INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Analysis of Glottal Stop in Assam Sora Language

Sishir Kalita[1], *Luke Horo*[2], *Priyankoo Sarmah*[2], *S.R.M. Prasanna*[1], *S. Dandapat*[1]

[1]Department of Electronics and Electrical Engineering
[2]Department of Humanities and Social Sciences
Indian Institute of Technology Guwahati, Guwahati-781039, India
(sishir, luke, priyankoo, prasanna, samaren)@iitg.ernet.in

Abstract

The objective of this work is to characterize the intervocalic glottal stops in Assam Sora. Assam Sora is a low resource language of the South Munda language family. Glottal stops are produced with gestures in the deep laryngeal level; hence, the estimated excitation source signal is used in this study to characterize the source dynamics during the production of Assam Sora glottal stops. From that, temporal domain voice source features, Quasi-Open Quotient (QOQ) and Normalized Amplitude Quotient (NAQ) are extracted along with spectral features such as H1-H2 ratio and Harmonic Richness Factor (HRF). One excitation source feature is extracted from the zero frequency filtered version of the speech signal to characterize the variations within the glottal cycles in glottal stop region. A recently proposed wavelet based voice source feature, Maxima Dispersion Quotient (MDQ) is also used to characterize the abrupt glottal closure during glottal stop production. From the analysis, it is observed that the features are salient enough to uniquely characterize glottal stops from the adjacent vowel sounds and may also be used in continuous speech. A Mann-Whitney U test confirmed the statistical significance of the differences between glottal stops and their adjacent vowels.

Index Terms: Assam Sora Language, glottal stop, zero frequency filter, maxima dispersion quotient.

1. Introduction

Speech sounds are produced with gestures in the sub or supralaryngeal level. However, some sound units are articulated only in the larynx without any effective gestures in the vocal tract. A glottal stop, defined as a stop made by the glottis, is an example of such sound that is produced by firmly adducting the vocal folds [1]. In the glottal continuum model [2], the glottal stop is considered to be an extreme form of glottal closure and is placed at the right edge of the continuum, while voiceless sounds are placed at the extreme left edge. However, it is suggested that a complete glottal stop is rare in continuous speech [3].

Apart from producing glottal stops as a phonological unit in a language, they can be produced as compensatory articulations. For example, Cleft Lip and Palate (CLP) patients produce glottal stops as compensatory articulations [4], while in English, glottal stop occurs as an allophone of the stop consonant /t/. At the same time, many Austro-Asiatic languages of the Mon-Khmer subfamily such as Khmer, Chong, Kammu, Car, Khasi, Pnar, Katu, Dannu, Mon, Bunong, Sedang and Kui as well as of the Munda subfamily such as Santali, Mundari, Kera?, Ho, Korku, Juang, Kharia, Sora, Gorum, Remo, Gutob and Gta? [5] [6] include glottal stops in their phoneme inventories. However, analysis and characterization of glottal stops with the help of a spectrogram is difficult [7] [8]. As there is no movement of the supralaryngeal articulator, information regarding articulation in the larynx cannot be obtained. Hence, a significant voice (excitation) source analysis is needed to characterize this sound unit.

Production of glottal stops has a variety of realizations ranging from a complete stop to a laryngealized realization. Similarly, acoustic characteristics of a glottal stop differs significantly depending on the context in which it occurs [9]. For instance, while it is suggested that intervocalically a dip in the pitch and amplitude contour are reliable cues for perceiving a glottal stop [10], it is argued that irregularity and aperiodicity of estimated source signals may also serve as dependable cues of identifying a glottal stop in the same region [7] [8] [11]. Moreover, in the production of a glottal stop, as the larynx primarily has an effective gesture and vocal fold vibration is significantly deviated from the adjacent voiced region, analysis of glottal stop may also be conducted using aerodynamic parameters, EGG signals and estimated voice source from speech signals. However, it is preferable to analyze glottal stops directly from the speech signal so that estimated voice source may provide a better way of characterizing the acoustic qualities of a glottal stop.

