Speech Communication 116 (2020) 30–43

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

Detection of glottal closure instant and glottal open region from speech signals using spectral flatness measure

Sudarsana Reddy Kadiri[a,*], RaviShankar Prasad[b], B. Yegnanarayana[c]

[a] Department of Signal Processing and Acoustics, Aalto University, Finland
[b] Speech and Audio Processing Group, IDIAP, Martigny, Switzerland
[c] Speech Processing Laboratory, IIIT-Hyderabad, India

ARTICLE INFO

Keywords:
Speech analysis
Excitation source
Glottal closure instants
Glottal open region
Single frequency filtering
Zero time windowing

ABSTRACT

This paper proposes an approach using spectral flatness measure to detect the glottal closure instant (GCI) and the glottal open region (GOR) within each glottal cycle in voiced speech. The spectral flatness measure is derived from the instantaneous spectra obtained in the analysis of speech using single frequency filtering (SFF) and zero time windowing (ZTW) methods. The Hilbert envelope of the numerator of group delay (HNGD) spectrum at each instant of time is obtained using the ZTW method. The HNGD spectrum highlights the important (like resonances) spectral characteristics of the vocal tract system at each instant of time. The dynamic characteristics of the vocal tract system can be tracked by the spectral flatness feature of the HNGD spectrum, thus bringing out the characteristics of the vocal tract system when the subglottal region is coupled with the supraglottal region during the open phase of the glottal cycle. The SFF spectra at each instant change significantly at the location of the GCI. The GCIs can be detected using the changes in the spectral flatness information derived from the SFF spectra. The proposed methods of detection of GCI and GOR is compared with several existing methods.

1. Introduction

During speech production, the major type of excitation is due to glottal vibration at the larynx. The airflow from the lungs is intercepted by the normally closed vocal folds, which are held together by the tension on the vocal folds. For sufficiently high pressure from the lungs, the tension is insufficient to hold the vocal folds together, and hence they are forced to open, releasing the air through the supraglottal vocal tract system. Just after the pressure is released, the vocal folds tend to close abruptly, thereby causing an impulse-like excitation of the vocal tract system. The opening and closing of the vocal folds take place in a quasi-periodic manner. The nature and frequency of vibrations depend on several factors, such as the mass and tension on the membranes of the vocal folds, besides the pressure difference on either side of the vocal folds. The objective of this paper is to study the effect of coupling during glottal vibration on the response of the vocal tract system.

Study of features of the glottal vibration by high speed photography may help to identify the contribution of different parts of the glottis during the vibration of the vocal folds (Mehta et al., 2011; Yan et al., 2006; Lohscheller et al., 2008; Larsson et al., 2000). But it is not possible to collect such data during normal speech production, where the vocal tract system is also time varying. Moreover, some of the visible features of the glottal vibration may not contribute significantly for production or perception of speech sounds. Measurements such as electroglottograph (EGG) will bring out only some specific characteristics, such as impedance across the folds during opening and closing phases in each glottal cycle (Abberton et al., 1989; Krishnamurthy and Childers, 1986; Childers and Krishnamurthy, 1984). The high impedance (or low conductance) values in the EGG signal indicate the open phase (less contact area between the vocal folds), and the low impedance (or high conductance) values indicate the closed phase (more contact area between the vocal folds). The open and closed phases are preceded by their onsets, which are attributed to the glottal opening and glottal closing instants, respectively. A schematic description of the EGG signals during one glottal cycle is shown in Fig. 1. Within a glottal cycle, the EGG signal can be described by four distinct phases (Henrich et al., 2004), namely, *closing, closed, opening* and *open* phases as follows:

Closing phase ($t_1 - t_3$): In this phase, contacting of the vocal folds starts at the lower margins (t_1 to t_2), then moves to the upper margins (t_2 to t_3). In general, closing is faster than opening, and the instant of maximum slope occurs at t_2, which can be seen as a strong negative peak (impulse-like behavior) in the differenced EGG (DEGG) signal (see Fig. 1(b)).

Closed phase ($t_3 - t_4$): In this phase, the vocal folds are in full contact, hence passage of air through the glottis is very less.

* Corresponding author.
E-mail address: sudarsana.kadiri@aalto.fi (S.R. Kadiri).

https://doi.org/10.1016/j.specom.2019.11.004
Received 3 July 2018; Received in revised form 25 October 2019; Accepted 12 November 2019
Available online 18 November 2019
0167-6393/© 2019 Elsevier B.V. All rights reserved.

