Itelmen Ejectives

Xiaotian Wang and Jonathan David Bobaljik

Harvard University

Revised August 2026

1. Introduction

This paper offers a preliminary study of ejective consonants in (Western) 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. Itelmen is the only Chukotko-Kamchatkan language–and the only language in the area– with an ejective consonant series. Our core finding is that very long voice onset time (VOT) and strong burst intensity (for stops but not affricates) are consistent properties of Itelmen ejectives, but we note that there is substantial inter-speaker variation regarding vowel quality, even in our sample of three speakers. Itelmen ejectives thus pattern mostly as “strong” under the classification in Kingston (1985) and subsequent work although with some qualifications.

1.1. Background

The Itelmen consonant inventory is given in Table 1, alongside which we recognize five full vowels (a,e,i,o,u) and one (possibly epenthetic) reduced vowel (ə). Existing scholarship disagrees about some aspects of the inventory, primarily the number of vowels, the number and phonetic quality of the fricatives and approximants (what we write as <β,ɣ> may be [ʋ, ɰ], or part of a richer spectrum), as well as the status of some consonants that occur primarily in loanwords, onomatopoeia, and some possibly suprasegmental phenomena. We have generally followed the dictionary forms in transcriptions below, unless otherwise noted.[1]

Table 1: Itelmen Consonant Inventory

			labial (labiovelar)	coronal	palatal	velar	uvular	glottal
nasal			m	n	ɲ	ŋ		
central	stops	pulmonic	p	t		k	q	ʔ
		glottalic	p’	t’		k’	q’	
	affricates	pulmonic		ʧ				
		glottalic		ʧʼ				

	fricatives	ɸ	s̠ z̠		x	χ	
	approximants	β (ʋ)		j	ɣ (ɰ)		
	rhotic		r				
lateral	fricative		ɬ				
	approximants		l	ʎ			

We focus here on the ejective series of consonants: stops: /p’,t’,k’,q’/, and affricate /tʃ’/, which occur alongside a corresponding series of voiceless oral stops /p,t,k,q/, and an affricate /tʃ/. We perceive the 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 VOT < 35ms), a fact that is relevant due to concerns about Russian influence on the pronunciation of Itelmen by bilingual speakers (see section 3.1.1 below). That is, the primary contrast in Itelmen stops and affricates appears to be voiceless aspirated (long-lag VOT) versus voiceless ejective (very long VOT). Voiced stops do not occur in the (etymologically) native vocabulary (apart from apparently onomatopoeic etyma), but centuries of contact with Russian have introduced numerous loanwords with voiced stops—any sizeable corpus of Itelmen will contain voiced stops.

Impressionistically, 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 in the second author’s experience there is some amount of intra-speaker variation.

Fallon (2002) and Percival (2019) report that ejectives in many languages have a restricted distribution. In Itelmen ejectives may occur in any position in a word (initial, medial or final) and may occur internal to consonant clusters as well as in pre-vocalic position (Itelmen tolerates complex clusters, though disallows sonorants in cluster-medial position; see Volodin, 1976; Bobaljik, 1998). 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 cluster	p’k’akas	t’sal	k’nəntx	q’lekas	tʃ’mamɬkas
	‘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’

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 (Volodin, 1976:88-89). Examples in (2a) illustrate this for the first person singular subject prefix *t-/t’*- Note that plain *t*- occurs freely (unsurprisingly) in prevocalic and prenasal position (2b).

(2) Ejective allomorphs of prefixes (examples from Volodin et al. 2021)

a.	t’-i-s-kitʃen	t’-nu-s-kitʃen	t-ɨ-kitʃen
	1SGS-go-PRES-1SGS	1SGS-eat-PRES-1SGS	1SGS-be-1SGS
	‘I am going’ /__V	‘I am eating’ /__N	‘I was’ /__C
b.	tinu-ʔn	tnum-k	
	‘relative-PL’	‘down-LOC’	

While there is some variation among sources (and possibly among speakers, in particular as to whether ejective allomorphs occur before schwa) the examples we examined in preparing this paper are clearly ejectives in pre-nasal and pre-vocalic positions.

It bears mention that at morpheme juncture, a plain stop can be followed by a glottal-stop-initial suffix, which yields a potential contrast between ...C’V and ...CʔV. Figure 1 shows the sound wave and spectrogram of the word *k’-ajtat-ʔan* ‘(s)he chased him/her/it out’, with the ejective allomorph of the participle prefix *k(’)-* and the glottal-initial transitive participial suffix -*ʔan.* Two stops, an ejective velar /k’/ and a plain alveolar /t/ followed by a glottal stop, are segmented. The ejective stop has a clear oral burst and glottalization before the periodic voicing of the vowel starts. The /tʔ/, on the other hand, does not have clear glottalization. Note that there does seem to be a more trivial burst-like component occurring after the burst of the alveolar /t/. However, since this burst is weak and does not have a clear closure preceding it, it is unclear whether it can be classified as a glottal stop based on the acoustic signal. We leave investigation of the potential contrast between ...C’V and ...CʔV for a future study.

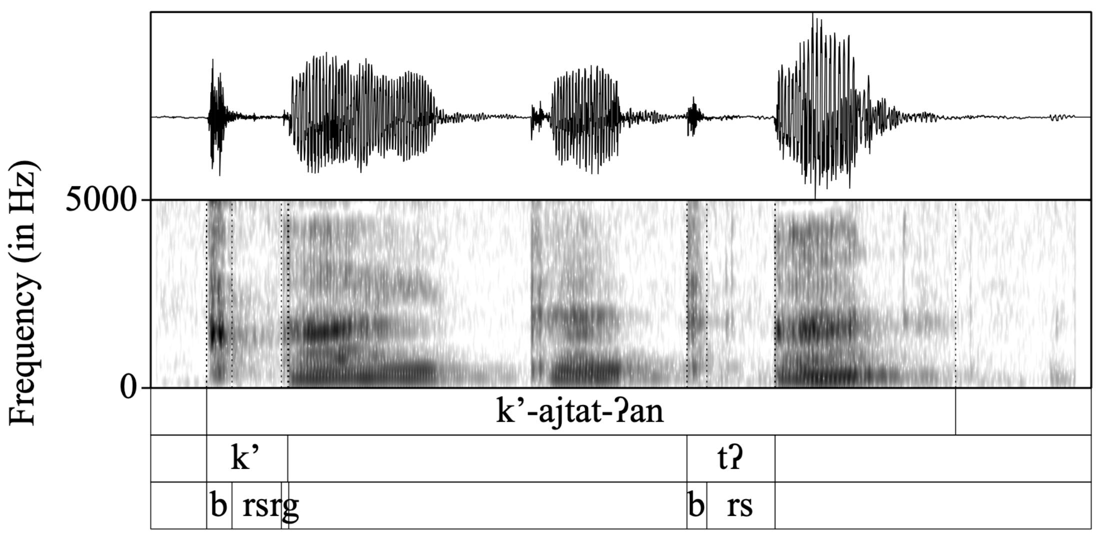


Figure 1: Sample Spectrogram of *k’-ajtat-ʔan* ‘(s)he chased him/her/it out’. The labels are as follows: b = burst, rg = glottalization, and rs = silence.

Table 2 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* (*PIRS,* Volodin et al. 2021).[2]

Table 2: Distribution of selected segments

	Initial	Medial	Final	Total		Initial	Medial	Final	Total	Ratio C:C’
p	254	408	32	694	p’	14	11	0	25	28:1
t	468	1559	109	2136	t’	47	123	1	171	12:1
k	630	1155	135	1920	k’	118	277	16	411	5:1
q	230	381	136	747	q’	10	27	6	43	4:1
tʃ	300	724	169	1193	tʃ’	125	168	5	298	4:1
SUM	1882	4227	581	6690		314	606	28	948	7:1

One notes that ejectives are infrequent compared to the corresponding plain obstruents although the ratio varies across place of articulation. Also, ejectives in this sample are relatively rare in word-final position (3% compared to 9% of the corresponding plain obstruents). For the record, we note that Asinovskij (2003:223-4) reported a different distribution over a (smaller but overlapping) corpus. His sample had fewer ejectives (n=227) than plain stops overall (n=2114), but ejectives in his sample show a stronger preference for initial position than in our sample.

