Analysis of Glottal Stops in Speech Signals

B. Yegnanarayana[1], S. Rajendran[1], Hussien Seid Worku[1] and Dhananjaya N[2]

[1]International Institute of Information Technology Hyderabad, India
[2]Indian Institute of Technology Madras, India

yegna@iiit.ac.in, su.rajendan@gmail.com, hussien@research.iiit.ac.in,
dhanu@cs.iitm.ernet.in

Abstract

During production of glottal stops the glottal vibration has unequal cycles and is caused by laryngealization. While one can perceive the features of laryngealization in the speech, it is difficult to analyse the signal to detect these source features from the standard spectrum-based analysis methods. In this paper we propose methods to extract the voice source vibration characteristics, and show that in the region of glottal stop, the pitch periods will be irregular, and the crosscorrelation coefficient of the signal in successive pitch periods will be low. This analysis enable us to locate the regions of glottal stops in continuous speech, and also help us to study the characteristics of creaky voice.

Index Terms: glottal stop, laryngealization, creaky voice, glottal vibration, voice source, pitch period.

1. Introduction

Both manner and place of articulation play important role in the production of different sound units in a given language. In particular, there are languages where the voice source vibration of the manner of articulation is used in a significant way to discriminate some sound units. The three broad categories of phonation types involving voice source vibration are: "(a) modal (the normal vibration type), (b) breathy (where the vocal folds are held apart so that the glottis is not closed completely and (c) laryngealized (where the folds are held stiffly and vibration is partially inhibited" [1,2]. We focus on the analysis of the laryngealized sounds in this study. Laryngealization can cause the arytenoid and ligamental portion alternating with higher and low amplitudes that may be perceived as creaky sound. In these cases the glottal vibration approaches closed phase abruptly, exciting frequencies throughout the spectrum. The glottal vibration typically has unequal cycles. Laryngealization is used to produce glottal stops (which are denoted in this paper by '?') and glottalization.

The glottal stop patterns are different from the supralaryngeal consonants, in the sense that the features of the vowel preceding the glottal stop spread across it into the following vowel, whereas the spreading does not occur across oral stop consonant [3]. For example, the articulation of the alveolar closure associated with /t/ can be represented as formant transitions on the immediately preceding vowel, total silence during the period of closure and formant transition on the immediately following vowel [3], as shown in Fig. 1(a). On the other hand, for the glottal stop there are no formant transitions on the flanking vowels as shown in Fig. 1(b). In the case of glottal stop F_0 may fluctuate.

If the two flanking vowels are different, then the formant

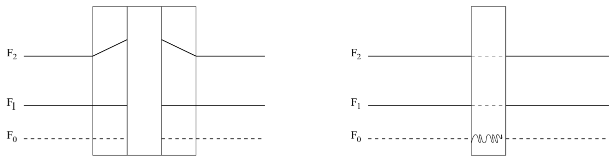

Figure 1: Illustration of formant transitions and pitch contours when the preceding vowel is same as the succeeding vowel. (a) VCV for oral stop and (b) V?V for glottal stop

transitions are caused by the place of the oral consonant(C), (see Fig. 2(a)), whereas the transition is caused by the following vowel in the $V_1?V_2$ for the glottal stop (Fig. 2(b)). The effect of the glottal stop on the speech signal waveform is primarily due to the voice source vibration. It may be possible to study the characteristics of the glottal stops from the correlation between airflow, electroglottograph and acoustic data. However, it is preferable to derive the glottal stop characteristics from the speech signal. Studies with speech signal so far are mostly limited to short-time spectrum analysis, either nonmodel-based or model-based (like linear prediction). The voice source excitation is also derived from the relative amplitudes of the harmonics and the spectral slope characteristics. The only explicit source information, separately derived and used, is the variation of the fundamental frequency F_0. Sometimes the uneven periods of the glottal cycles are also extracted whereever possible either directly or indirectly, and from the spectrogram.

