PERCEPTUAL CUES TO EJECTIVE STOPS ACROSS LANGUAGES

Maida Percival

University of Toronto
maida.percival@mail.utoronto.ca

ABSTRACT

The goal of this paper is to identify perceptual cues to an ejective - non-ejective stop contrast from Hul'q'umi'num' for listeners from four languages with ejectives: Hul'q'umi'num', Q'anjob'al, Dene, and Eastern Oromo, chosen for their typologically different stop laryngeal contrasts. Results indicated that listeners from each language were similar in perception: all used as primary cue to the perception of ejectives the period of silence following the stop burst and had similar usage of properties of the stop burst and coarticulation in a following vowel as secondary cues. Cross-linguistic differences in perception could mainly be explained based on inventory differences in terms of which other stop types occur across the languages. This is the first study to systematically test perceptual cues to ejectives and has implications for ejective typology.

Keywords: speech perception, ejectives, typology

1. INTRODUCTION

It has been proposed that there are "strong" (or tense or fortis) and "weak" (or slack or lenis) ejective stops in languages, which are distinguished from one another through groupings of acoustic, articulatory, and perceptual characteristics [1, 2]. For perception, the intuitions are that some languages' ejectives sound "weak", less "poppy", or more similar to non-ejective stops than others [3, 4], but these intuitions have never been tested and in fact no prior studies have systematically examined which acoustic cues are used to perceive ejectives. The contribution of this paper is to start to fill this gap by presenting the results of an experiment whose goals are to identify perceptual cues to an ejective - non-ejective contrast from Hul'q'umi'num' and to identify any cross-linguistic differences in the use of these cues between listeners from four languages with ejectives with typologically different stop laryngeal contrasts (Table 1). Two of these languages, Hul'q'umi'num' and Q'anjob'al, also impressionistically differ in the realization of their ejectives: they are thought to be on average strong in Hul'q'umi'num' but weak in Q'anjob'al (c.f. closely related languages [1, 5]).

It is expected that listeners will use as cues acoustic dimensions that differ between ejective and non-ejective stops. If languages have different inventories of stop laryngeal contrasts, then they might also be expected to differ in their use of cues to the two-way Hul'q'umi'num' contrast.

Language (Family)	Contrast
Hul'q'umi'num' (Coast Salish)	T' - T^h
Q'anjob'al (Mayan)	T' - T
Dene[1] (Dene/Athabaskan)	T' - T^h - T
Eastern Oromo (Cushitic)	T' - T^h - D

Table 1: Summary of languages' laryngeal contrasts. T' = ejective, T = voiceless unaspirated, T^h = voiceless aspirated, D = voiced

2. METHODOLOGY

2.1. Participants

100 participants took part in the experiment, with about equal numbers from each language. Table 2 gives the participant details by language. Because some of the languages had few speakers, second language and less fluent first language (L1) speakers were sometimes included as participants in the study. However, all participants grew up hearing their language in their homes and/or community even if they only spoke it a little themselves.

Language	# (# L1)	Gender	Mean age
Hul'q'umi'num'	26 (7)	18F, 8M	56 (24-87)
Q'anjob'al	25 (25)	13F, 12M	33 (19-61)
Dene	24 (19)	17F, 8M	51 (19-76)
Oromo	25 (23)	12F, 13M	42 (17-65)

Table 2: Summary of participants by language

2.2. Materials

A recording of a real-word ejective - non-ejective minimal pair (t^hahw [tax^w] "later, right now, then" and t'ahw [$t'ax^w$] "go downhill to water") by a female speaker of Hul'q'umi'num' was manipulated in Praat [6] to create stimuli for a forced-choice identification task. The two baseline tokens were modified in pieces (burst, post-burst release, following vowel), as described in the following paragraphs. Modification values were based on production [7]. The pieces were cross-spliced to make all possible combinations, a total of 80 stimuli.

Burst: The two bursts (baseline non-ejective and baseline ejective) were manipulated to have two intensities (low, where the mean intensity of the burst was set to 40 dB, and high, where it was set to 50 dB) for a total of four bursts. Burst intensity was manipulated because high intensity bursts are a characteristic of strong ejectives, and so listeners might be expected to use them in perception.

Post-burst release: The baseline non-ejective stop had an aspirated release while the baseline ejective stop had a release of silence following the burst. These two release types (aspiration, silence) were given three durations (0 ms, 50 ms, 120 ms) for a total of five releases (0 ms, 50 ms aspiration, 50 ms silence, 120 ms aspiration, 120 ms silence). 0 ms was included to see how listeners respond when there is no post-burst release: the lack of one is a characteristic of weak ejectives, as well of voiceless unaspirated stops (found in Q'anjob'al and Dene) and voiced stops (found in Oromo).