2. Methodology

2.1 Speakers and materials

For the analysis reported here, we made use of existing recordings from three speakers: the language activist Klavdia Nikolaevna Khaloimova (1934-2018, Sopochnoe), Galina Afanasievna Zaporotskaya (1935-2019, Moroshechnoe), and Georgi Dmitrevich Zaporotskij (1937-2014, Morposchechnoe). The villages of Sopochnoe and Moroshechnoe are part of the southern (sub-)dialect group of Western Itelmen. Recordings were made over the period from 2011-2013, by Tatiana Degai and Chikako Ono under the auspices of the Itelmen Audio-Visual Dictionary Project (PIs: J.D. Bobaljik, D. Koester), see Volodin et al. (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 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 from school age onwards. For some time during the Soviet period, children were not permitted to speak Itelmen in schools. In some villages, including Moroshechnoe, Itelmen was still widely spoken until the 1950s. 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. Speakers 2 and 3 reported using Itelmen in the home, and a colleague of K.N. Khaloimova (speaker 1) reported to the second author (in 2025) that Khaloimova spoke only Itelmen until she started school. Although all three subsequently shifted to Russian-dominant profiles in adulthood, Khaloimova was the leading figure of the language revitalization (and documentation) efforts starting in the 1980s.[3] Our recordings represent the Itelmen language of that generation, possibly with effects of bilingualism and language attrition. 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. 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.

A further qualification is that the recordings we made use of were intended for an online dictionary. The speakers read words and phrases aloud 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

A total of 290 stops and 59 affricates were selected for data analysis (see Table 3). We tried to balance the number of stops across places of articulation. However, as shown above in Table 2, the distribution of ejective and non-ejective stops is in general unbalanced and this is to some extent reflected in our target word/segment selection.

Table 3: Consonants Extracted for Analysis by place of articulation and laryngeal type

	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

Similarly, while we tried to balance the number of stops across positions, we ended up with an unbalanced position distribution. The data includes 175 initial stops, 88 medial stops, and 27 final stops. Only medial stops followed by vowels are included in the data analysis. We prioritized intervocalic stops, but a few stops preceded by consonants (mostly sonorants) were included to balance the distribution of data as much as possible. Consequently, with these limitations, we will not discuss the difference between stops across places of articulation and word positions in detail in this study.

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, which have been held (see Percival 2024) to be salient acoustic cues distinguishing ejectives from non-ejectives and for classifying ejectives into subtypes (strong, weak etc.). These measurements include VOT, burst intensity, F0 perturbation of the following vowel, and corrected H1-H2, a measure of spectral tilt taken to indicate differences in phonation (breathy, modal, creaky voice) which have been held to be an acoustic cue to ejective contrasts in the preceding vowel (Keating et al., 2015).. 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).

Following previous studies (Percival, 2019; Ham, 2008), release duration or VOT is measured from the release of the burst to the onset of 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. Burst intensity was measured as the mean burst intensity of the target stop from the onset to the offset. F0 and voice quality were measured with perturbation following Wright et al (2002) and Percival (2019). The formulas used are shown in Table 4.

Table 4: Perturbation Formulas

F0 Perturbation	mean F0 at vowel 10% point – mean F0 at vowel midpoint
H1*-H2* Perturbation	mean H1*-H2* at voicing onset – mean H1*-H2* at vowel midpoint

With the limitations of the data discussed above, we present graphs derived from both raw data 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 (2008), is intended to help minimize the inter-speaker variation for a phonetic study with a relatively small number of tokens like ours.

While ejectives are limited to the etymologically native Itelmen vocabulary, plain (non-ejective) stops are found in the dictionary both in the native stratum and in Russian loans with varying degrees of integration. 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, see 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

The release duration of ejective and non-ejective stops is shown in Figure 2 below. Outliers are removed from the plot. The line represents the median, and the dots correspond to the mean (reported for each speaker in Table 5). Each panel shows the data of one individual speaker. Speaker 3 produced stops with the shortest release duration while the first speaker produced the longest. The voiceless non-ejective stops from the first speaker have an especially long release duration. Across all speakers, the ejective stops consistently have a significantly longer VOT than the non-ejectives.

Table 5: Mean VOT of Stops

	Speaker 1	Speaker 2	Speaker 3
Ejective	117.1 ms	114.1 ms	92.0 ms
Non-ejective	87.6 ms	56.4 ms	55.2 ms

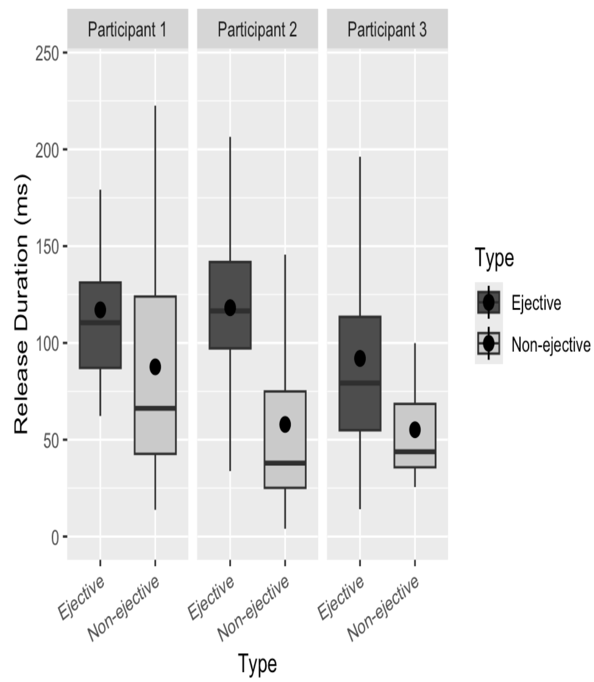


Figure 2a: Release Duration of Stops

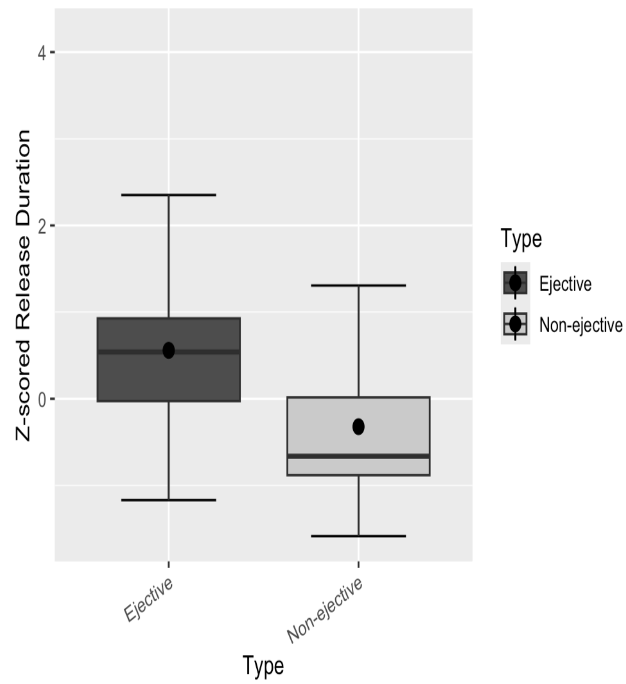


Figure 2b: Z-normalized Release Duration

Figure 2b above visualizes the z-normalized VOT across the two stop series. The mean z-score for ejective stops is positive while that for non-ejective stops is negative. A one-way ANOVA test indicates a significant difference in release duration according to laryngeal type ($F(1)=63.95$, $p<0.001$).