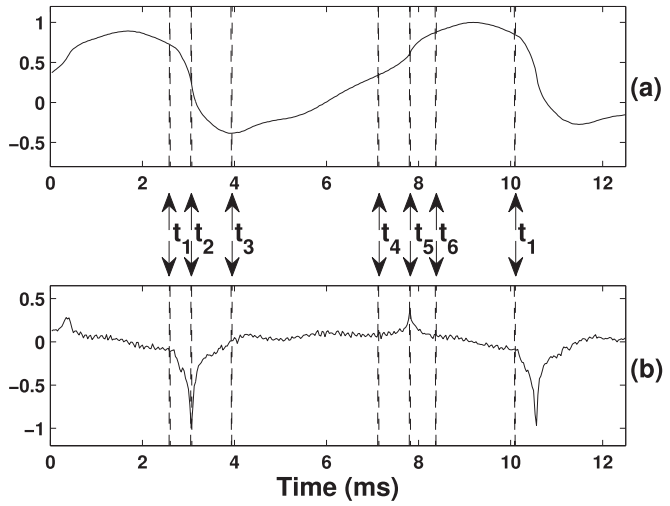

Fig. 1. Illustration of one glottal cycle as seen on an EGG signal in (a) and a DEGG signal in (b). Here, $t_1 - t_3$: closing phase, $t_3 - t_4$: closed phase, $t_4 - t_6$: opening phase, and $t_6 - t_1$: open phase.

Opening phase ($t_4 - t_6$): In this phase, the lower margins of the vocal folds slowly start to separate from each other (t_4 to t_5), followed by separation along the upper margins (t_5 to t_6). The instant of maximum slope occurs at t_5, which can be seen as the positive peak in the DEGG signal (see Fig. 1(b)).

Open phase ($t_6 - t_1$): In this phase, the vocal folds are apart. There is little variation in the impedance resulting in a relatively flat region signal in the EGG (see Fig. 1(a)).

The physiological correlates of the peaks observed in the DEGG signal, i.e., negative peak at t_2 and positive peak at t_5, are interpreted as glottal closure instant (GCI) and glottal opening instant (GOI), respectively. These instants serve as the reference/ground truth for evaluating the GCIs/GOIs derived from the speech signal. In general, the glottal opening is relatively a slower phenomenon, and hence does not exhibit any impulse-like behavior in the DEGG signal. Hence, the glottal opening cannot be interpreted as an instant property. However, sometimes a weak impulse-like behavior with opposite polarity is seen between two successive GCIs in the DEGG signal, and it is attributed to the glottal opening instant (GOI) (Henrich et al., 2004). The definitions of different phases of the glottal activity vary in the case of pathological voices (Henrich et al., 2004), where the presence of mucus strand bridge affects the EGG signal and also the glottal flow waveform derived from the speech signal. It is to be noted that the EGG signal does not give any information about the acoustic pressure variations or the glottal width (Stevens, 1977).

Studies on glottal activity features focus mainly on extracting the GCIs from the speech signal, as it is difficult to extract the GOIs due to their relatively weaker presence (Yegnanarayana and Gangashetty, 2011). Methods for detecting the GCIs rely mostly on the linear prediction (LP) residual of the speech signal, assuming that the prediction is low at the GCIs (Ananthapadmanabha and Yegnanarayana, 1979). Refinements were proposed to improve the GCI detection from the LP residual (Prathosh et al., 2013; Drugman et al., 2012b; 2014). In general, the inverse filtering in the LP analysis does not cancel out the effects of the vocal tract resonances completely in the LP residual.

Methods for detecting the GOI rely mostly on the identification of the GCI initially, and then a suitable duration is marked for the open phase (fixed value or certain ratio with respect to the period of the glottal cycle). Detection of the GOI is needed for closed phase LP analysis, and also for analysis of pathological speech due to its dependence on the knowledge of the open quotient (Lieberman, 1963; Silva et al., 2009). Detection of GOI is a challenging task, as there is no universally agreed definition of GOI (Thomas et al., 2012). Three main definitions of GOI are reported in (Thomas et al., 2012). The first one defines GOI as the instant at the end of the closed phase where an increased residual error is obtained in the LP analysis (Thomas et al., 2012; Wong et al., 1979). This definition is used for closed phase covariance LP analysis. In the second definition, the GOI is marked at the location of the maximum value of the differenced EGG (DEGG) signal, that corresponds to the maximum rate of change of the glottal conductivity, and not the air flow (Abberton et al., 1989; Henrich et al., 2004). This definition is used to assess the open quotient in pathological voices (Lieberman, 1963; Silva et al., 2009). In the third definition, the GOI is identified with the point at which the amplitude of the EGG waveform is equal to a percentage of the maximum value within a glottal cycle (Rothenberg and Mahshie, 1988). Each of these definitions is limited to specific studies of interest. Actually GCI and GOI are not point properties, but are regions of finite duration within a glottal cycle, although the closing phase is usually sharp enough to be considered as occurring at an instant (Abberton et al., 1989; Herbst et al., 2014). The opening phase of the glottis being more gradual, it may be possible to assign only a *region* for the open phase within a glottal cycle.

