INTERSPEECH 2009 BRIGHTON

Acoustic Characteristics of Ejectives in Amharic

Hussien Seid[1], *S. Rajendran*[2], *B. Yegnanarayana*[3]

International Institute of Information Technology Hyderabad, India

[1]hussien@research.iiit.ac.in, [2]su.rajendran@gmail.com, [3]yegna@iiit.ac.in

Abstract

In this paper, a preliminary investigation of the acoustic characteristics of Amharic ejectives in comparison with their unvoiced conjugates is presented. The normalized error from linear prediction residual and a zero frequency resonator output are used to locate the instant of release of the oral closure and the instant of the start of voicing, respectively. Amharic ejectives are found to have longer closure duration and smaller VOT than their unvoiced conjugates. Cross-linguistic comparisons reveal that no ejectives of two languages behave acoustically in a similar manner despite similarity in their articulation.

Index Terms: Amharic, ejectives, glottalized stops

1. Introduction

Ejectives are sounds rarely found in the languages of the world. In a survey of glottalized consonants distribution, Ian Maddieson has reported that only 16.3% of the world's languages have ejective sounds [1]. Amharic, which is a Semitic language, and probably the second largest spoken language next to Arabic in its family [2], has five ejectives: four stops /p'/, /t'/, /ʧ'/, /k'/ and one dental fricative /s'/. In Amharic, which is the official language of Ethiopia, each ejective has unvoiced and voiced conjugates, i.e., sounds with the same place of articulation [3]. Table 1 shows the list of ejectives with their respective conjugates according to the place of articulation.

Ejective consonants are produced in a similar vocal tract configuration of either their voiced or unvoiced conjugates. The difference is in the source configuration. When ejectives are produced the vocal folds are kept tightly closed so that no air comes from the lungs to the oral cavity. Hence the air in the oral cavity is used to produce the sound. Unlike the voiced and unvoiced stop sounds, in the production of ejectives, the piston-like upward movement of the larynx creates the pressure needed for excitation, and pushes the air to the place of articulation [4].

In this paper we analyze the acoustic characteristics of ejectives and compare with their unvoiced conjugates. Voiced conjugates are not included in the analysis, as they have negative voice onset time (VOT) and their voicing nature makes them distinct from both the unvoiced and ejective sounds.

Table 1: Ejectives with unvoiced (left column in each place of articulation) and voiced (right column) conjugates in Amharic.

Manner	Place of articulation			
	Labial	Dental	Alveolar	Velar
Plosive	p b	t d	ʧ ʤ	k g
Ejective stop	p'	t'	ʧ'	k'
Fricative		s z		
Ejective fricative		s'		

The paper is organized as follows: Section 2 describes the data collected for this study. Section 3 describes the analysis techniques used to study the acoustic characteristics of ejectives. The results of the study of ejectives in relation to their unvoiced conjugates are discussed in Section 4. In Section 5, the acoustic characteristics of ejectives in Amharic are compared with the studies of ejectives in other languages. Section 6 summarizes the study reported in this paper.

2. Speech material for the study of ejectives in Amharic

Some words of minimal pairs for each ejective and its unvoiced conjugate are collected from a free online dictionary of the Amharic language [5]. Data in different vowel environments was collected. Since labial ejective and labial unvoiced stop sounds are not native Amharic sounds, they are found only in borrowed words from different languages [3]. Hence, for labials, we took some words with at least one similar immediate neighboring phoneme as in /papaya/ and /p'ap'as/.

Each identified word is recorded while it is read by a native speaker (male, age 34) of the language twelve times continuously with a brief pause after each utterance. It was ensured that the speaker produced the words in natural constant tone and speaking style. To avoid the prosodic effects at the beginning and end of an utterance, the first and the last words are ignored in the recorded files.

The recording is carried out in a computer laboratory using an ordinary head phone set (TECH-COM SSD-HP-202) and an ACER Desktop computer with a Windows XP Sp2 operating system. For recording and spectral analysis, the *WaveSurfer* tool configured to 16-bit, 16 kHz sampling frequency was used with default settings. All other analysis were carried out using MatLab 7.0 version.