As we noted in section 1.1, the VOT for the non-ejective voiceless stops counts as long-lag (average of the average VOT across labial, alveolar and velar >35ms) which Chodroff et al. (2019) classify as aspirated. On this criterion (which is open to debate), Itelmen would have a two-way phonemic contrast in (voiceless) stops between aspirated and ejective. One might wonder (as a reviewer does) to what extent the long-lag VOT for the plain stops is a task effect. Since the dictionary includes many integrated Russian-source loanwords, we were able (for Speakers 1 and 2) to compare VOT directly, in the same task, for the native Itelmen vocabulary as against Russian-source loanwords. We did not have representative stops in loanwords from all places of articulation for all speakers, but where we did have Itelmen–vs.–loanword pairs, we report the mean VOT for non-ejective stops in Table 6: for bilabial and alveolar stops, and for the mean across positions (excluding uvulars, which Russian lacks), the VOT values for stops in Russian loanwords are similar to those reported elsewhere for Russian (Chodroff et al., 2019, supplemental data), namely short-lag (unaspirated), whereas those for etymologically Itelmen words are substantially longer. This suggests to us that there is a genuine VOT difference between Russian and Itelmen that is not entirely due to task effects, and moreover, that despite decades of Russian language dominance in the daily lives of these speakers, they retain subtle phonetic properties of Itelmen.

Table 6: VOT in loanwords vs Itelmen non-ejective stops

	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

3.1.2 Burst Intensity

The mean burst intensity values of ejective stops and non-ejective stops produced by all three speakers are demonstrated below (see Table 7 and Figure 3a). The stops collected from the third speaker are louder than those of the first two speakers. However, the pattern is consistent – the mean burst intensity of the ejectives is consistently slightly higher than that of non-ejectives.

Table 7: Mean Burst Intensity of Stops

	Speaker 1	Speaker 2	Speaker 3
Ejective	54.8 dB	55.9 dB	73.0 dB
Non-ejective	52.2 dB	54.9 dB	69.9 dB

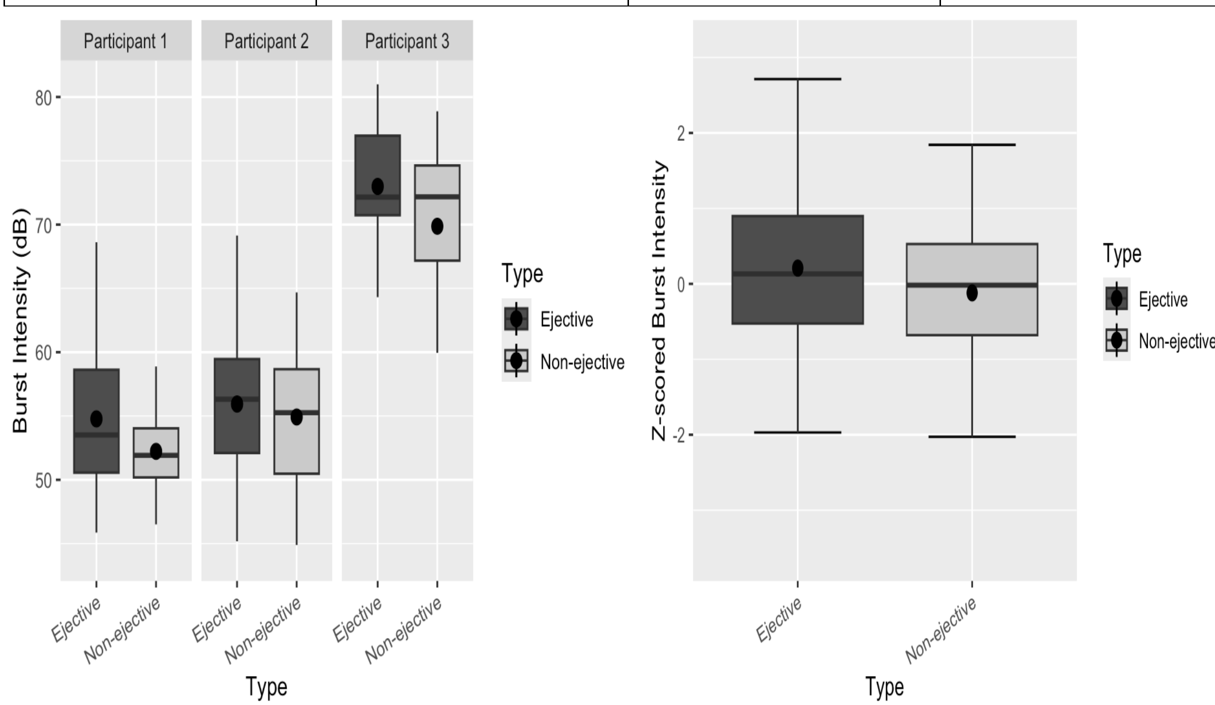


Figure 3a: Mean Burst Intensity of Stops **Figure 3b: Z-normalized Burst Intensity**

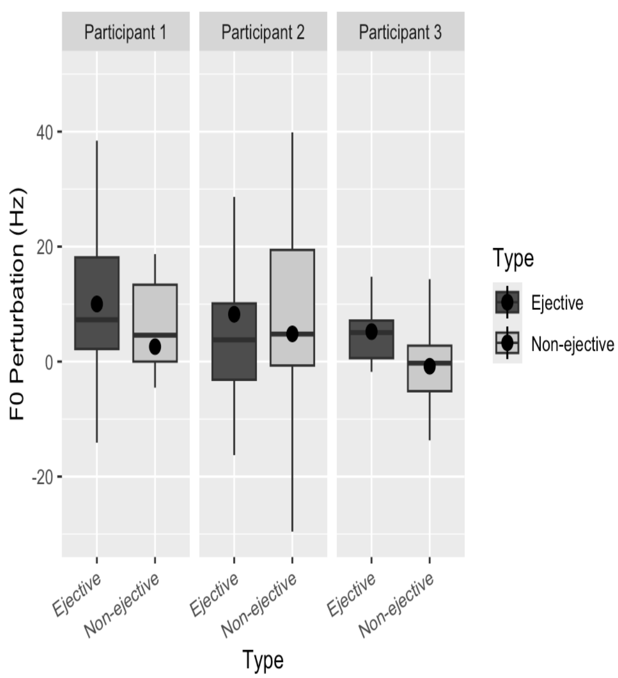
The same pattern can be seen in the z-normalized mean burst intensity (Figure 3b). The range of the box is similar across stops, with the ejectives having higher mean burst intensity. Nevertheless, we found a statistically significant difference between the mean burst intensity values of ejective and non-ejective stop through a one-way ANOVA test ($F(1)=7.475$, $p<0.01$).

3.1.3 F0

Figure 4a shows the data from all three speakers and the mean F0 perturbation (see also Table 8). For two of the three speakers, both ejectives and non-ejectives 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 non-ejectives.