Following vowel: The two vowels (baseline non-ejective, baseline ejective) were each manipulated to have two F0 patterns (raised, where F0 is raised by 25 Hz for the first 30 ms, and lowered, where F0 is lowered by 25 Hz for the first 30 ms) for a total of four following vowels. After the first 30 ms, the vowels' F0 was set to the mean F0 contour for the speaker who recorded the baseline tokens. The voiceless coda consonant was kept together with its baseline vowel type during cross-splicing (i.e. all tokens with baseline ejective vowels included [x^w] from the baseline ejective [t'ax^w]). Depressed F0 at the vowel onset is a characteristic of weak ejectives and so it was thought that listeners might use vowel F0 to distinguish ejective and non-ejective stops, especially if their language has weak ejectives.

2.3. Procedures

Participants listened to the stimuli in a quiet location: for Hul'q'umi'num', this was at a language school in Duncan, BC, Canada, for Q'anjob'al, this was at a home in Santa Eulalia, Huehuetenango, Guatemala, and for Dene (Déline, NT, Canada) and Oromo (Canada, mostly Toronto), this was typically at the participants' homes. Most participants wore headphones; except some did not have access to any.

The stimuli were presented in random order on a webpage using jsPsych javascript framework. For each trial, participants heard a stimulus play automatically and saw two boxes, one for each word, with the word written in Hul'q'umi'num' orthography. Participants would then select the box with the word that they heard. Most participants completed the task by themselves, but those with accessibility issues indicated their responses by pointing at or verbalizing the word they wanted to select and a researcher would select it for them.

The instructions participants received were to listen and indicate whether the word they heard was tahw [t^{h}ax^w] starting with a plain <t> or t'ahw [t'ax^w] starting with an ejective <t'>. Participants were shown the two words in advance and were played an example recording of them (by a different female speaker). The language groups other than Hul'q'umi'num' also had illustrations included with the text of the words and were given an explanation that there is a language called Hul'q'umi'num', which has ejective sounds reportedly like those of their language and that one of the goals of the study was to see whether they can hear these sounds like they do the equivalent sounds in their own language.

2.4. Statistical analysis

Generalized linear mixed regression models were performed in R [8] using the lme4 package [9] to test the extent of listeners' reliance on the manipulated acoustic dimensions in the stimuli for perception of ejectives. The template for the models is given in 1.

(1)

$$response \sim language * dimension + (1 \mid participant)$$

For each model, the response variable was the participants' response of ejective or non-ejective. The predictor variables were language (4 levels: Hul'q'umi'num', Q'anjob'al, Dene, Oromo) and the manipulated dimension for the given model (i.e., one of release type, release duration, burst type, burst intensity, vowel type, or vowel F0). The dimensions, all simple coded, are summarized in Table 3. The interaction between language and dimension was also included to test if the use of a dimension as a cue varied across languages. Separate models

Dimension	Levels	
Release duration	0 ms	*50 ms* vs. 120 ms
Release type		*silence* vs. aspiration
Burst type	*ejective* vs. non-ejective	
Burst intensity	*high* vs. low	
Vowel type	*ejective* vs. non-ejective	
Vowel F0	*lowered* vs. raised	

Table 3: Summary of dimensions

were done for responses to stimuli with zero (post-burst) release duration (1597 responses) and those to stimuli with non-zero (post-burst) release durations (6392 responses). For language, the reference level was Hul'q'umi'num', since that was the language the stimuli were from, and for all models, the significance level was p <0.05.

3. RESULTS

3.1. Post-burst releases

Figure 1 shows the results for release duration and release type together in one graph. % ejective response is on the y-axis, release duration is on the x-axis, release type is represented by the different line colours, and the responses for each language are grouped by panel. Each dot represents one participant's mean. Because stimuli of zero post-burst release duration have no release type, they are omitted from the models for release dimensions in this section, and separate models are performed on them in the following sections for the burst and following vowel to see if these dimensions play a bigger role in the absence of release type cues.

In general, the responses for each language group in Figure 1 are very similar to one another: listeners perceive as more ejective stimuli with a period of silence while they perceive as more non-ejective stimuli with aspiration. As for stimuli with neither aspiration nor silence, listeners perceived them about equally as ejective or non-ejective.