For the analysis of ejectives, two parameters are used: voice onset time (VOT) and closure duration. The VOT is defined as the time interval from oral release at the constriction of the stop sound to the first visible periodicity of the following vowel. The VOT for ejectives is also defined as the duration from oral release to the glottal release [6]. But since the voicing may start after some time delay from the instant of the glottal release our analysis is based on the time interval from oral release to the onset of periodicity. It is also important to include the gap between the glottal release and the delayed onset time of voicing, which is an important cue to identify ejectives [7]. The closure duration is measured as the time interval from the end of the previous vowel to the oral release.

These two measurements were done for singleton and geminated phonemes separately. For fricatives, only the overall duration is measured for singleton in the middle of the word (MOW), singleton at the beginning of the word (BOW) and geminated(GEM) cases. The mean (μ) and standard deviation

(σ) of each measurement for each phoneme as per its place of articulation are obtained, and the statistical significance of the difference between ejectives and their unvoiced conjugates was checked using the T-test and F-test, respectively, for a significance level of 95%.

3. Measurement of VOT and closure duration

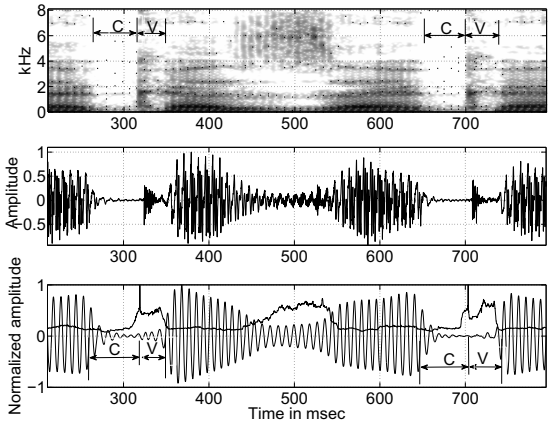

Figure 1: Wideband spectrogram (top panel), speech signal (middle panel) and zero frequency resonance filtered signal (bottom panel) with normalized LP residual for the word /zəkəzəkə/.

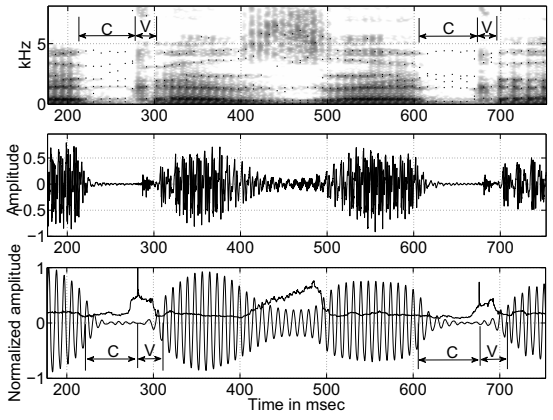

Figure 2: Wideband spectrogram (top panel), speech signal (middle panel) and zero frequency resonance filtered signal (bottom panel) with normalized LP residual of the word /zək'əzək'ə/.

To measure the VOT and closure duration of ejectives and other stops, it is necessary to determine the instant of burst release and the onset of glottal activity precisely. It is difficult to obtain these instants precisely from the waveform or a spectrogram (wideband or narrowband). The instant of glottal activity is difficult to observe due to burst and frication noise or due to background noise. Likewise the block processing of data in spectrum analysis smears the abrupt changes due to excitation source and vocal tract dynamics.

Recently, filtering of speech by zero frequency resonator was proposed to locate the instants of significant excitation of the vocal tract system, which are mostly caused by the glottal activity, and in particular due to rapid closure of the vocal folds in each glottal cycle. The method gives precise location of onset of glottal activity and hence periodicity. The instant of burst release can be located from the normalized error computed from linear prediction residual. The normalized error and the zero frequency filter output after DC removal are obtained as follows [8, 9]:

1. Difference the signal $s[n]$

$$x[n] \;=\; s[n] \;-\; s[n-1], \tag{1}$$

2. Calculate the LP residual, $e[n]$ as follows:

$$e[n] \;=\; x[n] \;-\; \sum_{k\,=\,1}^{p} a_k x[n-k], \tag{2}$$

where p is the LP order and a_k's are the LPCs. The LPCs are obtained by solving the following autocorrelation normal equations:

$$\sum_{k\,=\,1}^{p} a_k R[m-k] \;=\; -R[m] \quad m \;=\; 1, 2, 3, ..., p,$$

where $R[m]$ is autocorrelation sequence of the signal $x[n]$.