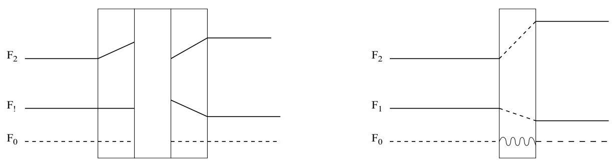

Figure 2: Illustration of formant transitions and pitch contours when the preceding vowel is different from the succeeding vowel. (a) V_1CV_2 for oral stop, and (b) $V_1?V_2$ for glottal stop

In this paper we propose analysis of glottal stops in speech signals to extract the characteristics of the voice source vibration. We show that it is indeed possible to detect and characterize the irregular voice source vibration in the $V_1?V_2$ context. In Section 2 we first discuss some acoustic characteristics specific to glottal stop sounds, and then discuss the need to develop suitable methods to extract this information. In Section 3 we present our proposed method of analysis, and illustrate how the method extracts the glottal stop characteristics in the $V_1?V_2$

context, which is more appropriate for study of glottalization that may occur in different languages [4,5]. In Section 4 we apply the proposed method for different situations of glottal and nonglottal V_1CV_2 sounds. Finally, Section 5 summarizes the work presented in this paper, and discusses some possible extensions of this study.

2. Acoustic characterization of glottal stops

Fig. 3 shows the spectrogram and the speech waveform of a glottal stop sound in the context of /isma?il/. The figure also contains the formants extracted using a method proposed in [6], the linear prediction (LP) (10th order) residual, the Hilbert envelope (HE) of the LP residual, and the pitch period contour obtained by autocorrelation of speech waveform [7]. From the spectrogram and the formant plots, we find that the continuity of formants indicate that there is no break in the shape of vocal tract as in the case of V_1CV_2 for a oral stop consonant (C). Fig. 4 shows the plots for V_1CV_2 for the stop consonant /t/, and Fig. 5 shows the plots for the diphthong V_1V_2. The formant transitions in V_1CV_2 are due to the dynamic vocal tract shape before and after the consonant. The formant transitions in Fig. 3 and Fig. 5 look similar, and are due to transition from V_1 to V_2. Therefore the glottal stop information cannot be inferred from the formant contours. Some discontinuities in harmonic structure due to aperiodicity can be seen in the spectrogram in Fig. 3, but it is difficult to relate it to the source characteristics precisely. The aperiodic behavior of the glottal vibration during the stop region can be seen to some extent in the HE of the LP residual.

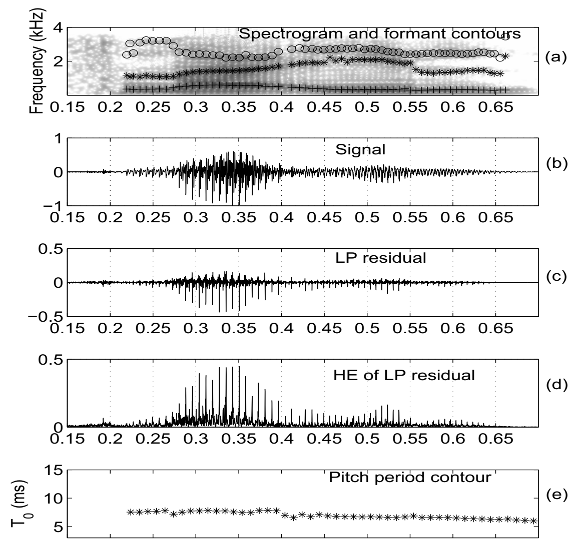

Figure 3: Analysis of the signal for the word /isma?il/. (a) Spectrogram and formant contours, (b) speech signal, (c) LP residual, (d) HE of the LP residual, and (e) contour of pitch period (T_0).

It is obvious that the voice source vibration during glottal stop is present in the source characteristics. In order to characterize the glottal stop, it is necessary to focus explicitly on the extraction of the source information from the speech signal. But most analysis methods to extract the source and system characteristics of production use mostly short-time spectrum analysis. In the next section we describe methods to extract predominantly source-related information.

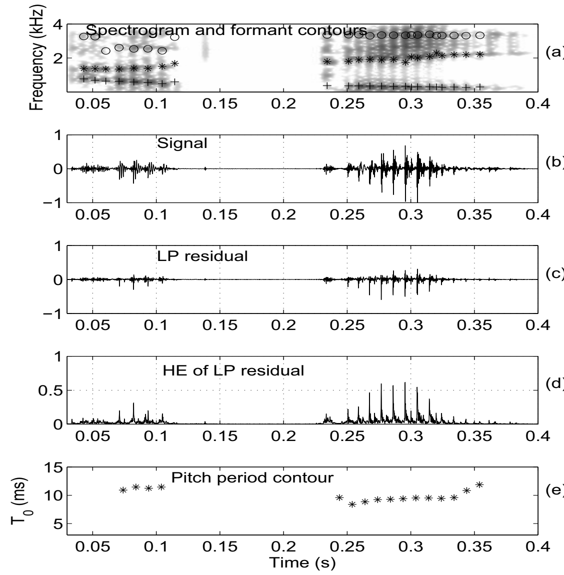

Figure 4: Analysis of the signal for the word /ati/. (a) Spectrogram and formant contours, (b) speech signal, (c) LP residual, (d) HE of the LP residual, and (e) contour of pitch period (T_0).