Methods for detection of GCI/GOI include changes in the spectra (Moulines and Di Francesco, 1990), changes in signal energy, Hilbert envelope of LP residual Ramesh et al. (2013), Frobenius norm (Ma and Willems, 1994), eigen value decomposition of Hankel matrix Jain and Pachori (2013, 2014); Jain and Pachori (2012), time-order representation based on short-time Fourier-Bessel series (Jain and Pachori, 2012), phase flatness (Degottex et al., 2009; 2010; 2011b) and group delay function (Rao et al., 2007). With the observation that the discontinuities in a signal are manifested as local maxima across multiple wavelet scales, the lines of maximum amplitude (LoMA) algorithm identifies singularities in a signal (D Alessandro and Sturmel, 2011). In (Thomas et al., 2012; Bouzid and Ellouze, 2009; Thomas and Naylor, 2009), the multiscale product of the decomposed wavelet signals has been shown to be effective for GCI/GOI detection from EGG and speech signals. The yet another GCI/GOI algorithm (YAGA) (Thomas et al., 2012) uses wavelet transform, group delay, glottal flow waveform and dynamic programming. Another method uses a mean-based signal and LP residual to detect GCIs/GOIs (Drugman et al., 2012b).

Estimation of the glottal source waveform by glottal inverse filtering (GIF) involves filtering the speech signal with the inverse of the estimate of the vocal tract transfer function (Alku, 2011; Walker and Murphy, 2007). The vocal tract system is considered stationary during the analysis frame ($20 - 30$ *ms*), which is not the case in each glottal cycle (Alku et al., 2009). To overcome this problem, iterative adaptive inverse filtering (IAIF) was proposed in (Alku, 1992). Variants of the IAIF such as weighted linear prediction (WLP), stabilized weighted linear prediction (SWLP), and quasi-closed phase weighted linear prediction (QCP-WLP) are proposed in (Airaksinen et al., 2014; Kafentzis et al., 2011). A second approach for inverse filtering uses joint source-filter optimization for capturing the vocal tract and glottal source components (Fu and Murphy, 2006; Schleusing et al., 2013). In this, models like Liljencrants-Fant (LF) model (Fant, 1995; Degottex et al., 2011a) and Rosenberg-Klatt (RK) model (Alku, 2011; Veldhuis, 1998) of the glottal source are used. A third approach is based on mixed-phase model and combination of causal (minimum phase) and anti-causal (maximum phase) components of speech. The complex cepstrum decomposition (CCD) and zeros of the z-transform (ZZT) techniques are two different methods which fall under this category (Drugman et al., 2011; 2012a). In these, the vocal tract impulse response and the glottal return phase are considered as causal signals, and the open phase of the glottal flow signal is considered as an anti-causal signal. Even though the accuracy of the GIF methods is difficult to quantify, several studies conclude that the glottal source estimates tend to become unreliable in the analysis of certain types of voice such as high pitched and expressive voices (Drugman et al., 2014; Alku et al., 2009). As of now, the

existing GIF methods are limited mostly to analysis of sustained sounds (Alku, 2011).

It is important to note that most of the methods for estimating the glottal source information from speech do not address the issue of coupling of the subglottal cavity with supraglottal cavity (Moulines and Di Francesco, 1990; Kadiri and Yegnanarayana, 2017). The coupling effects have been studied using either acoustic tube or vocal fold mechanics models (Lulich et al., 2009; Barney et al., 2007; Chi and Sonderegger, 2007). Also, these studies focussed mostly on the effect of coupling on the vowel formants or on the acoustic response of the supraglottal vocal tract. Moreover, the discrete Fourier transform (DFT) used on short segments of speech smears the effects of coupling due to averaging effect of the response of the vocal tract in the window of analysis.