Table 8: Mean F0 Perturbation of Vowels following Stops

	Speaker 1	Speaker 2	Speaker 3
Ejective	10.0 Hz	8.2 Hz	5.2 Hz
Non-ejective	2.6 Hz	4.8 Hz	-0.8 Hz



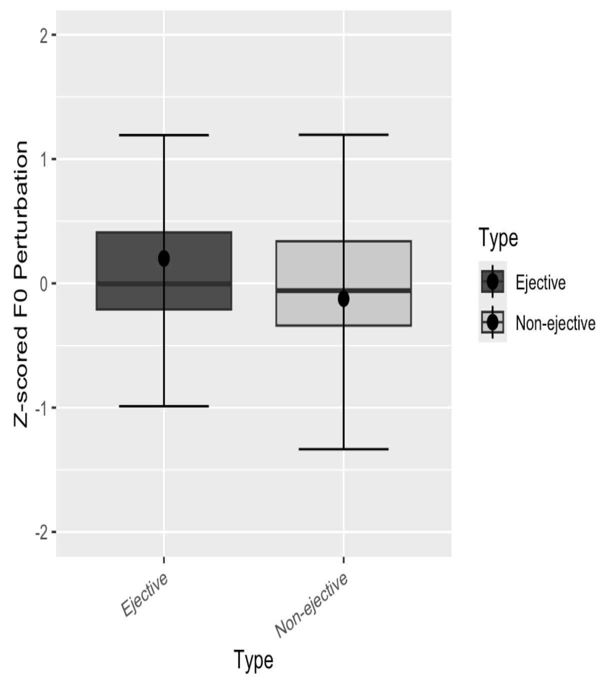


Figure 4a: F0 Perturbation **Figure 4b: Z-normalized F0**

The z-normalized F0 perturbation is displayed in Figure 4b. Although the z-score values across the two stops are close to each other(around 0), the z-normalized F0 perturbation by laryngeal type is statistically significant ($F(1)=5.663$, $p<0.05$).

The picture is less clear when we look at each speaker individually. Figure 5 below plots mean F0 at the 10%, 50%, and 90% points of the vowel following ejective and non-ejective 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. For Speaker 1 the vowels following ejective stops have a higher F0 than non-ejective stops. 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 have a slightly higher F0 when preceded by non-ejectives. The lowering trend is also only seen in the ejective stops. However, it is unclear whether these differences can be characterized directly as inter-speaker variation, since the tokens produced by the participants were not controlled. In addition to speaker variation, other factors, such as vowel quality and the prosodic condition of the vowels (i.e., stressed or not), might also affect the results.

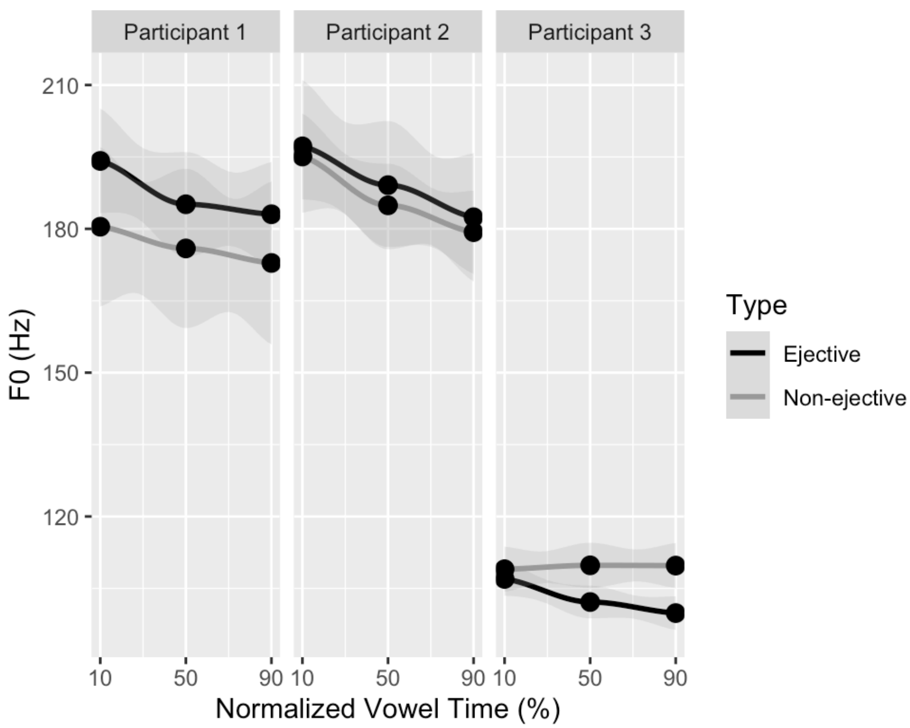


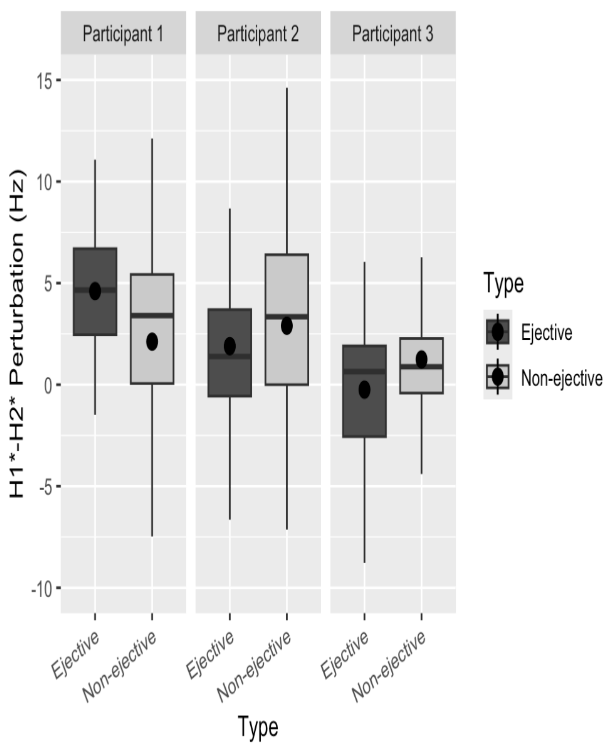
Figure 5: Mean F0 at the 10%, 50%, and 90% Points of the Vowels following Stops

3.1.4 H1-H2**

The H1*-H2* perturbation was measured by subtracting the initial corrected H1-H2 from the midpoint mean and is reported in Figure 6a. Mean H1*-H2* values are summarized in Table 9. Aside from Speaker 1, the other two speakers both had a higher H1*-H2* perturbation for non-ejectives 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.

Table 9: Mean H1*-H2* of Vowels following Stops

	Speaker 1	Speaker 2	Speaker 3
Ejective	4.6 dB	1.9 dB	-0.3 dB
Non-ejective	2.1 dB	2.9 dB	1.2 dB



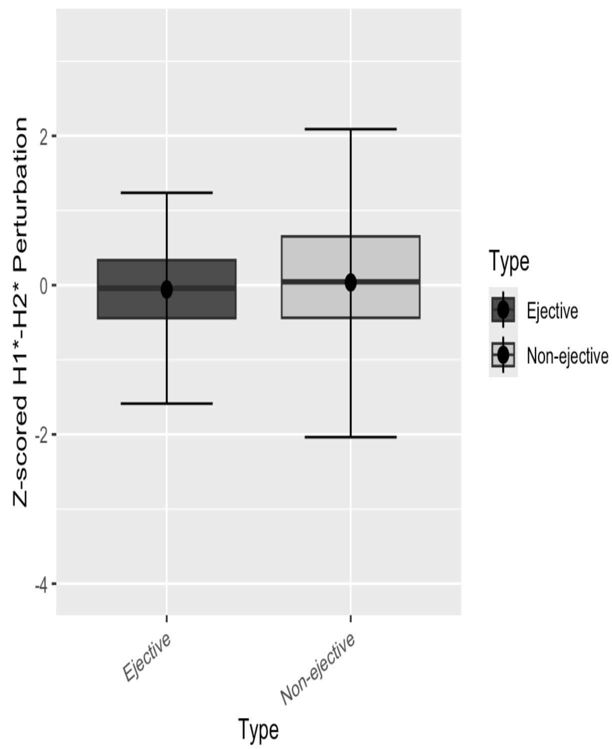


Figure 6a: H1*-H2* Perturbation **Figure 6b: Z-normalized H1*-H2* Perturbation**

The z-score values across all three speakers (Figure 6b) are close to the average value. An ANOVA test suggests that the difference is not statistically significant ($F(1)=0.472$, $p=0.493$). Figure 7 presents the change of H1*-H2* across different sections of the vowels. At vowel onset, the H1*-H2* of vowels following non-ejectives is higher than ejectives for Speaker 2 and 3; however, a different pattern is found in Speaker 1, where the onset of the vowel preceding ejectives is much more modal.