3. Analysis of voice source characteristics of glottal stops

As mentioned earlier, the voice source vibration is aperiodic, and somewhat irregular, in successive periods during the glottal stop region, compared to the voice source vibration in the vowel regions. In order to analyze the source characteristics, we need to reduce the effect of vocal tract response. Since the vocal tract response is not influenced significantly by the source characteristics (as seen from the formant contours in Fig. 3), we can derive a suitable (say 10^{th} order LP analysis) LP residual (see Fig. 6(b)) by using a fixed frame size of 20 ms and a frame shift of 5 ms. Note that the use of a fixed frame size and frame shift is reasonable as the vocal tract shape change is only due to the gradual change from one vowel to the other, and not due to excitation. Filtering the LP residual twice through a zero frequency resonator, and removing the trend, we get a waveform which looks like a glottal pulse sequence (Fig. 6(c))[8]. While some difference can be seen in the glottal pulse sequence in the vowel and glottal stop regions, it is difficult to derive any useful information consistently from this sequence. But the main information we can derive from this is the intervals (between two successive zero crossings) of successive aperiodic glottal cycles in the glottal stop region. While the segments of the speech signal in these regions are distinctly aperiodic, the aperiodicity is caused due to excitation source, and not due to the vocal-tract response. The difference in the source characteristics can be derived from the LP residuals in these regions. The normalized cross-correlation coefficient of successive segments can be computed, and is used as a measure of dissimilarity between successive segments (see Fig. 6(d)). The dissimilarity is less in the vowel regions compared to the dissimilarity in the glot-

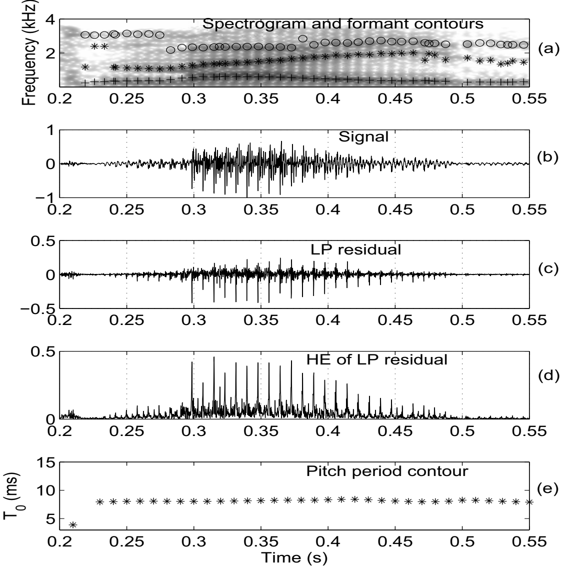

Figure 5: Analysis of the signal for the word /ismayil/. (a) Spectrogram and formant contours, (b) speech signal, (c) LP residual, (d) HE of the LP residual, and (e) contour of pitch period (T_0).

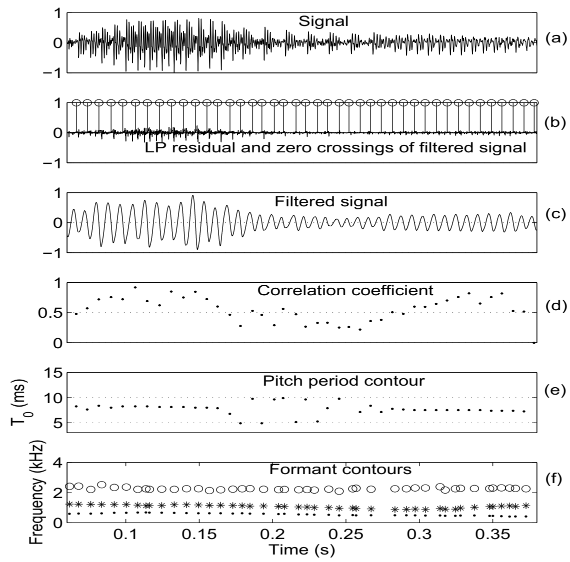

Figure 6: Glottal stop in V?V context. (a) Speech signal, (b) LP residual and zero crossings of filtered signal, (c) filtered signal derived from LP residual, (d) normalized correlation coefficient, (e) contour of pitch period (T_0), and (f) formant contours.

tal stop region. The successive glottal pulse intervals are also plotted as a pitch period ($T_0 = \frac{1}{F_0}$) plot (see Fig. 6(e)). The T_0 plot in the vowel regions is more steady compared to the T_0 plot in glottal stop region. Thus the cross correlation coefficient is low and the T_0 values are irregular in the glottal stop regions, compared to the vowel regions.

