INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Acoustic Characterization of Word-final Glottal Stops in Mizo and Assam Sora

Sishir Kalita[1], *Wendy Lalhminghlui*[2], *Luke Horo*[2], *Priyankoo Sarmah*[2,3], *S R Mahadeva Prasanna*[1,3], *Samarendra Dandapat*[1]

[1]Department of Electronics and Electrical Engineering
[2]Department of Humanities and Social Sciences
[3]Center for Linguistic Science and Technology
Indian Institute of Technology Guwahati
Guwahati 781039, India

(sishir, wendy, luke, priyankoo, prasanna, samaren)@iitg.ernet.in

Abstract

The present work proposed an approach to characterize the word-final glottal stops in Mizo and Assam Sora language. Generally, glottal stops have more strong glottal and ventricular constriction at the coda position than at the onset. However, the primary source characteristics of glottal stops are irregular glottal cycles, abrupt glottal closing, and reduced open cycle. These changes will not only affect the vocal quality parameters but may also significantly affect the vocal tract characteristics due to changes in the subglottal coupling behavior. This motivates to analyze the dynamic vocal tract characteristics in terms of source behavior, apart from the excitation source features computed from the Linear Prediction (LP) residual for the acoustic characterization of the word-final glottal stops. The dominant resonance frequency (DRF) of the vocal tract using Hilbert Envelope of Numerator Group Delay (HNGD) are extracted at every sample instants as a cue to study this deviation. The gradual increase in the DRF and significantly lower duration for which subglottal coupling is occurring is observed for the glottal stop region for both the languages.

Index Terms: glottal stops, Mizo, Sora, glottal stop detection

1. Introduction

A glottal stop is a pulmonic voiceless plosive which is produced with full or partial closure of the glottis by the vocal folds and also, sometimes with the constriction of the false vocal folds [1–3]. Glottal stops vary in terms of glottal constriction depending on individual speaker's voice quality, their position in the syllable and also at the spoken phrase. For example, in case of English speakers, it is observed that unlike in the onset position, a glottal stop in the coda position of a syllable is produced with additional ventricular constriction as voicing needs to be halted [2].

The present study aims to investigate glottal stop characteristics in two languages namely Mizo and Assam Sora. Mizo is a tonal language of the Tibeto-Burman language family, spoken in Mizoram, India by approximately one million speakers. On the other hand, Assam Sora is a South Munda language of the Austroasiatic family spoken primarily in Assam, India by approximately forty thousand people. Glottal stops in both the languages are phonemic, however; their distribution in the two languages is distinctive. While the glottal stop in Assam Sora occurs word medially and word finally, it occurs only word finally in Mizo. As mentioned earlier, Mizo is a tonal language with a tonal inventory of four lexical tones, namely, High, Low, Falling and Rising tone [4]. Thereby, like in most of the tonal languages, glottal stop in Mizo is commonly associated with the Low tone and it acts as a tone depressor that lowers the pitch of a word [4]. Additionally, it prohibits long vowels in syllables [5]. On the other hand, analysis of intervocalic glottal stops in Assam Sora has revealed that, intervocalically a glottal stop in Assam Sora may have three phonetic realisations (a) a complete glottal stop (b) a glottal stop with creaky phonation or (c) a voiced glottal stop [6]. Additionally, the intervocalic glottal stops show two types of pitch contours: a rising contour and a dipped contour [6].

From the literature it is observed that prior attempts to study the acoustic characteristics of word-final glottal stops were explored only for glottal stops emerging as an allophonic variation of the phoneme */t/* . For instance, the acoustic difference between /t/ glottatization and phrasal creak was studied previously using vocal quality features [7, 8]. Similarly, duration and quality of word-final glottal stops in Voro language was studied [9]. As far as acoustic characteristics of glottal stops are concerned, it has been reported that abrupt glottal closing, significant irregularities between the glottal cycles, and reduced open phase duration are the primary characteristics [2, 10–12]. In a previous study on Assam Sora, glottal stops in the intervocalic context was investigated by using various temporal and spectral voice source features [6].