It is obvious that different glottal phases within the period of a glottal cycle affect the resulting speech signal differently. The impulse-like excitation due to glottal closure produces speech as the response of the supraglottal vocal tract system. The strong excitation at the GCI is generally observed in the residual signal. Also, the spectrum due to the impulse at the GCI (if it can be computed) should appear more flat than at the other regions. When the vocal folds are open, speech is the response of the subglottal system coupled with the supraglottal system. As opening of the vocal folds is gradual in nature, it is difficult to determine its influence on the response of the vocal tract system. However, when the vocal folds are completely open, the subglottal system is coupled to the supraglottal system, and the resultant vocal tract has different dimensions in comparison with the vocal tract dimensions in the closed phase region. Thus the effect of opening on the response of the vocal tract system is likely to be different at different instants in the open region. When the opening is too small, then there may be some increase in the bandwidth of the first formant of the supraglottal vocal tract system. On the other hand, if the opening is large, the effective vocal tract length will be large due to coupling, and hence low frequency resonances appear along with increase in the bandwidth of other resonances. This may also result in flat spectrum in the response of the vocal tract system.

Response of the vocal tract system is generally examined using the spectral characteristics derived from the speech signal. Some attempts were made to exploit the changes in the spectral characteristics for determining the events in the glottal cycle, especially the GCIs from the speech signal. In (Moulines and Di Francesco, 1990), the abrupt changes in the short-time spectral characteristics are estimated using auto-regressive (AR) models. To detect the abrupt change from open to closed regions at the GCI, two statistical methods (likelihood ratio and divergence between short-term probability distribution function (PDF) and long-term PDF) were used (Moulines and Di Francesco, 1990). In these cases, the spectrum is estimated using lower order AR models to improve the temporal resolution. But lower order AR model gives smoothed spectral envelope, thus reducing the discrimination between open and closed regions. Moreover, an AR model tries to fit the spectrum of the entire windowed segment, and thereby capturing only the overall short-time spectrum characteristics. Recently in (Kadiri and Yegnanarayana, 2017), the magnitude spectral properties of the time domain impulse are used for detecting the GCIs, as the effect of the time domain impulse is spread across all the frequencies resulting in a flatter spectrum. In (D Alessandro and Sturmel, 2011), wavelet transform has been used to highlight the discontinuities in voiced speech for different time-scales.

In this paper, we intend to study the effect of glottal closure and glottal opening on the coupling of the vocal tract system within a glottal cycle in order to determine the GCI and glottal open region (GOR). We focus on GOR instead of GOI, as it is difficult to define GOI precisely. It is to be noted that the present study attempts to interpret these features in terms of changes in the vocal tract system, rather than the changes in the excitation characteristics. We attempt to determine the GCI and GOR by exploiting the changes in the spectral characteristics derived from single frequency filtering (SFF) (Aneeja and Yegnanarayana, 2015) and zero time windowing (ZTW) (Yegnanarayana and Gowda, 2013) methods, respectively. We exploit the distinctive features that the GCI is reflected prominently in the excitation component of the signal, and the GOR is reflected through the changes in vocal tract system characteristics due to coupling of the subglottal system.

The key contributions of the present study are as follows:

- A new approach for determination of GCI and GOR from speech signals is proposed based on the changes in the spectral characteristics within a glottal cycle, using two signal processing methods, namely, ZTW and SFF.
- The spectral flatness parameter derived from the SFF spectra highlights the impulse-like characteristics at the GCI.
- The spectral flatness derived from the ZTW spectra highlights the glottal opening, as there are significant changes in the effective vocal tract length in the GOR.
- The proposed method of GCI detection is compared with several existing methods in clean speech and telephone quality speech scenarios. The results of the proposed method is comparable to the existing methods in clean speech and better in telephone quality speech.
- The proposed method of GOR detection is shown to be comparable with the existing methods.

The organization of the paper is as follows: Section 2 briefly describes the ZTW and SFF methods for processing signals. Derivation of glottal source features from the ZTW and SFF is discussed in Section 3. Effect of analysis parameters in these methods is discussed in Section 4. Section 5 gives the methods for detection of GCI and GOR from speech signals. The results of the proposed methods in comparison with the existing methods are discussed in Section 6. Finally, Section 7 gives a summary of the study.

2. SFF and ZTW methods

Two recently proposed signal processing methods are used in this study to determine the GCI and GOR in each cycle of glottal activity. A brief description of these two methods is given in this section.

2.1. Single frequency filtering (SFF) method

In the SFF method, the amplitude envelope of the speech signal is obtained at any desired frequency by frequency shifting the signal and filtering the signal using a single-pole filter. The following are the steps involved in obtaining the amplitude envelope of the signal at the k^{th} desired frequency (f_k Hz) (Kadiri and Yegnanarayana, 2017; Aneeja and Yegnanarayana, 2015).