A few attempts have been made to automatically characterize a glottal stop using speech signal processing. These studies have mostly used excitation source information to characterize a glottal stop. In one such study, the irregularity during a glottal stop region using a normalized cross correlation between two adjacent glottal cycles is quantified in a linear prediction residual of the speech signal [8]. Also, in order to detect the glottal stop in continuous speech of Amharic, normalized jitter and logarithm peak normalized excitation strength (LPNES) at each glottal closure instant (GCI) is computed [7]. Additionally, pitch synchronous integrated linear prediction residual is also used as voice source representation to characterize the glottal stop in intervocalic context [11]. This has helped in capturing the variation in the abruptness of glottal pulses using the ratio between the strength of excitation (SoE) at two consecutive epoch locations and temporal energy distribution using waveform peak factor (WPF). The asymmetric behavior of each glottal cycle is extracted using higher order statistical (HOS) measures.

The current study proposes a characterization method for intervocalic glottal stops in a South Munda language called Assam Sora, spoken by approximately 5000 people in Assam of North East India. Assam Sora has emerged due to the migration of Sora speakers from Orissa to Assam in the 19^{th} century. While the presence of a glottal stop in Sora has been reported [12], its presence in Assam Sora is also observed [13].

http://dx.doi.org/10.21437/Interspeech.2016-877

Glottalized vowels in Sora are reported by [12] and [14]. However, a discriminatory feature of glottalized vowel in Sora or in Assam Sora is unknown. Additionally, Anderson and Harrison suggest that a glottal stop in Sora may also function phonologically to add a mora to monosyllabic nouns to complete the minimal word template. The phonetic realization of a glottal stop in Sora or in Assam Sora is unknown. However, close observation of Assam Sora glottal stops, by trained phoneticians, have revealed that variability in the production of glottal stops is significant in the language (see Section 2). Hence, this study considers Assam Sora as a suitable language for classifying glottal stops through signal processing methods.

Due to the lack of a large dataset to build an acoustic model for glottal stops, this study is limited to the analysis of glottal stop characteristics using different voice source features only. The iterative adaptive inverse filtering based glottal flow waveform and traditional linear prediction residual are used as the representation of the estimated voice source for this analysis. The abrupt adduction of glottis in a glottal stop region and its resultant phenomena of increased harmonics in voice spectrum is captured using temporal and spectral domain source features. Additionally, to capture the variations within the glottal cycles of zero frequency filtered version of the signal, cosine distance between each half of the glottal cycle is computed. Since the glottal closing is more abrupt in glottal stops than in vowels, it is expected that the maxima following the wavelet decomposition of LP residual in glottal stops will be less dispersed than the vowels. This information is captured by the recently proposed voice source feature maxima dispersion quotient (MDQ). The irregularity of adjacent glottal cycles in voice source signal is captured by computing the normalized cross-correlation coefficient between them.

The paper is organized as follows. In Section 2 characteristics of the glottal stop in Assam Sora is discussed. Different source features used in this work are elaborately discussed in Section 3. In Section 4, a detailed discussion of dataset preparation and experimental results are provided. Finally, in Section 5, the paper is concluded by summarizing the present work.

2. Glottal stops in Assam Sora

The data in this study reveals that glottal stops in Assam Sora have three different realizations. 79% of the samples in the dataset show that glottal stops in Assam Sora are partial stops. A partial glottal stop occurs due to coarticulatory laryngealized voicing [9], and the most common characteristic caused by this is the distribution of features of the preceding vowel of the glottal stop into the following vowel. While 13% of the samples in the dataset contain creaky phonation in the glottal stop region, only 8% of the samples show a complete glottal stop. However, samples containing complete glottal stop are not taken due to the limited scope of this study.

It was mentioned earlier that a glottal stop in an intervocalic region can be perceived by a dip in the pitch contour and a dip in the amplitude contour [10]. However, the data in this study reveals that intervocalic glottal stops can have two different pitch contours. It is found that the pitch contour during a glottal stop production in an intervocalic region can either be a dip or a peak. Figure (1) shows the two different pitch contours in the speech sample produced by two male speakers examined in this study. The pitch contour is estimated by SE-VQ ([15]) algorithm and is cross examined with the widely used speech analysis tool, Praat [16].

On the other hand, the amplitude dip during the glottal stop production in intervocalic region is invariably present in data taken in this study. Only a few samples in the dataset show a flat amplitude contour from the preceding vowel to the following vowel. However, apart from these commonly fundamental frequency and amplitude analysis, better characterization method is required to accurately describe a glottal stop in an intervocalic context of Assam Sora. Hence, several voice related features are explored in the following sections that are expected to provide more insights to the source dynamics in the glottal stop region.