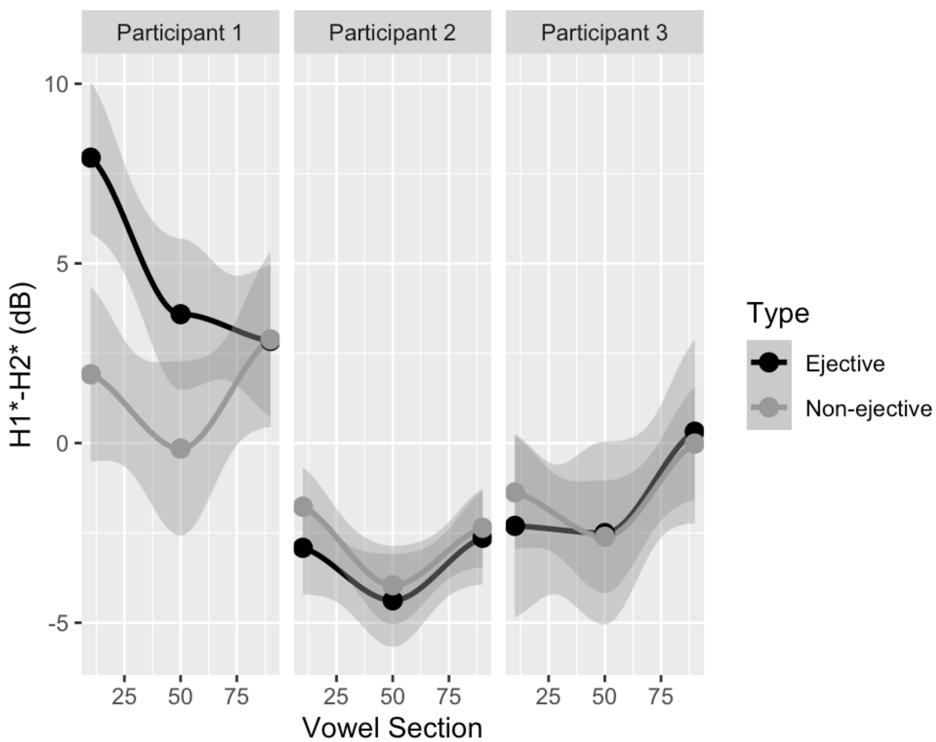


Figure 7: Mean Corrected H1-H2 at the 10%, 50%, and 90% Points of the Vowels following Stops

3.2 Affricates

We also examined the acoustics of both ejective and non-ejective affricates. As with stops, release duration, intensity, and the following vowel quality were considered. As the number of words with affricates produced by Speaker 1 and 2 is disproportionate, with Speaker 1 producing only less than 10 % of the tokens, we continue to present both the raw data and the z-normalized results.

3.2.1 Release Duration

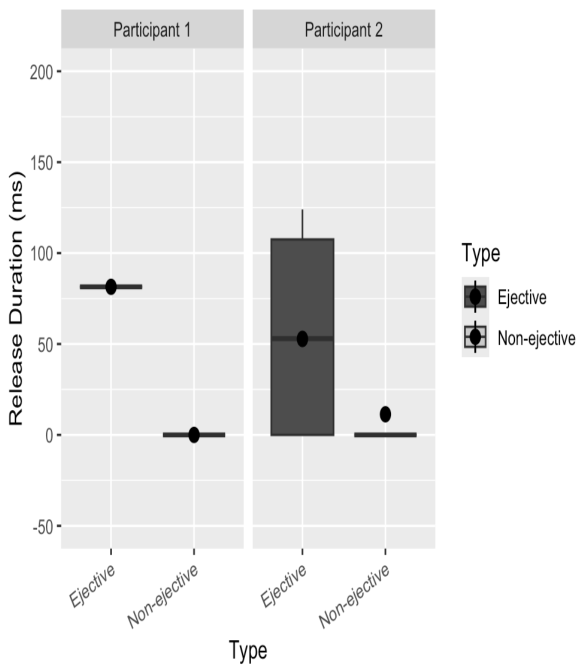
Following Abramson & Whalen (2017), the VOT of affricates was measured as the duration from the end of the frication to the onset of the voiced part. Additionally, the frication duration was also included in the analysis to enhance the compatibility between our results and those of Kye (2024). The mean VOT values are shown in Table 10. The ejective affricates have a longer VOT than the non-ejectives (represented as NE in the figures) for both speakers. Note that the non-ejective voiceless affricates have a much lower VOT than those reported in Kye (2024).

Table 10: Mean VOT of Affricates

	Speaker 1	Speaker 2

Ejective	81.4 ms	52.8 ms
Non-ejective	0 ms	11.1 ms

Figure 8a below 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.



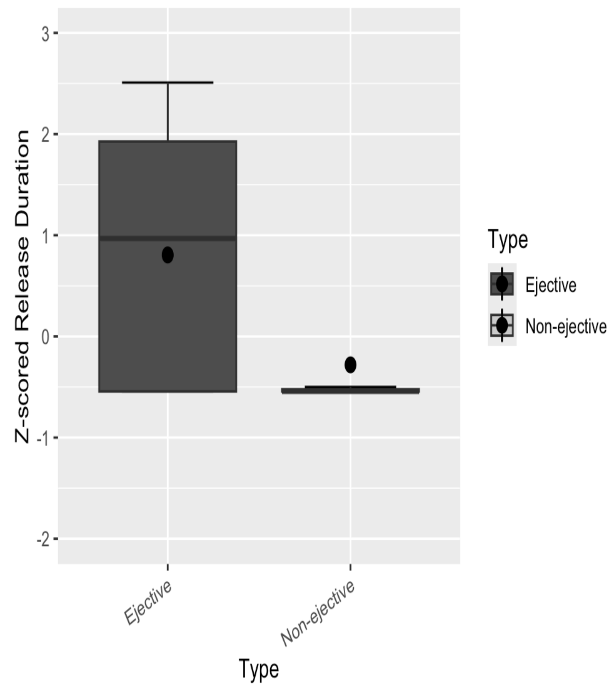


Figure 8a: VOT of Affricates **Figure 8b: Z-normalized VOT**

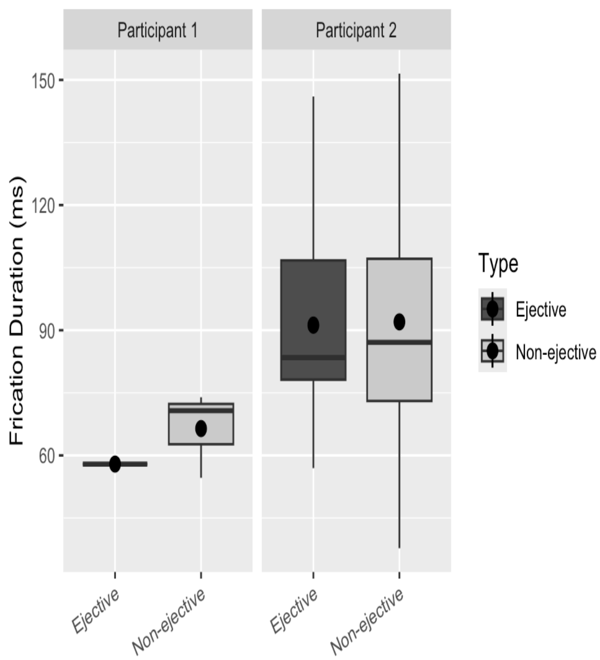
The z-scored VOT (Figure 8b) indicates that while the VOT of ejective affricates is higher than the mean value, that of non-ejective affricates tends to be lower. The difference between ejective and non-ejective affricates is statistically significant ($F(1)=17.14$, $p<0.001$).