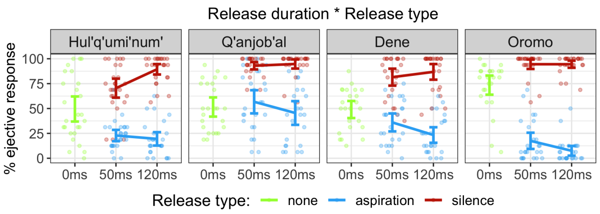

Figure 1: % ejective response to releases by language

Two exceptions to this are reflected across all the models by significant effects of language. First, Q'anjob'al listeners responded with significantly more ejective responses overall than Hul'q'umi'num' listeners for stimuli with non-zero post-burst release durations, as they did not consistently classify stimuli with aspiration (in blue) as non-ejective. Second, Oromo listeners gave significantly more ejective responses for stimuli with zero post-burst release duration (in green).

Figure 2, where dimensions are on the x-axis and languages are represented by line colour, reflects the results of the models for release type and release duration. For release type, stimuli with silence received significantly more ejective responses than those with aspiration (β = -3.468, z = -40.501, p <0.001) in all languages. However, Oromo listeners (in yellow) relied on release type to a greater extent than Hul'q'umi'num' (in red), as indicated by a significant interaction between language and release type (β = -2.433, z = -9.639, p <0.001).

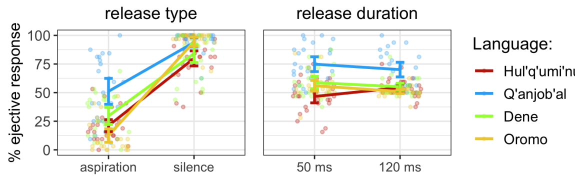

Figure 2: % ejective response to release type and release duration by language

For release duration, no significant difference was found in listeners' responses to stimuli with releases of 50 ms and 120 ms, but there were significant interactions for release duration for Hul'q'umi'num' and each of the other languages. These interactions are a reflection of the patterns seen in Figure 1: for Hul'q'umi'num' there is a greater increase in % ejective response for releases of silence (in red) between 50 ms and 120 ms than for the other languages, while for the other languages the % ejective response to releases of silence is very high (even at 50 ms), but there is a greater decrease in % ejective response for releases of aspiration (in blue) between 50 ms and 120 ms.

3.2. Bursts

Figure 3 (top) shows the results for burst type. In stimuli with post-burst releases, there was a significant effect of burst type (β = -0.416, z = -7.829, p <0.001) and no significant differences across language. For each language, the stimuli with baseline ejective stop bursts were perceived as more ejective than stimuli with baseline plain stop bursts. In stimuli with zero post-burst release duration, the significant effect of burst type remains and in fact listeners use burst type more (β = -1.222, z = -7.829, p <0.001), especially Q'anjob'al listeners (in blue) (β = -0.931, z = -2.71, p = 0.007).

Burst intensity results are shown in Figure 3 (bottom). Burst intensity was significant in stimuli with both non-zero (β = -0.268, z = -9.701, p <0.001) and zero post-burst release durations (β = -0.333, z = -2.85, p = 0.004). Stimuli with high burst intensity were perceived as more ejective than those with low burst intensity. However, there were differences across languages in the extent that this was so in stimuli with non-zero post-burst release durations. Significant interactions for Hul'q'umi'num' compared to Dene (β = 0.471, z = 3.25, p = 0.001) and Oromo (β = 0.469, z = 3.292, p <0.001) suggested that while listeners from Hul'q'umi'num' and Q'anjob'al do use burst intensity, Dene and Oromo listeners do not differ in % ejective response based on high vs. low intensity bursts. These interactions were not significant in stimuli with zero post-burst release duration; Dene listeners used burst intensity similarly to Hul'q'umi'num' and Q'anjob'al listeners, but with Oromo listeners' mean % ejective response (in yellow) being lower with high burst intensity, it's not clear, despite the model, that they use it at all.

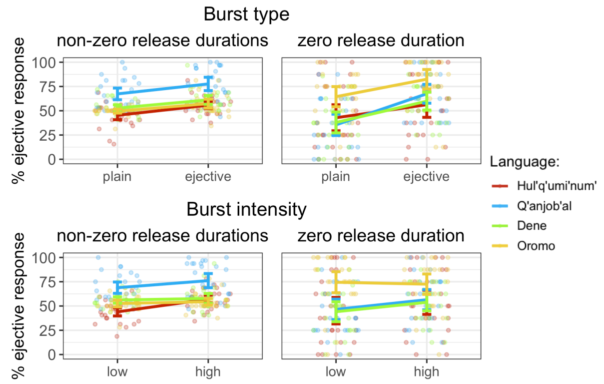

Figure 3: % ejective response to burst type (top) and intensity (bottom) by language in stimuli with non-zero (left) and zero (right) post-burst release durations

3.3. Following vowels

The results for vowel type (Figure 4) resemble those for burst type but with slightly smaller effects. Stimuli with baseline ejective vowels received significantly more ejective responses than those with baseline plain vowels (β = -0.206, z = -3.898, p <0.001). In stimuli with zero post-burst release duration, vowel type's significant effect was greater (β = -0.984, z = -8.005, p <0.001), especially for Q'anjob'al (β = -0.875, z = -2.532, p = 0.011).