There are no acoustic studies that have explored the word-final glottal stops in Mizo and Assam Sora. Hence, their characteristics in the two languages are largely unknown. Secondly, although the deviation in source characteristics has been studied in general, the affect of changes in the source dynamics in vocal tract characteristics (VTC) during a glottal stop production has not been explored in other languages as well. It has been reported previously that, final glottal stops need more glottal and ventricular constriction [2], leading us to assume that there will be a reduction in the effective length of the vocal tract during the closed phase of the glottal cycle. It is also reported that the opening to closing phase transition is considerably abrupt, with the open phase often reducing, during the glottal stop production [2, 10, 13]. In case of modal phonation, it is observed that the subglottal coupling during the open phase changes the VTC in terms of reducing the dominant resonance frequency (DRF) of the vocal tract [14]. The reduction of the open phase and increased constriction in the laryngeal level will change the characteristics of DRF in the glottal stop. This motivates us to study the acoustic characteristics of the word-final glottal stops, exploiting the changes in VTC, arising due to the changes in glottal source behavior.

The dominant resonance frequency of the vocal tract sys-

http://dx.doi.org/10.21437/Interspeech.2017-604

tem is derived using Zero Time Liftering (ZTL) based spectral estimation method. Their characteristics are extracted from the glottal stop region at every sample instant and subjected to further analysis [15]. Apart from that, excitation source characteristics, in terms of strength of excitation (SoE) and H1-H2 ratio, computed from the linear prediction (LP) residual, are also considered for the present study. Finally, as the realization of glottal stops vary widely across languages, this work also compares the word-final glottal stops in two unrelated languages, namely Mizo and Assam Sora, using the extracted acoustic features.

The rest of the paper is organized as follows: Section 2 provides a brief description of the database used in this study, Section 3 describes the acoustic analysis of word-final glottal stops in terms of dynamic VTC and excitation source information. Experimental results are presented in Section 4 and finally, Section 5 presents the summary of the present work and outlines possible future directions.

2. Speech database

The Mizo data used for the present study is obtained from four native Mizo speakers, who were asked to produce a set of 56 words (see 2.2 for details). The speakers were recorded in a sound attenuated booth using a Tascam linear PCM recorder connected to a unidirectional Shure head-worn microphone in the Phonetics and Phonology Lab of IIT Guwahati. Annotation of the speech data was done manually using Praat. Assam Sora speech data is from an existing speech database that was recorded in Sessa and Singrijhan tea estates of Assam and subsequently curated by the third author of the present study. Although the Assam Sora speech data is recorded in field, the recording equipment and the data annotation procedures are similar to that for the Mizo speech data and the recording quality is comparable. All recordings were done at 44.1 kHz sampling rate.

2.1. Subjects

Four native speakers of Mizo (2 female and 2 male) from Mizoram who are currently pursuing post graduation degree in Assam participated in this study. They had a mean age of 25 years and were reported to be free from any speech and hearing disorders. On the other hand the speech data for Assam Sora included data from fourteen Assam Sora speakers (5 female and 9 male) from Sessa and Singrijhan tea estates of Sonitpur district in Assam. The Assam Sora speakers had a mean age of 53 years and were reported to be free from any speech and hearing disorders.

2.2. Stimuli

Speech data for Mizo consists of a word list of 56 unique words that includes 33 words with CVC syllables and 23 words with CV syllables. While the data with CV syllables do not have a glottal stop, in CVC syllables, the glottal stop occurs in the coda position. Consonants in the onset position were /t,d,m,l,r,s,z/ and syllable nucleus were monophthongs /e,i,a,o,u/. Moreover, each word was produced three times in isolation by 4 native Mizo speakers resulting in 672 tokens used for this study. On the other hand the Assam Sora speech data with word-final glottal stop consists of 84 tokens including 27 unique words that have the VC.CVC, V.CVC, CV.CVC and CVC.CVC syllables, where coda of the second syllable is always a glottal stop and coda of the first syllable is a sonorant consonant. Whereas, Assam Sora data with open syllable includes a set of 21 unique words having syllable types CV.CV and CVC.CV.

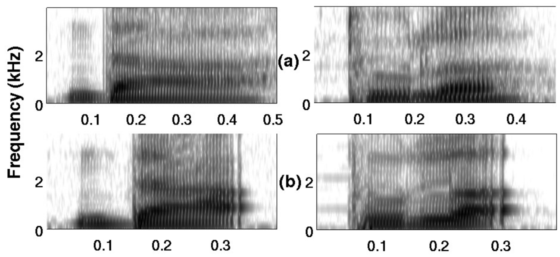

Figure 1: *Spectrogram of speech signal without glottal stop ((a), (c)) and with glottal stop ((b), (d)) for Mizo and Assam Sora respectively. Mizo samples: (a) /da:/ (low tone), (b) (daʔ). Assam Sora samples: (c) /tula/, (d) /tulaʔ/. Strong glottal irregularity and reduced pitch are observed in the glottal stop region for both languages.*