3. Calculate the normalized error $\eta[n]$ as follows:

$$\eta[n] \;=\; \frac{\sum_{m\,=\,n\,-\,N/2}^{n\,+\,N/2} e^2[n+m]}{\sum_{m\,=\,n\,-\,N/2}^{n\,+\,N/2} x^2[n+m]}, \tag{3}$$

where $N+1$ is the total number of samples in each analysis frame.

4. Pass the residual signal $e[n]$ through the zero frequency resonance filter twice

$$y_1[n] \;=\; -\sum_{k\,=\,1}^{2} a_k y_1[n-k] \;+\; e[n], \tag{4}$$

and

$$y[n] \;=\; -\sum_{k=\,1}^{2} a_k y[n-k] + y_1[n], \tag{5}$$

where a_1 = -2 and a_2 = 1.

5. Remove the trend in $y[n]$ by subtracting the mean of $y[n]$ over 10 msec at each sample, and the resulting signal $z[n]$ is the zero frequency filtered signal, or the filtered signal.

$$z[n] \;=\; y[n] \;-\; \frac{1}{2N+1} \sum_{m\,=\,-\,N}^{N} y[n+m], \tag{6}$$

where $2N+1$ is the length of the 10 msec window.

Figs. 1 & 2 show the waveform, spectrogram, filtered signal for the LP residual and the normalized errror for a word with unvoiced stop and ejective respectively. The positive zero crossing instants correspond to instants of significant excitation. From

the figures it is obvious that the burst release can be detected from the normalized error plot by its prominence in amplitude, and the onset of glottal activity from the filtered signal by the appearance of periodicity. The time interval between burst release and onset of glottal activity is the VOT (as marked as *V*) in Figs. 1 & 2 bottom panel). The closure duration (*C*) is measured from the last significant excitation of the preceding vowel to the burst release.

These parameters are also derived from the waveform and spectrogram as marked in Figs. 1 & 2 (top panel). We call these parameters as those derived from spectral method. The voice onset time and closure duration derived from normalized error and the filtered signal is referred to as non-spectral method.

For the spectral method, the closure duration of both the ejective and unvoiced stop is marked as the duration of very low spectral energy after the preceding vowel. The duration from a burst of broad spectral energy to the first visible periodic vertical striation of the spectrogram is used to mark VOT. The amplitude and the periodicity of the waveform is also used as suplementary information for the decision to mark the closure duration and the VOT.

4. Results and discussion

Both spectral and nonspectral methods are used for studying the acoustic characteristics of ejectives. As can be seen from Table 2, although both the methods give almost similar mean values, the spectral method (results shown in parentheses) gives larger variation in the parameters (i.e., standard deviation) than the non-spectral method. Hence, for further analysis the nonspectral method is used.

In this experiment, it is found that ejectives have less duration of frication than their unvoiced conjugates. This may be because the amount of air for excitation is limited to the volume of the oral cavity, unlike the unvoiced consonants, where the air comes from the lungs. As can be seen in Fig. 2, it is also observed that the amplitude of the vowel following the ejectives grows more slowly than for the unvoiced phonemes. Since the vocal folds are closed during the production of ejectives, it takes time to build up to the normal vibration mode for the production of following vowel. This in turn may be the cause of creaky phonation.

All ejectives are found to be followed by a creaky (irregular periodicity) vowel, which is the result of change of the state of vocal cords from complete closure to normal phonation [7]. Domolin [10] reported that ejective stops have a higher amplitude burst than pulmonic stops, but our data did not confirm that result. However, fricative ejectives have shown 6-8 dB higher energy than that of the unvoiced fricative conjugates. We have found that the ejective stops have significantly less VOT and longer closure duration than their unvoiced conjugates in both singleton and geminated cases (Figs. 3 and 4). But the overall duration measurements are not as systematic as VOT and closure measurements. Velar stops in both cases and dental stops in singleton case have shown no significant difference, while other ejectives, with the exception of alveolar stops, have longer overall duration than their unvoiced pairs. However, alveolar ejectives seem to have less duration than that of the unvoiced conjugates.

The ejective fricatives also seem to have less duration than the unvoiced fricatives. The speaker was asked to sustain the articulation of fricatives in /is/ and /is'/ for longer duration without a break. While he could produce /is/ for 10 sec without any difficulty, he could not produce /is'/ longer than about 0.6 sec.