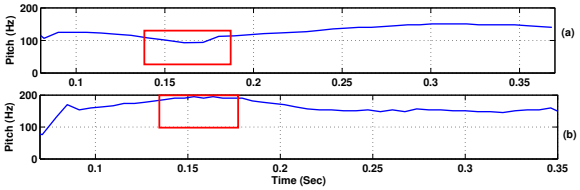

Figure 1: *Illustration of two realizations of pitch contour observed in Assam Sora (Glottal stop region is marked by a red rectangle).*

3. Feature extraction

The features used in this analysis characterizes the voice source signal to extract information regarding the intervocalic glottal stop in Assam Sora. All features are computed GCI synchronously from the representation of the voice source signal. Iterative adaptive inverse filtering (IAIF) based method is used to estimate the source signal in this work [17]. Along with this, linear prediction residual is also used to compute a wavelet based voice source feature. Since the voice behavior shifts to the non-modal category during the glottal stop region, a robust GCI detection method to handle this situation is needed. A recent GCI detection method, SEDREAMS algorithm, modified to better handle voice qualities, (SE-VQ) is used for this purpose [15]. This method shows better performance in non-modal phonation. Different source characteristics are explored using the following features:

Quasi-Open Quotient (QOQ): It is a time domain feature describing the relative open time of the glottis [18]. This parameter is preferred over the traditional open quotient (OQ) because it does not use information regarding the location of significant excitations. It is defined as the timespan during which the glottal flow is above 50 % of the difference between the maximum and minimum flow, normalized to the pitch period.

Normalized Amplitude Quotient (NAQ): This feature characterizes the glottal closing phase and it is defined as the ratio between the maximum of the glottal flow waveform and the minimum of its derivative. It is then normalized with respect to the pitch period. Since the closing phase constitute the main excitation of the vocal tract, this feature reflects phonation and vocal intensity changes robustly [19].

H1-H2 ratio: This feature describes the spectral slope of voice source spectrum. It was found that more the voice becomes tense or loud, the H1-H2 ratio decreases [20].

Harmonic Richness Factor (HRF): The harmonic content in the voice source spectrum is changed with respect to the variation in the vocal fold vibration. This parameter is defined to quantify the amount of harmonics present in the amplitude spectrum of voice source [21]. The HRF is defined in Equation (1).

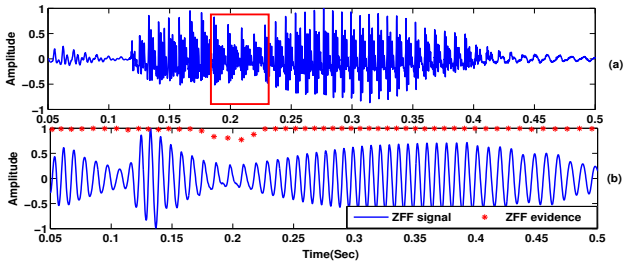

Figure 2: *Illustration of ZFF evidence in* ***/aʔa/****. context. (a) Speech signal, (b) ZFF evidence. (Glottal stop region is marked by red rectangle)*

$$HRF = \frac{\sum_{i \geq 2} H_i}{H_1} \tag{1}$$

Maxima Dispersion quotient (MDQ): This recently proposed source feature has significant link to the sharpness of the glottal closure. To compute MDQ, the LP residual of the signal is processed by a dyadic wavelet transform with seven scaled version of the wavelet function. Then, from each glottal closure instant (GCI), the locations of the maxima are searched within a predefined search interval. The distances from the GCI to each maxima is measured and then mean of these distance is computed. At last, the MDQ feature is computed by normalizing this mean value by pitch period. For breathy voice the maxima were observed to be more dispersed than the tense voices. In our study it is used to observe the sharpness of the glottal closure for the CLP speech [22].

To characterize the period to period irregularity in the derivative of the glottal flow during the glottal stop region, normalized cross-correlation coefficient (NCC) between two adjacent glottal cycles is computed. The NCC is computed as follows.

$$NCC = \frac{\langle V_1, V_2 \rangle}{\|V_1\| \|V_2\|} \tag{2}$$

where, V_1 and V_2 are two successive glottal cycles of residual signals. The number of samples in V_1 and V_2 residual segments are n_1 and n_2, respectively. If n_1 is greater than n_2, then $n_1 - n_2$ zeros are added at the end of the V_2 segment and vice versa.

