
Category	Strong	N/A	-	-	

Table 17: Summary of Results from Affricates

Ejective and voiceless affricates did not differ as much as stops. We only observed a significant difference in VOT. All other acoustic dimensions were similar and therefore do not help us distinguish between strong and weak ejectives in this case.

4. Discussion

Itelmen oral stops and affricates make, for each place of articulation, a two-way contrast between voiceless aspirated and ejective segments. The largest acoustic contrast between the two series (and the only measure that was statistically significant for the affricates) is VOT. In fact, compared to ejectives in other languages, Itelmen ejectives have an especially long VOT. A possible explanation is that since there is only a two-way laryngeal contrast between long-lag voiceless aspirated stops and ejectives, the VOT must be extra-long for speakers to adapt in production and perception. This aligns, perhaps, with existing perception studies such as Percival (2023), which investigates perceptual cues to ejective stops used by listeners from four languages with typologically different laryngeal contrast and concludes that all speakers used the period of silence following the stop burst as their primary cue (Percival, 2023).

In the binary typology summarized in Table 1, long VOT is associated with strong ejectives. Burst intensity and pitch on vowel onset also show patterns corresponding to the description of strong ejectives with more extreme larynx raising. The voice quality at the onset of the following vowels, however, did not differ across laryngeal types. Corrected H1*-H2* mean values across speakers are consistent with the absence of the creaky voice that would arguably characterize Itelmen ejectives as weak, but as noted above, there is variation among the speakers regarding whether vowels after ejectives or after voiceless aspirates are creakier.

If one were to posit a binary division between strong and weak ejectives, then Itelmen ejective stops would be classified as strong. But our results, especially when the affricates are considered alongside the stops, are also consistent with many previous studies which suggest that the acoustic cues to ejectives in Table 1 do not reliably cluster into two discrete groups across languages. Note also that the patterns are not identical across speakers. Similar inter-speaker variation is also found in Ingush, a Caucasian language: at least 5 variants of VOT and F0 perturbation combinations, including long VOT and raised F0, intermediate VOT and raised F0, long VOT and depressed F0, and so on, are reported (Wright et al., 2002). Additionally, the Ingush ejectives can be classified as strong when considering the F0 and voice quality dimensions of the following vowel, but the short release duration (26.2 ms for ejective stops and 45.1 ms for pulmonic voiceless stops) and slightly weaker burst than the pulmonic stops pattern with the description of weak ejectives (Warner, 1996). The results of the acoustic and articulatory aspects of ejectives in various Caucasian languages, including Avar, Ingush, Georgian, and others (Grawunder, Simpson & Khalilov, 2010), as well as Tsilhuqut'in (Ham, 2007), Witsuwit'en (Wright, Hargus & Davis, 2002) suggest that the various acoustic cues associated with ejectives fail to constitute a binary typology, and may vary at least to some extent independently of one another. Ladefoged (1973) points out that the glottalic egressive sounds are part of a scale or range. While some of our data may seem relatively consistent (i.e., the mean values never reflect depressed F0 at vowel onset), clear inter-speaker variation, especially in the production of affricates, reveals that simple classification as strong or stiff for the ejective series in the language as a whole does not do justice to the more fine-grained variation we observe.

5. Afterword: On the origin of Itelmen ejectives

A major remaining question about Itelmen ejectives is their diachronic origin. Why does Itelmen alone among the Chukotko-Kamchatkan (and neighbouring) languages have ejectives? Fortescue (2005:7-8) suggests that ejectives in Itelmen generally arise diachronically from syncope, often of a preceding consonant or syllable. Some forms do appear to conform to Fortescue's suggestion—Itelmen forms with ejectives are given here alongside relevant cognates in C (Chukchi), A (Alutor), and K (Koryak) (as identified by Fortescue in the case of (d)).