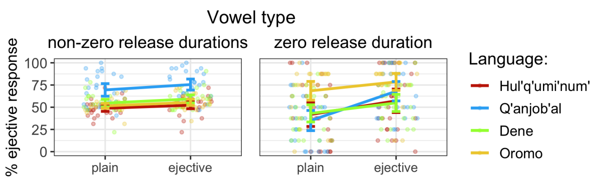

Figure 4: % ejective response to vowel type by language in stimuli with non-zero (left) and zero (right) post-burst release durations

F0 at the vowel onsets was not significant in stimuli with any release duration, nor were any interactions of vowel F0 and language significant. This suggests that vowel F0, at least as manipulated for this study was not used in perception of ejectives.

4. DISCUSSION AND CONCLUSIONS

Listeners from each language were remarkably similar in perception: they all used as primary cue to ejectives the period of silence following the stop burst. Burst type (likely correlated with spectral properties of the burst and/or change in amplitude over time) and vowel type (likely correlated with creaky (ejective) and breathy (aspirated) voice quality from stop coarticulation) were secondary cues in each of the languages, that listeners relied on even more in stimuli with no period of silence. Burst intensity was also a secondary cue but perhaps as it encompassed only a single acoustic dimension, it was used less consistently.

While both strong (long release, loud burst) and weak (creaky vowel) ejective characteristics consistently cued ejectives, small differences across language groups were present, which can be attributed to their inventories of laryngeal stop types. Q'anjob'al listeners more often hearing aspiration as ejective is likely due to the lack of aspirated stops in their language, and their greater burst type and vowel type cue usage in stimuli without aspiration or silence may reflect greater exposure to short stops, given that Q'anjob'al only has voiceless unaspirated pulmonic stops. Dene and Oromo listeners relying less consistently on burst intensity may be due to intensity being too subtle to consistently differentiate three stop laryngeal types. Oromo listeners classifying stimuli with no post-burst release as mainly ejective may relate to their lack of voiceless unaspirated pulmonic stops (their only pulmonic stops without post-burst releases are voiced). Their greater use of release type may also be evidence of the importance of aspiration to voiceless non-ejective stops for them. An alternate explanation for Oromo's greater use of release type, and Q'anjob'al's of vowel and burst type is demographics: Hul'q'umi'num' and Dene's L1 listeners being fewer and older than Q'anjob'al and Oromo's may have led to less consistent cue usage.

Given that the present study was based on non-native speech for three language groups, further experiments of language-specific perception and production are underway for each language to add to this study's main finding: that there seems to be little evidence for cross-linguistic differences in the perception of ejectives, outside of how they contrast within a language's stop laryngeal type inventory.

5. REFERENCES

[1] J. Kingston, "The Phonetics and Phonology of the Timing of Oral and Glottal Events," Ph.D. dissertation, University of California Berkeley, 1985.

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

[3] S. Bird, "Dakelh ejectives: Evidence for new ways of classifying sounds," 76th Meeting of the Linguistic Society of America.San Francisco, CA, 2002.

[4] P. D. Fallon, *The Synchronic and Diachronic Phonology of Ejectives*. New York, NY: Routledge, 2002.

[5] S. Russell, "Some acoustic characteristics of pulmonic and glottalic stops in Mam," Master's thesis, Simon Fraser University, 1997.

[6] P. Boersma and D. Weenink, "Praat," http://www.praat.org, University of Amsterdam, 1992−2017. [Online]. Available: ttp://www.fon.hum.uva.nl/praat/

[7] M. Percival, "Contextual variation in the acoustics of Hul'q'umi'num' ejective stops," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, M. T. . P. W. Sasha Calhoun, Paola Escudero, Ed., 2019, pp. 3270–3274.

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

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

[1] This language is also called Sahtúot'ine kedé by its speakers in Déline but I shall refer to it as Dene in this paper, because this term is shorter and because it is a term now used for both the language family and the particular language.