A zero frequency filter (ZFF) based feature is derived to characterize the intervocalic glottal stop in Assam Sora. ZFF method is proposed to compute the GCI locations from the speech signal [23]. In this method the speech signal is passed twice through a zero frequency resonator. The transfer function of such a resonator is determined by the equation in (3).

$$H(z) = \frac{1}{1 - 2z^{-1} + z^{-2}} \tag{3}$$

This resonator de-emphasizes the characteristics of vocal tract system. The resonator output decays or grows as a polynomial function of time. The trend of the output signal is removed by subtracting the local mean over 10 ms at each sample. The resulting signal obtained after subtracting the local mean is called zero-frequency filtered signal. A window size of average pitch period is used for computing the local mean. Generally time instants of negative to positive zero crossings of the ZFF signal are regarded as GCI locations. In case of intervocalic glottal stops, the average pitch period is computed from the vowel region. Then the trend removal for the glottal stop region will not be proper due to the inherent variation of the pitch period in that region. Each half of one glottal cycle for zero frequency filtered signal in the glottal stop region is not similar as shown in Figure 2(b). To compute this variation, cosine distance between two halves within the glottal cycle is computed. The illustration of this feature is given in Figure 2(b). Apart from the above features, we have also explored normalized jitter and logarithm peak normalized excitation strength (LPNES) to characterize the glottal stop in Assam Sora. The robustness of these features to characterize the intervocalic glottal stop is reported in [7]. LPNES feature is derived from the ZFF output of speech. It is defined as the logarithm of excitation strength at each GCI location of the speech divided by the local peak of ZFF output. The slope of the ZFF output at each GCI location is defined as the excitation strength [23]. In the vowel region the LPNES value will be nearly constant, while during the glottal stop region it will have higher variation.

A recent freely available software repository known as COVAREP [24] is used in this analysis to estimate the voice source using IAIF and MDQ parameter. This repository is also used for the SE-VQ GCI detection method.

4. Experiments

4.1. Dataset

Speech samples used for analysis in this study are collected from 12 native Assam Sora speakers, both male and female speakers between 23 to 40 years of age. The dataset includes 14 Assam Sora disyllabic words that have an intervocalic glottal stop between two identical vowels. A description of the dataset is shown in Table 1. The targeted words were recorded twice in isolation and twice in a sentence frame "I saw X written". Recordings were done in Singrijhan tea estate of Sonitpur district in Assam using a Tascam linear PCM recorder connected to a Shure unidirectional head-worn microphone. The sampling frequency was 44.1 kHz, 24 bits in WAV format. All the speech samples were later manually annotated in Praat [16].

Table 1: *Description of the Assam Sora database*

#	Sora	English	#	Sora	English
1	iʔi	*louse*	8	siʔi	*hand*
2	uʔu	*hair*	9	soʔo	*rotten foul smell*
3	daʔa	*water*	10	toʔo	*mouth*
4	luʔu	*ear*	11	zaʔa	*snake*
5	moʔo	*eye*	12	zeʔe	*red*
6	muʔu	*nose*	13	ziʔi	*tooth*
7	raʔa	*elephant*	14	zoʔo	*fruit*

4.2. Results and Discussion

This section discusses the extracted features for both classes by plotting the distribution for all the speech samples in the dataset. To compute the spectral domain features, a frame size of three pitch periods around the GCI location is taken from the derivative waveform of glottal flow. The hanning window is used for this GCI centric feature extraction. Figure 3 shows the distribution of all the features computed at the GCI locations. Open quotient decreases in case of glottal stop region in comparison to the vowel region. This signifies the firm adductive behavior of glottis in the glottal stop production. The small extra peak in