(3) a. t'sal 'fox' C: jatjol; A: tatol; K: jajol(a)

b. it'e 'when' K: tite

c. k'e 'who' K: meki

d. tʃ'e- 'enter' K: tʃəɣel– 'step/go in (to a river)'

But it remains unclear how systematic this is – one can easily find Itelmen ejectives for which cognates provide no evidence of a "missing" C:

(4) a. tʃ'atʃ'a- 'red' K: tʃetʃtʃə-

b. tʃ’it ‘bow,gun’ K: əjət, C:ərət

c. p’axp’aŋ ‘boiled’ A: pəɣəpɣ-

And even for the examples in (3) where there may have been a lost consonant, similar forms where an initial consonant may have been lost did not yield ejectives in Itelmen. Compare (3) to (5).

(5) a. te- ‘bring’ C: ret-; A: tat-; K: jet-

b. sis ‘needle’ K/C/A: titi-

c. maʔ ‘where’, mank(e) ‘where to’ K/C/A: miŋkə

For (3d), Fortescue relates the initial ejective to the loss of the second syllable (in the corresponding Koryak verb root), but note that Chukchi and Alutor have *re-* ‘go in’, a closer cognate both in meaning and in phonology: Chukchi /r-/ (in women’s speech [tʃ]) often corresponds to Itelmen /tʃ/, a better cognate candidate, but for the ejective.

Note also that Fortescue's suggestion (as he recognizes) does not encompass the allomorphic distribution of ejectives in the prefixes. The prefixes that undergo these alternations have clear cognates in the Chukotkan languages, with no “missing” consonants or syllables. The participial prefix *k(’)-* in Itelmen corresponds to Chukotkan *ɣ(e)-,* and the 1sg indicative to *t(ə)-* (in both, the C- form is pre-vocalic and the CV- form before consonants). This appears therefore to be a more recent, Itelmen-internal development.

In sum, despite tantalizing speculation, the origin of the ejective consonant series in Itelmen remains obscure.[5]

6. Conclusion

This chapter has offered a preliminary study of ejectives in Itelmen. We have summarized some distributional observations about the ejective series of voiceless stops (and affricates) and presented some acoustic measurements in service of allowing Itelmen ejectives to be situated in a cross-linguistic context. Among the measures reported, Itelmen ejective stops pattern primarily as strong ejectives, while for the affricates, only VOT serves as a statistically significant cue for distinguishing ejectives from corresponding voiceless aspirated affricates. Nevertheless, we have reported more fine detail on each of the acoustic measures, as we find that even with the small

[5] We lay aside here the suggestion in Everett 2013 that ejectives are statistically correlated with languages that are spoken within 200km of elevations above 1500m. Although the Kamchatka peninsula is indeed mountainous (including the highest active volcano in the Northern Hemisphere at 4,750m and two mountain ranges with peaks above 3,000m), this offers no real explanation. There is no evidence that the Itelmens ever lived at altitude, as opposed to the coastal regions and river valleys, and some evidence that they traditionally avoided the mountains due to the volcanic activity being associated with evil spirits (Krasheninnikov 1774), so any causal mechanism in terms of reduced physiological effort needed to produce ejectives due to lower ambient air pressure at higher altitude would not apply.

sample we have investigated, there is interesting variation among speakers that a simple classification into strong versus weak does not adequately reflect.

Acknowledgments

We are first and foremost grateful to the members of the Itelmen community and Itelmen speakers who have shared their knowledge with the second author over many years. For this paper, we thank especially the late Klavdia Nikolaevna Khaloimova, Galina Afanasievna Zaporotskaya, and Georgi Dmitrevich Zaporotskij. For useful feedback in preparing this work, we thank Katie Franich, Robert Xu, and members of the Harvard PhonLab. Our thanks also to Maida Percival for the initial invitation to contribute to this project and for assistance with the Praat script. This material is based upon work (recordings) supported in part by the Endangered Language Fund and by the National Science Foundation under Awards No. BCS 1065038 (Bobaljik) and 1065619 (PI: D. Koester) "Integrated Audio-Video Dictionary of Itelmen."

References

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of phonetics*, *63*, 75-86.

Asinovskij, Alexander S. 2004. Сопоставительная фонетика чукотско-камчатских языков. [Comparative phonetics of the Chukotko-Kamchatkan Languages] Ph.D. Dissertation. St. Petersburg: Institute of Linguistic Studies, Russian Academy of Sciences.