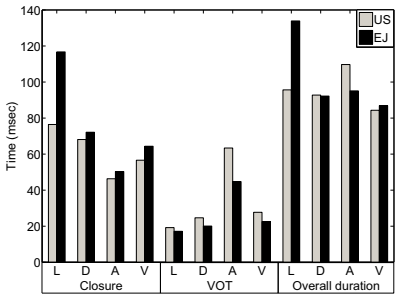

Figure 3: Closure, VOT and overall duration difference between ejective stops and their unvoiced conjugates (singleton) (L: Labial, D: Dental, A: Alveolar and V: Velar).

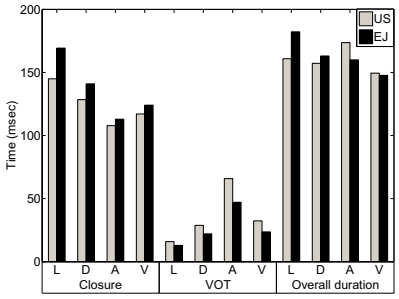

Figure 4: Closure, VOT and overall duration difference between ejective (EJ) stops and their unvoiced (US) conjugates (geminated).

This is because the air to produce ejectives is limited to the vocal tract volume only. Further, our results show that fricatives in general have significantly longer overall duration when they appear at the beginning than in the middle of a word, and even longer when they are geminated (Table 3).

5. Cross-linguistic comparison

To compare the Amharic ejectives with the ejectives of other languages, we have adopted Warner's table [4], and we appended observation on our data to it (Table 4).

In general, Amharic ejectives are followed by creaks for a few periods in the following vowel. Moreover, the amplitude of the following vowel also grows slowly as in Ingush and Quiché but Amharic ejectives are better characterized by low velar VOT and higher average closure to VOT ratio which is not reported for any other languages. In Kiowa, it is reported that ejectives have longer VOT than that of unvoiced stops, which is contrary to Amharic ejective stops [11]. In Amharic, ejectives are immediately followed by the next vowel after the glottal release. Hence short VOT and no silence before the start of the following vowel is observed.

Table 3: Overall duration measurements (in msec) of fricatives. Numbers in parentheses are standard deviations.

Fricative	Overall Duration (msec)		
	Middle	Beginning	Geminated
Unvoiced /s/	96.10 (7.35)	158.93 (16.92)	174.21 (7.87)
Ejective /s'/	85.31 (7.31)	117.10 (6.38)	160.79 (11.25)

Table 2: Closure duration, VOT and overall duration (in msec) from non-spectral analysis and spectral analysis. The numbers in parentheses are from spectral analysis.

Sound	Measurement	Closure		VOT		Duration	
		sin.	gem.	sin.	gem.	sin.	gem.
/p/	μ	76 (78)	145 (144)	19 (20)	16 (17)	95 (98)	161 (161)
	σ	7.1 (7.1)	5.5 (9.9)	2.2 (7.1)	2.4 (6.1)	6.9 (11.1)	6.0 (11.2)
/t/	μ	68 (65)	128 (128)	25 (25)	29 (29)	93 (90)	157 (157)
	σ	5.8 (11.5)	5.2 (7.9)	2.3 (6.5)	5.3 (8.8)	6.7 (14.5)	7.3 (11.7)
/ʧ/	μ	46 (44)	108 (107)	63 (62)	66 (65)	109 (106)	173 (172)
	σ	6.9 (7.5)	5.3 (8.7)	5.4 (11.0)	4.8 (9.5)	8.9 (13.1)	7.0 (12.3)
/k/	μ	57 (55)	117 (159)	28 (27)	32 (32)	84 (83)	149 (192)
	σ	5.3 (9.3)	8.3 (8.4)	4.8 (9.1)	5.3 (9.0)	6.8 (12.4)	10.8 (13.6)
/p'/	μ	117 (118)	169 (172)	17 (16)	13 (12)	134 (134)	182 (184)
	σ	6.0 (6.1)	4.1 (6.9)	1.5 (6.5)	1.5 (6.3)	6.3 (9.3)	4.0 (10.1)
/t'/	μ	72 (71)	141 (137)	20 (20)	22 (22)	92 (91)	163 (159)
	σ	3.2 (8.5)	5.7 (8.6)	1.3 (4.7)	2.7 (8.0)	6.7 (11.0)	6.0 (12.2)
/ʧ'/	μ	50 (48)	113 (111)	45 (46)	47 (47)	95 (95)	160 (159)
	σ	3.1 (7.2)	5.3 (8.7)	4.3 (8.3)	4.5 (7.3)	5.6 (11.5)	6.3 (11.0)
/k'/	μ	64 (66)	124 (139)	22 (16)	24 (24)	87 (81)	148 (164)
	σ	4.3 (8.5)	6.3 (7.8)	4.5 (5.6)	4.3 (5.4)	6.8 (10.9)	8.2 (10.9)