3. Methods for acoustic analysis

Abrupt cessation of voicing is reported to be the primary acoustic cue for the word-final glottal stops. The abrupt cessation is achieved by increased glottal and ventricular constriction, requiring more glottal efforts [2]. Such increased articulatory efforts affect the duration of the preceding vowel inversely. This phenomenon is clearly noticed in case of Mizo and Assam Sora as demonstrated in Figure 1. Spectrograms of syllables, with and without glottal stops, in Mizo and Assam Sora, are shown in Figure 1. The spectrograms clearly show the evidence of abrupt cessation of voicing and resultant reduced vowel duration in the glottal stop contexts. In this section, apart from the features mentioned above, word-final glottal stop region is also studied in terms of dynamic VTC and excitation source features.

3.1. Dynamic vocal tract characteristics in terms of source behavior in glottal stop regions

Vocal tract characteristics change, with respect to different phases of the glottal cycle, during the production of voiced sounds [14]. During the open phase of the glottal cycle, the supraglottal cavity is coupled with the subglottal cavity, resulting in an increase in the effective length of the vocal tract. This decreases the first formant frequency (F1) during this phase as compared to the closed phase duration of the glottal cycle, when the F1 is only from the supraglottal cavity. [14]. Generally, F1 is considered to be the dominant resonance frequency (DRF) of the vocal tract [14]. However, irregularity of the glottal cycles, abruptness in the glottal closing and reduced open quotient during the glottal stop production will affect the characteristics of the DRF, unlike in the case of modal phonation.

As the current work focuses on the analysis of changes in VTC due to the changes in different phases of glottal cycle in glottal stops, a spectral estimation method with high temporal resolution is required. For this purpose, Zero Time Liftering (ZTL) based spectral estimation method is used to derive the vocal tract spectral characteristics at every sample instant [14,15]. In this method, speech signal is first multiplied by a highly decaying impulse like window function to achieve better temporal resolution. However, this window function drastically smoothens the formant evidences. Therefore, to highlight the formant peaks, high resolution and additive properties of the Hilbert envelope of numerator group-delay (HNGD) function is exploited and the corresponding spectrum is termed as HNGD spectrum [15]. In this work, the HNGD spectrum is computed using a window of 4 msec at every sample instant. For the com-

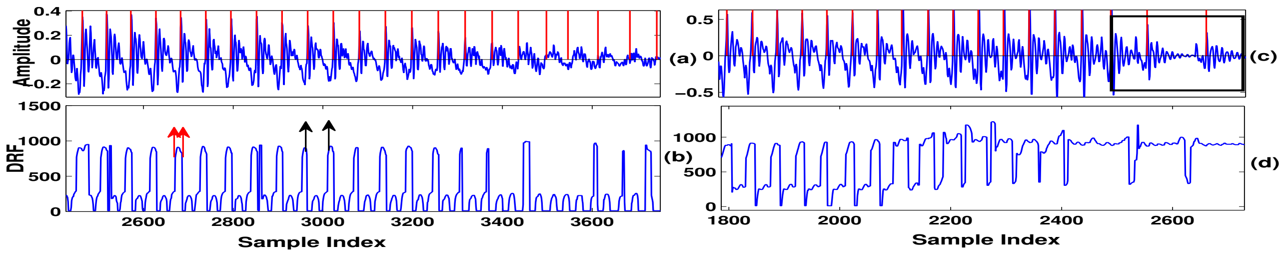

Figure 2: *Dominant resonance frequency (DRF) contours. (a)-(b), and (c)-(d) speech signal and DRF contour for non-glottal stop and glottal stop contexts respectively.*

putation of DRF, the HNGD spectrum is smoothened by using a 4-point median filter at each sample instant. The strongest spectral peak corresponding to the positive to negative zero crossing of the differenced smoothened HNGD spectrum is taken as DRF. Further, the contour of DRF is smoothened by using a 5-point median filter, in order to remove outliers.