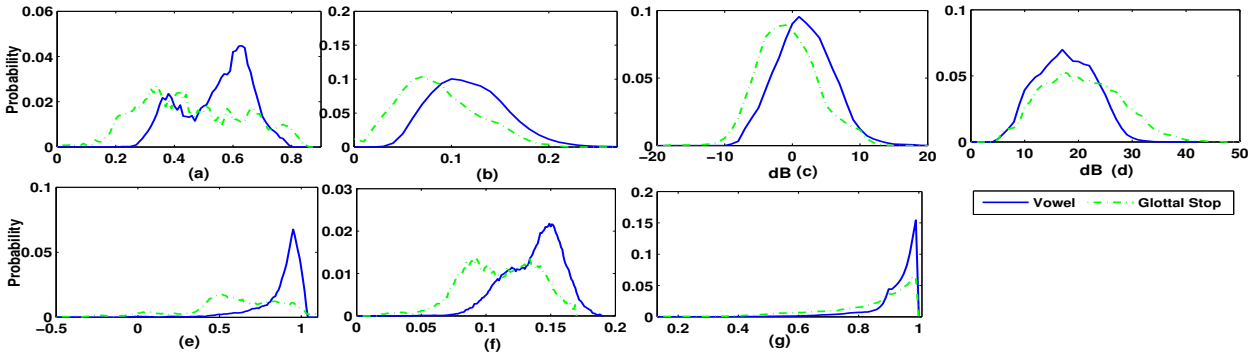

Figure 3: *Distribution of features values (a) QOQ, (b) NAQ, (c) H1-H2 ratio, (d) HRF, (e) NCC, (f) MDQ and (g) ZFF evidence for vowel regions and glottal stop regions.*

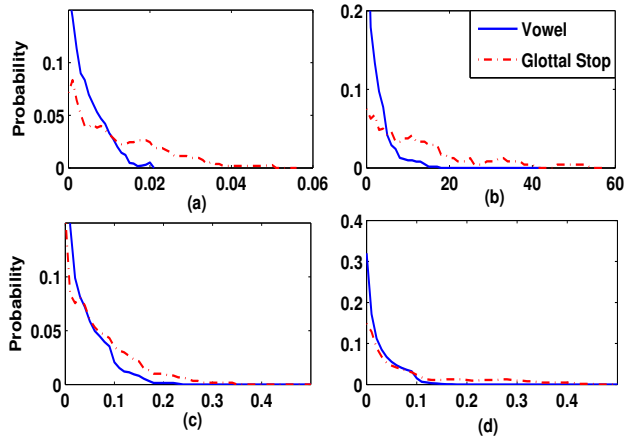

Figure 4: *Distributions of the variance values of (a) QOQ, (b) HRF, (c) NCC and (d) LPNES in glottal stop regions and in vowel regions.*

the distribution for the vowel region may be due to the feature values of vowel region adjacent to the glottal stop. The abrupt glottal closing during the glottal stop production is characterize by the NAQ parameter, as its value decreases for the glottal stop region (Figure 3 (b)). The abrupt glottal closing increases the higher order harmonics and decreases the slope of the voice spectrum (Figure 3 (c), (d)). These results are supported by the reduced value of MDQ feature. The period to period irregularities during the glottal stop region is captured by the NCC feature. Hence, the inherent irregularity of glottal cycles observed in languages analysed previously are also applicable to the glottal stops in Assam Sora. The irregularity of the glottal cycles may be due to the irregular vocal folds vibration caused during the production of a glottal stop. The feature extracted from zero frequency filtering also provides some amount of discrimination between the vowel and glottal stop region.

We computed all the features at every GCI location of the speech samples. In order to analyze the variability of the feature values of vowels as well as of the glottal stop regions, variance of the feature values for those regions is computed separately. The QOQ, HRF and NCC features show higher variability in the glottal stop region (Figure 4). The variance values show significant discrimination among the two classes and this can be used as another feature for spotting a glottal stop in continuous speech. Though the other features show discrimination between the two classes, the variability during the glottal stop region is lesser.

From these experiments, it is found that LPNES feature values in the vowel region is almost constant, whereas in the glottal stop region this feature shows higher variation. Figure 4(d) clearly shows that the LPNES value is less consistent for the glottal stop region then the vowel region. Since the excitation strength is not regular during the glottal stop due to the inappropriate vocal folds vibration more variation in the LPNES values for the glottal stop is noticed.

The normalized jitter value is computed for glottal stop and its adjacent vowels for all the samples in our dataset to observed the pitch period variation within the glottal stop region. It is found that the mean of all the jitter values computed from vowel (0.36) is significantly less than glottal stop (1.081). The variance of jitter for the glottal stop (1.43) is very high compared to the jitter of the vowel region (0.09). A Mann-Whitney U test between the glottal stop and its adjacent vowel regions for each features was performed and a significant difference (< 0.005) is observed. Hence, the intervocalic glottal stops in Assam Sora can be characterized robustly using the QOQ, NAQ, H1-H2 ratio, HRF, NCC, MDQ, ZFF based, LPNES features as well as the jitter values.