The frication duration values of affricates are displayed Figure 9a and the mean values are reported in Table 11.

Table 11: Mean Frication Duration of Affricates

	Speaker 1	Speaker 2
Ejective	57.9 ms	91.2 ms
Non-ejective	66.4 ms	89.8 ms

The opposite pattern is observed regarding frication release duration in Speaker 1's production, where the non-ejective (voiceless) affricates have a higher frication duration than ejectives. The frication duration in Speaker 2's production is similar across affricate types.



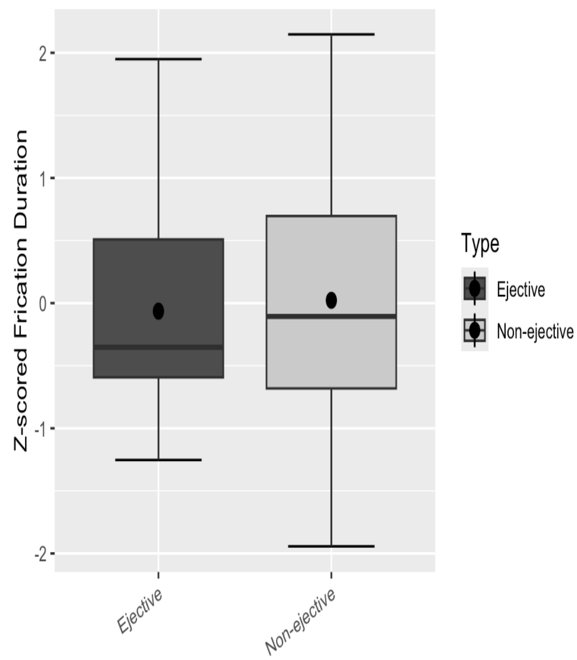


Figure 9a: Frication Duration of Affricates **Figure 9b: Z-normalized Frication Duration**

The z-scores (Figure 9b) for ejective and non-ejective affricates suggest that relative to each participant's overall distribution of frication duration, the difference between the two stop types is trivial. Ejectives with a longer VOT and a shorter frication duration compared to the voiceless affricates are also reported in Kye (2024). Even though we did not observe a consistent pattern, the direction of the change, where the difference in the mean frication duration between the two types becomes smaller than that in VOT, is consistent with Kye's (2024) description of /tʃ/ and /tʃʼ/.

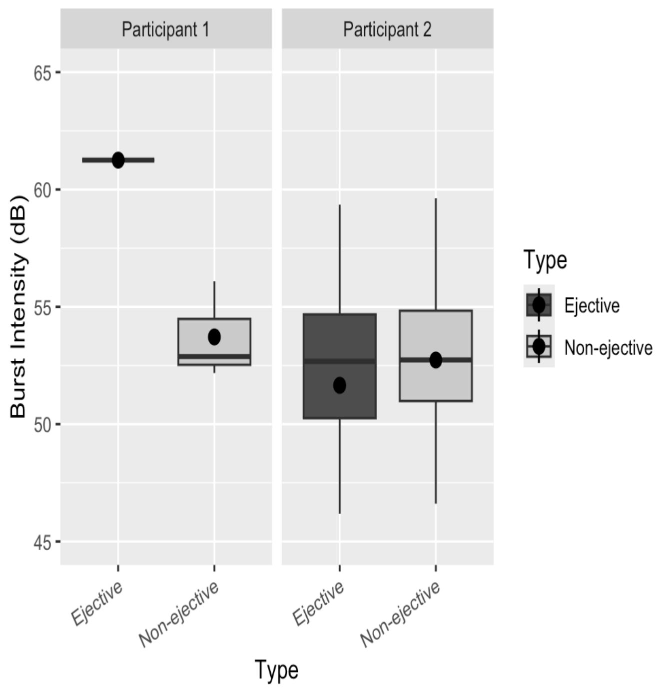
3.2.2 Burst Intensity

Since not all affricates have a visible burst, the burst intensity of affricates here is measured as the intensity of the affricates' release frication, following Kye (2024). The mean frication intensity is calculated by averaging over the intensity at the 10%, 30%, 50%, 70%, and 90% points of the frication. The mean values of burst intensity across speakers and affricate types are listed in Table 12.

Table 12: Mean Burst Intensity of Affricates

	Speaker 1	Speaker 2
Ejective	61.25 dB	51.7 dB
Non-ejective	53.7 dB	52.7 dB

As reflected in Figure 10a, the ejective affricates have a higher mean burst intensity than the voiceless affricates for Speaker 2. However, for Speaker 2, the non-ejective affricates have a slightly higher mean frication intensity than the ejective affricate. The difference between the two series is again more trivial for Speaker 2 than for Speaker 1.



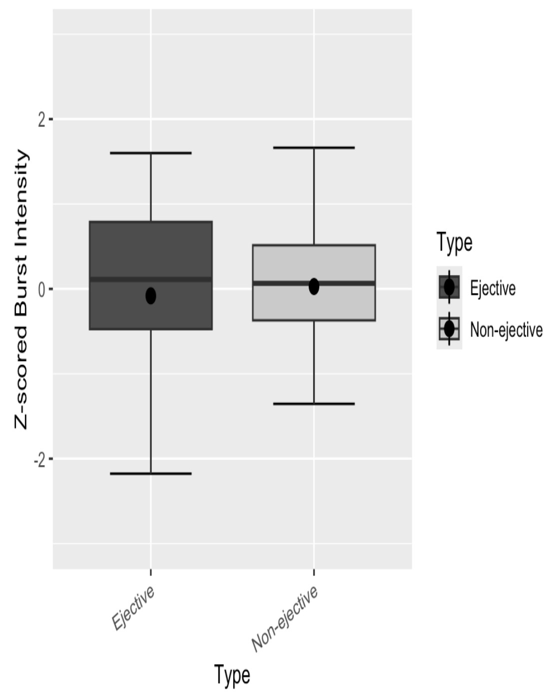


Figure 10a: Mean Burst Intensity of Affricates **Figure 10b: Z-normalized Burst Intensity**

The normalized data (Figure 10b) show identical patterns. We did not find a significant difference of frication intensity across the two series ($F(1)= 0.149$, $p=0.701$).

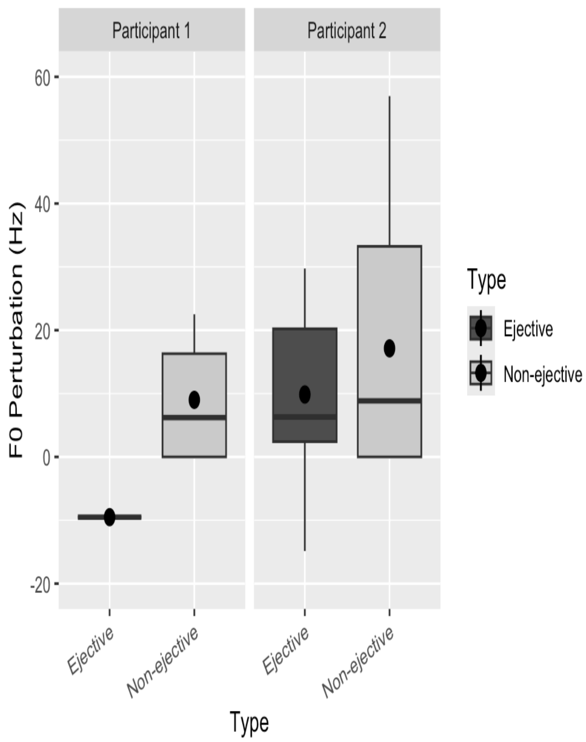
3.2.3 F0 Perturbation

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

Table 13: Mean F0 Perturbation of Vowels following Affricates

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

The mean values, given in Table 13 along with Figure 11a 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.