Table 4: Cross linguistic comparison of ejective sounds (Ingush, Hausa, Quinche, Navajo, Tigrinya) [4], Waima's [7].

Criteria	Ingush	Hausa	Quiche	Navajo	Tigrinya	Amharic	Waima'a
F0 of the following vowel	Higher		Lower		Higher		
Approx. VOT-velar	50 msec	25 msec	50 msec	80 msec	$\sim$80 msec	$\sim$23 msec	66 msec
Ratio of closure to VOT	1.94	$\sim$2		$\sim$1.5		$\sim$3	$\sim$1.7
Post-burst noise amplitude	Low	Low	Low	Low	Silence	Low	
Rise to full amplitude of vowel	Slow	Normal	Slow	Very fast	Fast	Slow	
Voice quality onset of vowel	Aperiodic	Aperiodic	Creaky	Modal	Modal	Creaky	Creaky

6. Summary and conclusion

Amharic is one of the few languages to have ejectives among the world languages. It has 4 ejective stops and one ejective fricative. The VOT and closure duration were measured for both types of ejectives and their stop conjugates for comparison. As conventional spectral analysis can not provide the instants of burst release and the onset of voicing precisely, the present study used non-spectral analysis (zero frequency resonator and normalized LP residual) to locate these significant instants. The results obtained by both the methods were presented and compared. The non-spectral analysis provides less variation compared to the spectral analysis. The ejective stops have significantly less VOT and larger closure duration than their unvoiced conjugates. The ejective fricatives have longer overall duration at the word initial position than at the word medial position. The results obtained for the Amharic ejectives were juxtaposed with the studies of ejectives in other languages for cross-linguistic comparison.

7. References

[1] I. Maddieson, "Glottalized consonants," in *World Atlas of Language Structures*, M. Haspelmath, M. S. Dryer, D. Gil, and B. Comrie, Eds. Oxford: Oxford University Press, 2005, pp. 34–37.

[2] K. Hayward and R. J. Hayward, "Amharic," in *Handbook of the International Phonetic Association*. Cambridge: International Phonetic Association, 1999.

[3] W. Leslau, *Reference Grammar of Amharic*. Wiesbaden: Otto Harrassowitz, 1995.

[4] N. Warner, "Acoustic characteristics of ejectives in Ingush," in *International Conference on Spoken Language Processing*, 1996, pp. 1525–1528.

[5] [Online]. Available: http://www.amharicdictionary.com

[6] T. Cho and P. Ladefoged, "Variation and universals in VOT: evidence from 18 languages," *Journal of Phonetics*, vol. 27, pp. 207–229, 1999.

[7] M. Stevens and J. Hajek, "A preliminary investigation of some acoustic characteristics of ejectives in Waimaa: VOT and closure duration," in *Proceedings of the Tenth Australian International Conference on Speech Science and Technology*, S. Cassidy, F. Cox, R. Mannell, and S. Palethorpe, Eds. Sydney: ASSTA, 2004, pp. 277–282.

[8] B. Yegnanarayana, K. S. R. Murty, and S. Rajendran, "Analysis of stop consonants in Indian languages using excitation source information in speech signal," in *Proceedings of ISCA ITRW Speech analysis and processing for knowledge discovery*, P. Dalsgaard, C. F. Pedersen, and O. Andersen, Eds., Aalborg, Denmark, 2008, p. 20.

[9] K. S. R. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 16, no. 8, pp. 1602–1613, 2008.

[10] D. Demolin, "Acoustic and aerodynamic characteristics of ejectives in Amharic," *Journal of the Acoustical Society of America*, pp. 2610–2610, 2004.

[11] R. Billerey-Mosier, "An acoustic analysis of Kiowa stops," 2003, UCLA Field methods term paper.