5. Conclusion

This study proposes a method to characterize the source behavior in the production of intervocalic glottal stops in Assam Sora. In Assam Sora, glottal stops are phonemic sound units that occur intervocalically and word finally. In this study the intervocalic glottal stops are used for source characterization. Most of these intervocalic glottal stops showed pitch dip characteristics, while a few were produced with increased fundamental frequency.

Two representations of the voice source, IAIF and LP residual, are used to extract temporal and frequency domain features separately. It is found that the open phase duration in the glottal stop region is very small and additionally opening to closing phase transition is very sharp when compared to the adjacent vowel regions. Significant irregularities in the adjacent glottal cycles in the voice source signal during the glottal stop production is observed. The LPNES features show higher variability and the normalized jitter value is high for the glottal stop regions. Hence, all the voice source features provide effective characterization of the intervocalic glottal stops in Assam Sora.

6. References

[1] J. Esling, K. Fraser, and J. Harris, "Glottal stop, glottalized resonants, and pharyngeals: a reinterpretation with evidence froma laryngoscopic study of nuuchahnulth (nootka)," *Journal of Phonetics*, vol. 33, no. 4, pp. 383–410, October 2005.

[2] M. Gordon and P. Ladefoged, "Phonation types: a cross-linguistic overview," *Journal of Phonetics*, vol. 29, pp. 383–406, 2001.

[3] 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.

[4] D. W. Warren, "Conpensatory speech behaviors in individuals with cleft palate: A regualtion/control phenomenon?" *Cleft Palate Journal*, vol. 23, no. 4, pp. 251–260, October 1986.

[5] M. Jenny, T. Weber, and R. Weymuth, 2014, vol. 1, ch. The Austroasiatic languagesa typological overview, pp. 13–143.

[6] G. D. Anderson, 2014, ch. Overview of the Munda Languages, pp. 364–414.

[7] 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.

[8] 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.

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

[10] 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.

[11] S. Kalita, S. Prasanna, and S. Dandapat, "Analysis of glottal stops using pitch synchronous integrated linear prediction residual," March 2016.

[12] G. D. Anderson and K. D. Harrison, *The Munda languages*, 2008, ch. Sora, pp. 299–380.

[13] L. Horo and P. Sarmah, "Acoustic analysis of vowels in assam sora," *North East Indian Linguistics*, vol. 7, pp. 69–88, 2015.

[14] N. H. Zide, *The Munda languages*, 2008, ch. Korku, pp. 256–298.

[15] J. Kane and C. Gobl, "Evaluation of glottal closure instant detection in a range of voice qualities," *Speech Communication*, vol. 55, no. 2, pp. 295–314, 2013.

[16] P. Boersma, *Praat, a system for doing phonetics by computer." Glot international 5.9/10 (2002): 341-345.*, 5th ed., Glot international, 2002.

[17] P. Alku, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech Commun.*, vol. 11, no. 2, pp. 109–118, 1992.

[18] J. Kane, S. Scherer, L.-P. Morency, and C. Gob, "A comparative study of glottal open quotient estimation techniques," in *Proceedings of INTERSPEECH*, Lyon, France, August 2013.

[19] P. Alku, T. Bckstrm, and E. Vilkman, "Normalized amplitude quotient for parametrization of the glottal flow." *J Acoust Soc Am.*, vol. 112, no. 2, pp. 701–710, August 2002.

[20] H. M. Hanson and E. S. Chuang, "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," *J. Acoust. Soc. Am.*, vol. 106, no. 2, pp. 1064–1077, August 1999.

[21] D. Childers and C. Lee, "Voice quality factors: Analysis, synthesis and perception." *Journal of the Acoustical Society of America,*, vol. 90, no. 5, pp. 2394–2410, 1991.

[22] J. Kane and C. Gobl, "Wavelet maxima dispersion for breathy to tense voice discrimination," *IEEE Trans. Audio Speech & Language Processing*, vol. 21, no. 6, pp. 1170–1179, 2013.

[23] B. Yegnanarayana and K. Murty, "Event-based instantaneous fundamental frequency estimation from speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 614–624, 2009.

[24] G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer, "Covarep - a collaborative voice analysis repository for speech technologies," in *In Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, 2014.