Thus the proposed method of analysis clearly highlights the irregular voice source characteristics in the glottal stop region compared to the steady region. In other words, regions, where the normalized cross-correlation coefficient between successive cycles is low, and the durations of the glottal cycles are also irregular compared to the vowel regions, correspond to the glottal stop region in the sequence $V_1?V_2$. Notice that there are no such fluctuations in these two parameters in the case of the diphthong V_1V_2.

The main steps in the computation of the voice source characteristics are summarized below [8]:

- Difference the speech signal $s[n]$

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

- Compute the LP residual for each frame of 20 ms shifted by 5 ms, i.e.,

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

where $\{a_k\}$ are the LPCs obtained by solving the normal equations

$$\sum_{k=1}^{p} a_k R[n-k] = -R[n] \qquad n = 1, 2, \ldots, p \tag{3}$$

where $R[n]$ is the autocorrelation sequence computed from the differenced signal $x[n]$

- Pass the LP residual signal twice through an ideal resonator at zero frequency. That is equivalent to integrating the LP residual four times.

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

$$y_2[n] = -\sum_{k=1}^{2} b_k y_2[n-k] + y_1[n] \tag{5}$$

where $b_1 = -2$ and $b_2 = 1$.

- Remove the trend in $y_2[n]$ by subtracting the average over 10 ms at each sample. Note that the choice of 10 ms is not very critical. Any value in the range of 0.8 to 1.5 of average pitch period is adequate. The resulting signal is

$$y[n] = y_2[n] - \frac{1}{2N+1} \sum_{m=-N}^{N} y_2[n+m], \tag{6}$$

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

- Locate the negative zero crossing instants, and mark those segments in the LP residual.
- Compute the normalized cross correlation coefficient of the LP residual signal between two successive cycles

$$r = \frac{\sum_{m=1}^{M} x_1[m] x_2[m+n]}{\sqrt{\sum_{m=1}^{M} x_1^2[m] \sum_{m=1}^{M} x_2^2[n+m]}} \tag{7}$$

where only M samples($= \min(N_1, N_2)$), and N_1 and N_2 are the number of samples in the successive cycles, n is the small relative shift in the number of samples that may be necessary to get the maximum cross-correlation coefficient.

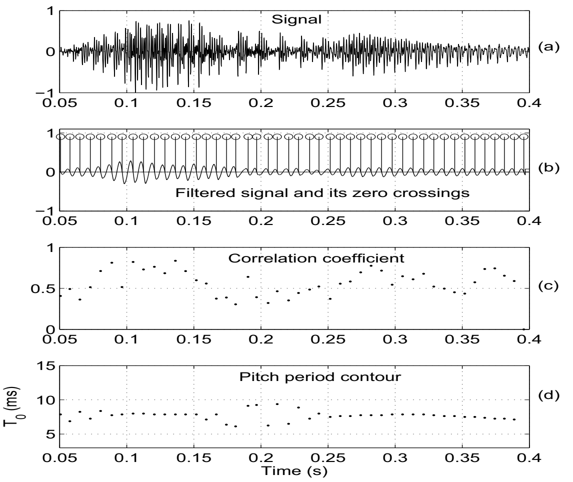

Figure 7: Glottal stop in the word /a?o/. (a) Speech signal, (b) filtered signal derived from LP residual, and its zero crossings, (c) normalized correlation coefficient, and (d) contour of pitch period (T_0).

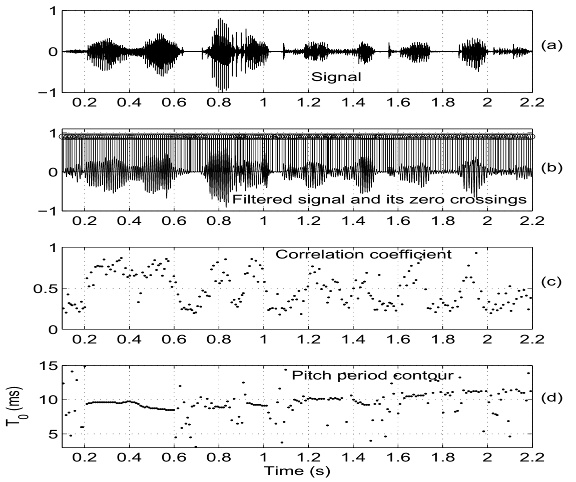

Figure 8: Glottal stop in an Amharic sentence. (a) Speech signal, (b) filtered signal derived from LP residual, and its zero crossings, (c) normalized correlation coefficient and (d) contour of pitch period (T_0).