The DRF contour of speech signal for the /da:/ sample in Mizo is shown in Figure 2(b). The increased DRF value between the two red arrows in DRF contour (Figure 2(b)), corresponds to the closed phase of the glottal cycle. However, the region between the black arrows in the same figure corresponds to the opened phase of the glottal cycles, where, the DRF value is relatively low due to subglottal coupling. This uniformity in the DRF contour can be observed in the /daʔ/ sample in Mizo in the initial to middle part of the vowel. The following DRF contour characteristics are considerably changed due to changes in the source behavior (Figure 2). In the initial region of /daʔ/, the lowered DRF regions between two adjacent GCIs are uniform, whereas, a gradual decrease is observed approximately from the 2200 sample index. This signifies the reduced open phase during the production of the glottal stop. Another remarkable observation is that, DRF value during the open phase is higher for the final region then for the initial region. The gradual decrease of the open phase region and increase in the DRF value for the closed region before the glottal stop signifies the intention to stop the voicing with more laryngeal constriction that starts well before the glottal stop is actually produced. However, these characteristics are not observed in the non-glottal stop context. The increase in the duration of the DRF value is more in the glottal stop region which exhibits an increase in the closed phase. To quantify the increased DRF value during the word-final glottal stop region, the mean DRF value within two glottal closure instants (GCIs) is derived. The GCIs are estimated using the zero frequency filtering (ZFF) approach [16]. Figure 3 shows the mean DRF value for the non-glottal stop and glottal stop samples of Mizo. The figure clearly shows the gradual increase of the DRF values in the glottal stop context, (Figure 3(d)) whereas, a gradual decrease in the non-glottal stop context (Figure 3(b)).

3.2. Source features

Apart from the analysis of changes in VTC arising due to source behavior, the source features estimated from the excitation source signal are also used to characterize word-final glottal stops. Earlier studies have shown the significance of source features for the characterization of intervocalic glottal stops in Assam Sora [6]. Therefore, the temporal source feature, strength of excitation (SoE) is used to characterize the abrupt glottal closure in the glottal stops. Also, to characterize the deviation in the source spectrum, the ratio of the first and the second harmonics (H1-H2) is computed, which represents the spectral slope of the source spectrum. Both these features are extracted epoch synchronously from the LP residual. For the computation of SoE, the first GCIs are detected and a segment of 3 ms from the HE of the LP residual around each GCI location is calculated [17]. The ratio between mean and standard deviation of the samples of the segment is considered to be the SoE.

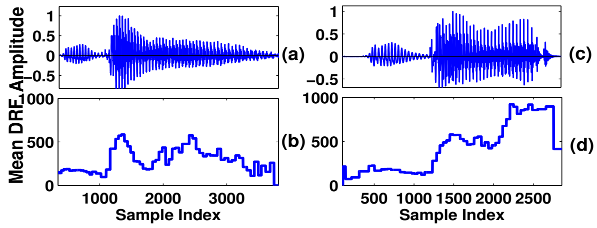

Figure 3: *Illustration of increased DRF in glottal stop region. (a)-(b), and (c)-(d) speech signal and mean of DRF between two adjacent GCIs for non-glottal stop and glottal stop contexts respectively.*

Table 1: *Mean values of the three features (SoE, H1-H2, and DRF) for both Mizo and Assam Sora*

	SoE		**H1H2**		**mean DRF**	
	NGS	**GS**	**NGS**	**GS**	**NGS**	**GS**
Mizo	0.59	0.88	8.87	-1.58	204.2	480.07
Assam Sora	0.75	0.72	5.54	1.88	244.5	356.61

4. Results and discussion

Both vocal tract and source features are used for acoustic characterization of the word-final glottal stops in Mizo and Assam Sora. In this analysis, vocal tract and source features of the final portions in a syllable, with and without a glottal stop coda are compared for both the languages. All features are averaged over the glottal stop region and in the final 40 msec of the vowel for the glottal stop and non-glottal stop contexts respectively. The terms, "*NGS*", and "*GS*" in all figures and tables denote non-glottal stop and glottal stop contexts, respectively. Figure 4(a) and Figure 4(b) show the distributions of SoE and H1-H2 values for Mizo. Significant difference in NGS and GS contexts are observed for both the features. The SoE value is lower in the non-glottal stop context and higher in glottal stop regions. In non-glottal stop contexts, the cessation of voicing is not as rapid as in the glottal stop context, where glottis is rapidly abducted with increased force. The abruptness of the glottal closure results in higher SoE values, in case of glottal stops. The vocal folds vibrate irregularly and they become considerably stiff during the production of word-final glottal stops. Such vocal