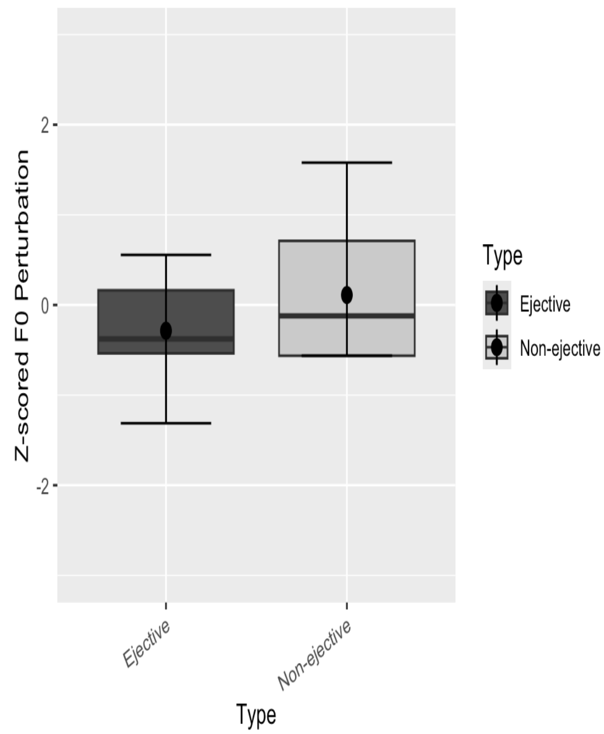


Figure 11a: F0 Perturbation **Figure 11b: Z-normalized F0 Perturbation**

The z-score indicates that F0 perturbation of vowels following ejective affricates is generally below the mean values (Figure 11b), but the positive z-score of non-ejective affricates represents an overall distribution higher than the mean. The difference is not significant ($F(1)=1.407$, $p=0.242$).

The change in F0 over time further visualizes F0 perturbation as in Figure 12. 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 the 10% point when the vowel follows non-ejective affricates, aligning with the results of Kye (2024).

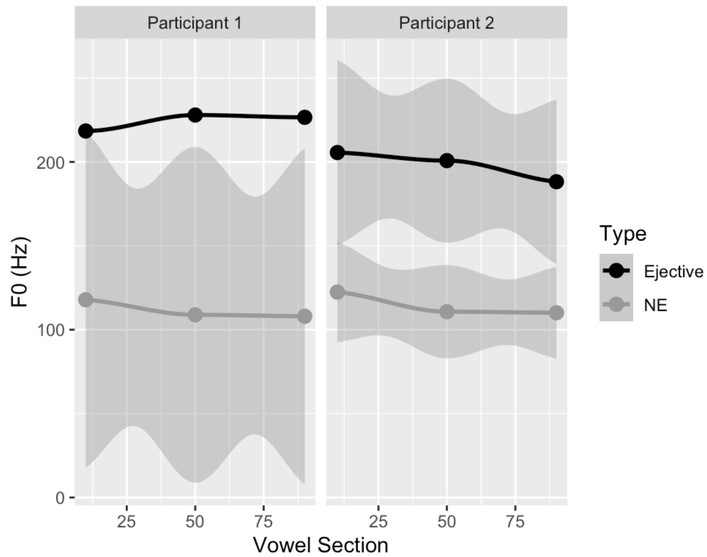


Figure 12 Mean F0 at the 10%, 50%, and 90% Points of the Vowels following Affricates

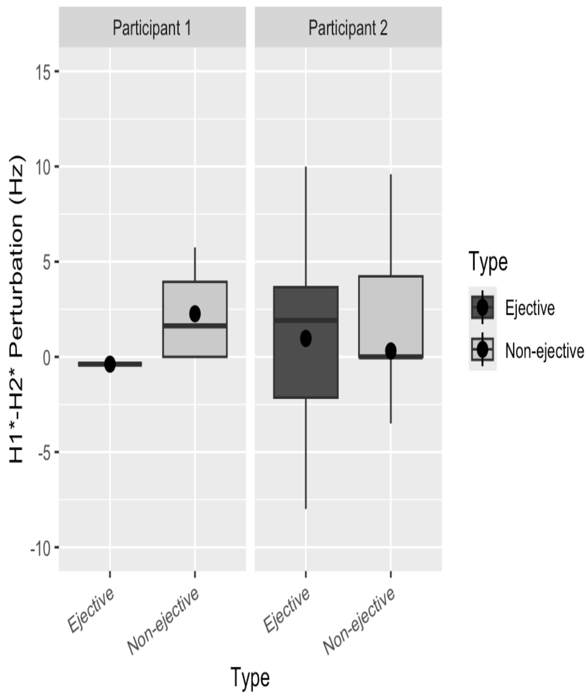
3.2.4 H1-H2**

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 Speaker 2 (see Table 14).

Table 14: Mean Values of H1*-H2* of Vowels following Affricates

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

We see in the boxplot (Figure 13a) 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.



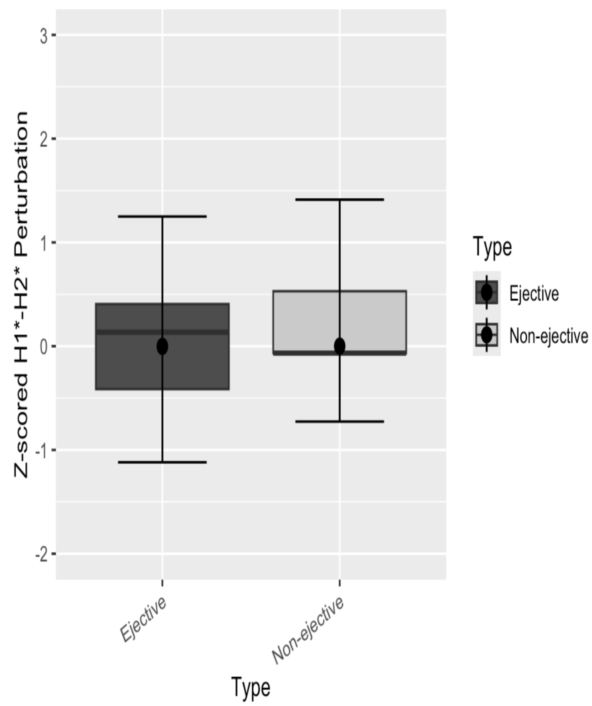


Figure 13a: H1*-H2* Perturbation **Figure 13b: Z-normalized H1*-H2* Perturbation**

The z-scores of the two categories are very similar, both close to zero (see Figure 13b). This observation is further confirmed by a one-way ANOVA test ($F(1)=0$, $p=0.993$), indicating that the difference is not statistically significant.

The trajectories of H1*-H2* over time reflect the same pattern (see Figure 14). Aside from the difference of H1*-H2* at vowel midpoint between ejectives and non-ejectives for Speaker 1, the overall contour shapes of both affricate series are identical for both speakers.