- Compute the interval of each cycle in number of samples to show the variation of the glottal cycle duration.

4. Illustrative examples

The proposed method is applied for different examples of sound segments which contain glottal stops. Fig. 7 shows the results for the word /a?o/ containing glottal stop. The method is also applied to continuous speech signal which contains some glottal stops. The idea is to verify if the proposed method can bring out the features of glottal stops through the extracted source features. Fig. 8 shows the results for a continuous speech utterance in which the glottal stop is in the region 0.8 to 1.0 sec. Finally, Fig. 9 shows the results of analysis for a creaky voice [9]. The irregularity in the pitch period contour and the low values of the normalized crosscorrelation coefficient clearly bring out the nature of the creaky voice.

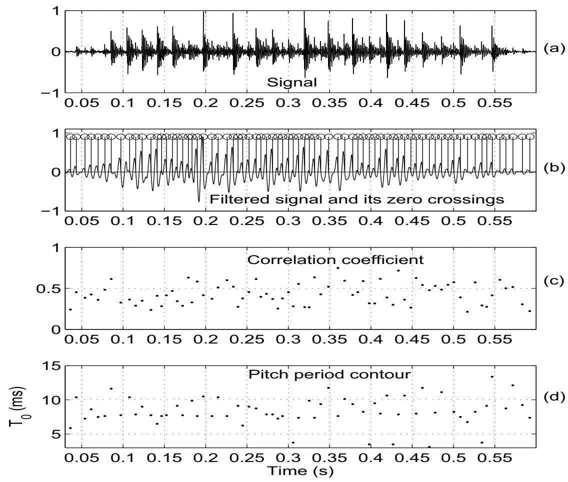

Figure 9: Creaky utterance 'had'. (a) Speech signal, (b) filtered signal derived from LP residual, and its zero crossings, (c) normalized correlation coefficient and (d) contour of pitch period (T_0).

5. Summary and conclusions

We have proposed a method to extract the characteristics of the voice source vibration with the objective of locating the glottal stop regions in $V_1?V_2$ context and in continuous speech. The characteristics of glottal stop are the irregular pitch periods and low values of the cross correlation coefficient between LP residual signals in successive glottal cycles. The proposed method is effective when applied to clean signals. It is a challenge to extract the voice source characteristics from degraded speech signal. Currently we are exploring methods to spot the glottal stop regions in continuous speech.

6. References

[1] Barbara Blankenship, "The timing of nonmodal phonation in vowels", Journal of Phonetics, vol. 30, no. 2, pp. 163–191, 2002.

[2] Matthew Gordon and Peter Ladefoged, "Phonation types: a cross-linguistic overview", Journal of Phonetics, vol. 29, no. 4, pp. 292–406, 2001.

[3] Marianne L. Borroff, "A landmark underspecification account of the patterning of glottal stop", Ph.D Thesis, Stony Brook University, New York, 2007.

[4] K. J. Kohler, "Glottal Stops and Glottalization in German," Phonetica, vol. 51, pp. 38–51, 1994.

[5] P. Hedelin and D. Huber, "Pitch period determination of aperiodic speech signals," Intl. Conf. on Acoust. Speech and Singal Processing," pp. 361–364, 1990.

[6] Anand Joseph M., Guruprasad S. and Yegnanarayana B., "Extracting formants from short segments of speech using group delay functions", Proc. Int. Conf. Spoken Language Processing (INTERSPEECH) 2006, pp. 1009-1012, Pittsburgh PA, USA, 17-21 Sept. 2006.

[7] P. Boersma, "Accurate short term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound", Proc. Institute of Phonetic Sciences, 17, Univ. Amsterdam, pp. 97-110, 1993.

[8] B. Yegnanarayana, K. Sri Rama Murty and S. Rajendran, "Analysis of stop consonants in Indian languages using excitation source information in speech signal", accepted for publication in Proceedings of Speech Analysis and Processing for Knowledge Discovery, to be held in Aalborg, Denmark during June 4-6, 2008.

[9] Peter Ladefoged, "http://www.linguistics.ucla.edu/faciliti/uclaplab.html".