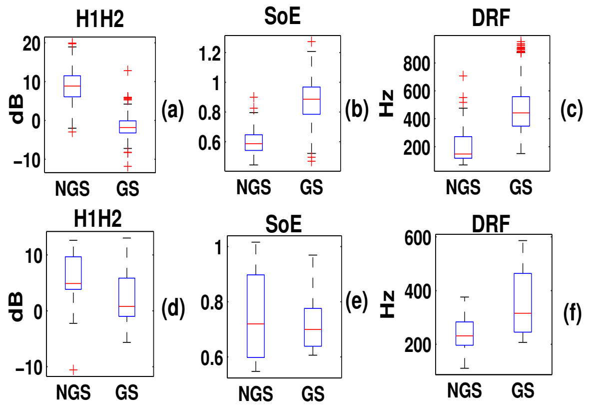

Figure 4: *Distributions of all features for Mizo (a,b,c) and Assam Sora (d,e,f).*

fold dynamics result in the increase of harmonics in the glottal source spectrum. In case of Assam Sora, smaller differences are observed for both the features (Figure 4(d) and Figure 4(e)). The SoE values in the glottal stop context are overlapping with the non-glottal stop context. Some SoE values in the non-glottal stop context are higher than glottal stop regions. On further investigation, it was observed that creakiness in the final vowels of the disyllabic words in Assam Sora results in an increase of the SoE values and decrease in the H1-H2 values. The spectrograms of the Assam Sora tokens were visually inspected and tokens were auditorily examined, before confirming the existence of creakiness. Further, the glottal stops in the word-final position are not clearly articulated in some cases.

Figure 4(c) and Figure 4(f) show the distributions of the mean DRF in Mizo and Assam Sora, respectively. In Mizo, for glottal stop regions, the DRF value is higher than for the non-glottal stop regions. However, in case of Assam Sora, DRF values may be less salient, although the median value is higher for the glottal stop region. A statistical analysis using mean of the features for both the languages are carried out and shown in Table 1. The mean values of all the three features for glottal stop regions are considerably different from the non-glottal stop contexts in Mizo. In Assam Sora, except for the SoE feature, salience is observed in H1-H2 and DRF features. As Mizo shows glottal stop and non-glottal stop distribution convincingly using the three features, a 3D plot for the three features is shown in Figure 5. As seen in the figure, glottal stop and non-glottal stop areas are well distinguished on a 3D plane.

Another experiment carried out in this study observes the temporal variations of DRF in both glottal and non-glottal contexts in the two languages. For this analysis, final glottal stops and their preceding vowel regions are considered as a single unit and divided into three equidistant parts labeled as initial, medial, and final. DRF contour is averaged for each portion and three averaged DRFs are derived for all tokens in the database and averaged. In the non-glottal stop context, only the final vowel of the word is considered and DRF values are extracted over the course of only that vowel. The measurement of the temporal variations in DRF reported in this work is based on the analysis of the high front vowel /i/ in CV and CVʔ context in Mizo and Assam Sora. Figure 6(a) shows the increasing trend of averaged DRF from preceding vowel to the following glottal stop. The averaged DRF is low in the initial part and highest in the final part in the glottal stop context for Mizo as well as for Assam Sora (Figure 6(b)). The final part contains the glottal

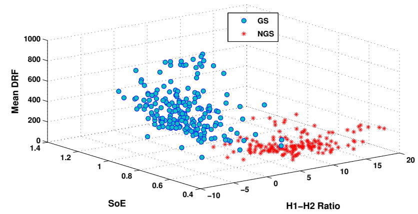

Figure 5: *3D distribution of three features (DRF, SoE, and H1-H2 ratio) for GS and NGS contexts in Mizo*

stop region where the average DRF becomes higher. However, in the non-glottal stop context, a decreasing trend is observed in Mizo. In case of Assam Sora, DRF is increasing from initial to final regions in the non-glottal stop context, due to the existence of a creaky phonation in most of the final vowels in Assam Sora. This increases the glottal closure duration, resulting in an increase in the average DRF in the final region.