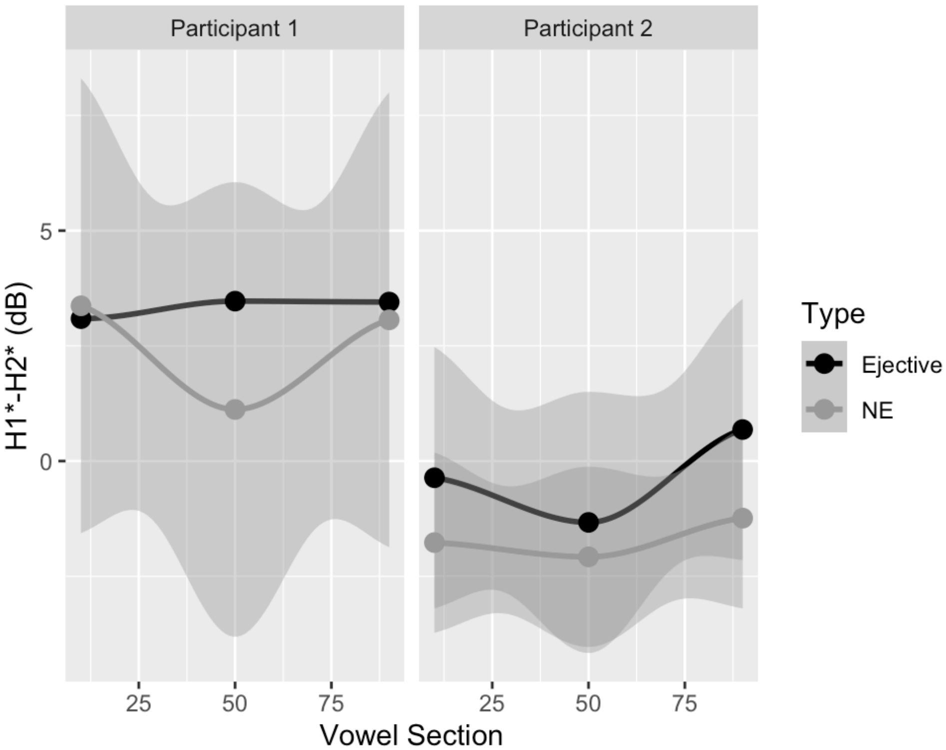


Figure 14: Mean H1*-H2* at the 10%, 50%, and 90% Points of the Vowels following Affricates

4. Discussion

Itelmen oral stops and affricates make, for each place of articulation, a two-way contrast between non-ejective and ejective segments (in the native, non-loanword vocabulary). 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 (possibly aspirated) non-ejectives 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 (2024), 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.

According to the binary typology (Kingston, 1985), 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. Table 15 below summarizes the combined results from all speakers, including the mean values and binary category supported by the data.

Table 15: Summary of Results from Stops

	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	-

If one were to accept 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 (see Table 16 for details), are also consistent with many previous studies which suggest that the acoustic cues to ejectives do not reliably cluster into two discrete groups across languages. Note also that the patterns are not identical across speakers.

Table 16: Summary of Results from Affricates

	VOT	Frication Duration	Burst Intensity	F0 Perturbation	H1*-H2* Perturbation
Voiceless Affricates	10.6 ms	90.2 ms	52.8 dB	8.2 Hz (Raised F0)	0.57 dB (Modal)
Ejective Affricates	54.7 ms	89.0 ms	52.3 dB	15.8 Hz (Raised F0)	0.88 dB (Modal)
Significance	Yes	N/A	No	No	No
Category	Strong	N/A	-	-	

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. Conclusion and Future Directions

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 affricates. Nevertheless, we have reported more fine detail on each of the acoustic measures, as we find that even with the small sample we have investigated, there is interesting variation among speakers that a simple classification into strong versus weak does not adequately reflect.

Our study, as we have noted, has been preliminary in some ways, limited in part by the material available. Various aspects of Itelmen ejectives might profitably be considered in more detail in follow-up studies. For example, we mentioned in Section 1.1 that at morpheme juncture, a plain stop may be followed by a glottal-stop-initial suffix. In theory this yields a contrast between C’V and CʔV, investigation of which we must leave to future work. Another point that remains unresolved is the diachronic origin of Itelmen ejectives. 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 the loss of a preceding consonant or syllable (compare Itelmen *t’sal* ‘fox’ to Chukchi *jatjol*, Itelmen *k’e* ‘who’ to Koryak *meki*). But it remains unclear how systematic this is since many ejectives (and especially the ejective allomorphs of the prefixes) do not have “missing” consonants compared to their cognates, and syncope does not always yield ejectives in Itelmen. The origins of ejectives thus remain unfrotunately obscure.

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, and to Maida Percial, Bonny Sands, Inna Sieber and Mark Zimin for helpful feedback on the first draft. 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, A. S. (2003). Сопоставительная фонетика чукотско-камчатских языков [Comparative phonetics of the Chukotko-Kamchatkan Languages]. [Unpublished doctoral dissertation]. Institute of Linguistic Studies, Russian Academy of Sciences.

Bobaljik, J. D. (1998). Mostly Predictable: Cyclicity and the distribution of schwa in Itelmen. In V. Samiian (Ed.), *Proceedings of the Western Conference on Linguistics 9 (1996)* (pp. 14-28). California State University Fresno.

Bobaljik, J. D. (2025). Affective palatalization in Itelmen. [Conference presentation]. *5th Conference on Uralic, Altaic, and Paleoasiatic Languages*. St. Petersburg, Russia.

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

Chodroff, E., Golden, A., & Wilson, C. (2019). Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization. *Journal of the Acoustic Society of America, 145*(1): EL109–EL115. https://doi.org/10.1121/1.5088035

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

Fortescue, M. (2005). *Comparative Chukotko-Kamchatkan Dictionary*. Mouton de Gruyter.

Gallagher, G. (2010). *The perceptual basis of laryngeal cooccurrence restrictions.* [Unpublished doctoral dissertation]. Massachusetts Institute of Technology.

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*,

Ham, S. (2008). *Tsilhqut'in ejectives: a descriptive phonetic study.* [Unpublished doctoral dissertation]. University of Victoria.

Keating, P., & Garellek, M. (2015). Acoustic analysis of creaky voice. [Conference presentation]. Linguistic Society of America Annual Meeting, Portland, OR.

Kingston, J. C. (1985). *The phonetics and phonology of the timing of oral and glottal events*. [Unpublished doctoral dissertation]. 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.

Moran, S., & McCloy, D. (2019). PHOIBLE 2.0. *Jena: Max Planck Institute for the Science of Human History*, *10*.

Percival, M. (2019). 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. (2024). Production and perception of ejective stops across languages. [Unpublished doctoral dissertation]. University of Toronto.

Volodin, A. P. (1976). *Ительменский язык*. [Itelmen]. Nauka.

Volodin, A. P., Ono, C., Bobaljik, J. D., Koester, D. & Krauss, M. (2021). *Полный ительменско-русский словарь* [Comprehensive Itelmen-Russian Dictionary]. Kulturstiftung Sibirien and Institute of Linguistic Studies, Russian Acadmy of Sciences. Contents available at: Itelmenslovar.org

Warner, N. (1996). Acoustic characteristics of ejectives in Ingush. In *Proceedings 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.

Zimin, M. M., & Mets, T. M. (2025). «Кружочек сверху»: проблема ительменской ‘просодической лабиализации’ в экспериментальной и исторической фонетике [“The little circle above”: the problem of Itelmen “prosodic labialization” in experimental and historical phonetics]. [Conference presentation]. *5th Conference on Uralic, Altaic, and Paleoasiatic Languages*. St. Petersburg, Russia.

---

[1] The coronal fricatives <s,z> have a retracted (postalveolar) articulation and vary somewhat in their phonetic realization: s ~ ʃ, z ~ ʒ. As these are not contrastive, we follow all published sources in omitting the underscore in transcriptions. There is much more variation across sources as to what the symbols given here as <ɸ,x> and <β, ɣ > represent phonetically: transcriptions of individual words vary along a range including ɸ ~ w̥ ~ x^{w} ~ x and correspondingly β ~ w ~ ʋ ~ ɣ^{w} ~ ɣ ~ ɰ (Volodin 1976:33, Bobaljik field notes). It is unresolved how many phonemes these represent, though most authors settle on two voiced and unvoiced fricatives. The palatals ʎ and ɲ are perhaps secondary, occurring only as allophones of l, n before palatals, in loanwords, or as the result of affective palatalization (Bobaljik 2025). See Zimin and Metz (2025) on the Itelmen vowel inventory.

[2] These counts were performed using a sandboxed ChatGPT on a spreadsheet representing a pre-final version of the dictionary. A subset of the ChatGPTs counts were checked manually.

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