Bird, S. (2002). Dakelh ejectives: evidence for new ways of classifying sounds. *Paper delivered at LSA*.

Bobaljik, Jonathan David. 1998. Mostly Predictable: Cyclicity and the distribution of schwa in Itelmen. In Vida Samiian ed., *Proceedings of the Western Conference on Linguistics 9 (1996)*. Fresno: California State University. 14-28.

Boersma, Paul & Weenink, David (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.20.

Chodroff, Eleanor, Alessandra Golden, Colin Wilson. 2019. Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization. *J. Acoust. Soc. Am.* 1 January 2019; 145 (1): EL109–EL115. https://doi.org/10.1121/1.5088035

Fallon, P. D. (2013). *The synchronic and diachronic phonology of ejectives*. Routledge.

Fortescue, Michael. 2005. *Comparative Chukotko-Kamchatkan Dictionary*. Berlin: Mouton de Gruyter.

Gallagher, Gillian. 2010. *The perceptual basis of laryngeal cooccurrence restrictions.* Ph.D. Dissertation, MIT.

Garellek, M. (2022). Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality. *Journal of Phonetics*, *94*, 101155.

Glass, G. V., & Hopkins, K. D. (1995). Multiple comparisons and trend analysis. *Statistical methods in education and psychology*, *3*, 444-481.

Grawunder, S., Simpson, A. P., & Khalilov, M. (2010). Phonetic characteristics of ejectives–samples from Caucasian languages. *Turbulent sounds. An interdisciplinary guide*, *21*, 209-244.

Ham, S. (2007). Tsilhqut'in ejectives. In *2007 Annual Conference of the Canadian Linguistic Association* (pp. 1-15).

Hargus, S. (2011). *Witsuwit'en grammar: Phonetics, phonology, morphology*. UBC Press.

Keating, P., & Garellek, M. (2015). Acoustic analysis of creaky voice. *Poster apresentado em sessão especial sobre voz crepitante no Encontro Anual da Linguistic Society of America em Portland (OR)*.

Kingston, J. C. (1984). *The phonetics and phonology of the timing of oral and glottal events*. University of California, Berkeley.

Kye, T. (2024). Affricates in Lushootseed. In *Proc. Interspeech 2024* (pp. 3699-3703).

Ladefoged, P. (1973). The features of the larynx. *Journal of phonetics*, *1*(1), 73-83.

Ladefoged, P., & Maddieson, I. (1996). The sounds of the world's languages. *(No Title)*.

Lindau, M. (1984). Phonetic differences in glottalic consonants. *Journal of Phonetics*, *12*(2), 147-155.

Moran, Steven & McCloy, Daniel (eds.) 2019. PHOIBLE 2.0.

Percival, M. (2019, August). Contextual variation in the acoustics of Hul'q'umi'num'ejective stops. In *Proceedings of the 19th International Congress of Phonetic Sciences. Melbourne, Australia* (pp. 3270-3274).

Percival, M. (2023). Perceptual cues to ejective stops across languages. In *Proceedings of the 20th International Congress of Phonetic Sciences, Prague, Czech Republic* (pp. 282-286).

Volodin, Aleksandr P. 1976. *Ительменский язык*. [Itelmen] Moscow: Nauka.

Volodin, Aleksandr P., Chikako Ono, Jonathan David Bobaljik, David Koester and Michael Krauss. 2021. *Полный ительменско-русский словарь.* [Comprehensive Itelmen-Russian Dictionary] Fürstenberg/Havel and Saint Petersburg: Kulturstiftung Sibirien and Institute of Linguistic Studies, Russian Acadmy of Sciences. Contents available at: Itelmenslovar.org

Warner, N. (1996, October). Acoustic characteristics of ejectives in Ingush. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96* (Vol. 3, pp. 1525-1528). IEEE.

Wright, R., Hargus, S., & Davis, K. (2002). On the categorization of ejectives: data from Witsuwit'en. *Journal of the International Phonetic Association*, *32*(1), 43-77.