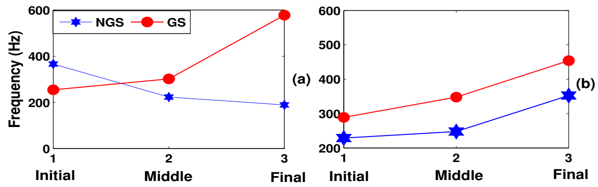

Figure 6: *Changes in DRF over the course of vowel (non-glottal stop context, NGS) and combine vowel and glottal stop (glottal stop context, GS), (a) Mizo, and (b) Assam Sora. "Final" mostly represents the glottal stop region in GS context*

5. Summary and future direction

In this study, we propose a set of acoustic features to characterize word-final glottal stops in two languages, namely, Mizo and Assam Sora. The changes in the vocal tract behavior due to deviations in the glottal source dynamics during the production of a glottal stop, is studied using DRF of the vocal tract. Increased DRF values are noticed during the closed and open phases of the glottal cycle in the glottal stop region. Moreover, the duration, in which subglottal coupling occurs, is found to be reduced in the glottal stop region. The mean DRF and excitation source features demonstrate higher deviation in the glottal stop regions when compared to the vowel final regions of non-glottal stop contexts in Mizo. Decreased salience is observed in case of Assam Sora due to a tendency among its speakers to produce creakiness in the final vowels.

Further exploration is required to study the interaction between Mizo tones and glottal stops, since tonal effect in the production of glottal stop is not considered in the current work. For automatic detection of glottal stops in the word-final position, further analysis is required to compare the characteristics of the non-glottalic stop consonants in the coda position.

6. Acknowledgment

This work is part of the ongoing project titled "Acoustic and Tonal Features based Analysis of Mizo" funded by the Department of Electronics & Information Technology (DeitY), Ministry of Communication & Information Technology (MCIT), Govt. of India.

7. References

[1] J. Hillenbrand and R. Houde, "Role of f0 and amplitude in the perception of intervocalic glottal stops." *J Speech Hear Res.*, vol. 39, no. 6, pp. 1182–1190, December 1996.

[2] M. Garellek, "Production and perception of glottal stops," Ph.D. dissertation, UCLA: Linguistics 0510, 2013. [Online]. Available: Retrievedfrom:http://escholarship.org/uc/item/7zk830cm

[3] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages*. BLACKWELL, 1996.

[4] L. Fanai, "Some aspects of the lexical phonology of mizo and english: An autosegmental approach," Ph.D. dissertation, CIEFL, Hyderabad, India, 1992.

[5] E. J. A. Henderson, "Notes on the syllable structure of lushai," *Bulletin of the School of Oriental and African Studies, University of London*, vol. Vol. 12, No. 3/4, pp. pp. 713–725, 1948.

[6] S. Kalita, L. Horo, P. Sarmah, S. R. M. Prasanna, and S. Dandapat, "Analysis of glottal stop in assam sora language," in *INTERSPEECH*, 2016.

[7] M. Garellek and S. Seyfarth, "Acoustic differences between english /t/ glottalization and phrasal creak," in *Proceedings of Interspeech 2016*, 2016, pp. 1054–1058.

[8] S. Seyfarth and M. Garellek, "Coda glottalization in american english," in *in Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, 2015.

[9] S. Iva, "Glottal stop in voro south estonian," *Linguistica Uralica*, pp. 123–133, 2005.

[10] B. Yegnanarayanaa, S. Rajendran, S. W. Hussien, and N. Dhananjaya, "Analysis of glottal stops in speech signals," in *Proc. INTERSPEECH*, Brisbane, Australia, September 2008, pp. 1481–1484.

[11] S. Kalita, S. Prasanna, and S. Dandapat, "Analysis of glottal stops using pitch synchronous integrated linear prediction residual," in *In proceedings of National Communication Conference (NCC 2016)*, March 2016.

[12] H. Seid, B. Yegnanarayanaa, and S. Rajendran, "Spotting glottal stop in amharic in continuous speech," *Journal of computer speech and language*, vol. 26, pp. 293–305, 2012.

[13] J. Pierrehumbert and D. Talkin, *Lenition of /h/ and glottal stop*. In: Docherty, G. Ladd, D. R. (Eds.), Papers inLaboratory Phonology. II. Gesture, Segment, Prosody. Cambridge University Press, Cambridge, UK, 1992.

[14] R. S. Prasad and B. Yegnanarayana, "Determination of glottal open regions by exploiting changes in the vocal tract system characteristics," *J. Acoust. Soc. Am.*, vol. 140, no. 1, pp. 666–677, July 2016.

[15] B. Yegnanarayana and D. N. Gowda, "Spectro-temporal analysis of speech signals using zero-time windowing and group delay function," *Speech Communication*, vol. 55, no. 6, pp. 782–795, 2013.

[16] K. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1602–1613, 2008.

[17] G. Seshadri and B. Yegnanarayana, "Perceived loudness of speech based on the characteristics of glottal excitation source," *J. Acoust. Soc. Am.*, vol. 126, no. 4, pp. 2061–2071, October 2009.