(a) The speech signal $s[n]$ is differenced to remove low frequency variations.

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

(b) The signal $x[n]$ is multiplied with a complex exponential ($e^{j\bar{\omega}_k n}$), where $\bar{\omega}_k = \pi - \omega_k = \pi - \frac{2\pi f_k}{f_s}$. The resulting frequency-shifted signal is given by

$$x_k[n] = x[n]e^{j\bar{\omega}_k n}. \tag{2}$$

The Fourier transform of the frequency-shifted signal is given by $X_k(\omega) = X(\omega - \bar{\omega}_k)$, where $X_k(\omega)$ and $X(\omega)$ are the Fourier transforms of $x_k[n]$ and $x[n]$, respectively.

(c) The signal $x_k[n]$ is passed through a single-pole filter whose transfer function is given by

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

The output of the filter $y_k[n]$ is given by

$$y_k[n] = -ry_k[n-1] + x_k[n], \tag{4}$$

where $r \approx 1$ for a root close to the unit circle in the z-plane. Here $y_k[n] = y_{kr}[n] + jy_{ki}[n]$, where $y_{kr}[n]$ and $y_{ki}[n]$ are the real and imaginary parts of $y_k[n]$, respectively.

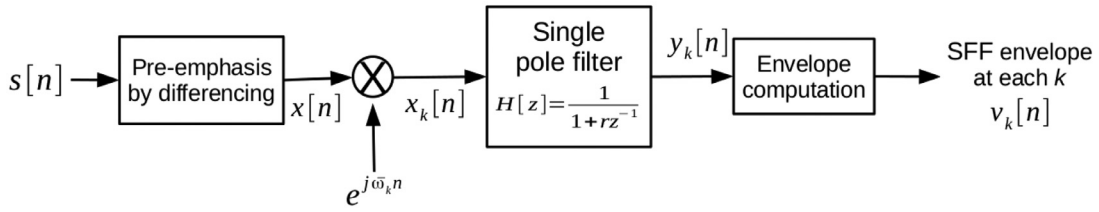

Fig. 2. Schematic block diagram of computations in the SFF method.

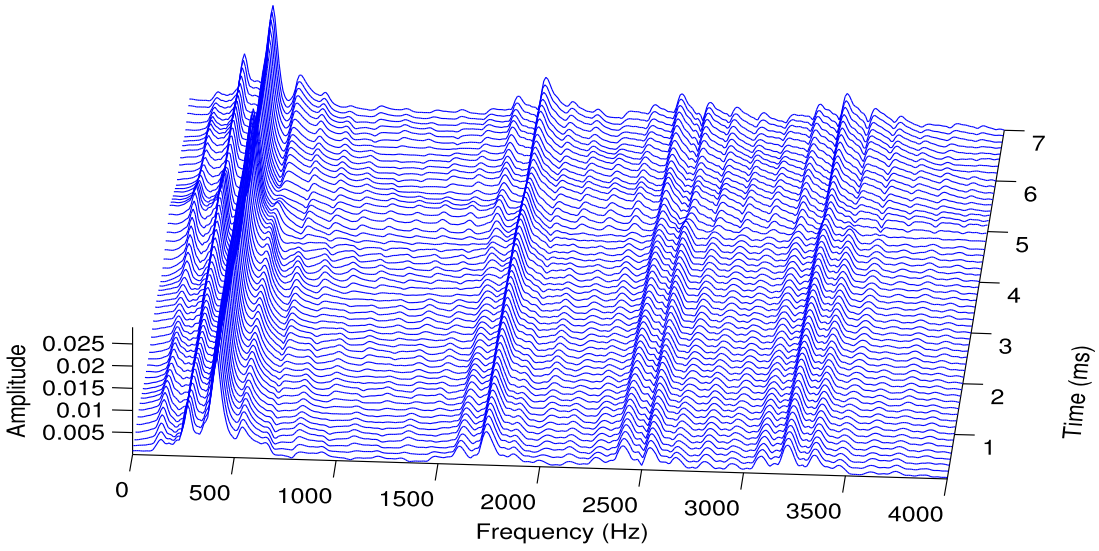

Fig. 3. Illustration of 3-D view of SFF spectra for a complete glottal cycle, starting from glottal open region. The significant change that occurs near GCI (around 5 *ms*) is not clearly visible in this spectra, due to harmonic structure of the spectra. But the change around GCI can be seen clearly in the spectral flatness contour computed from SFF spectra.

(d) The amplitude envelope $v_k[n]$ of the signal is given by

$$v_k[n] = \sqrt{y_{kr}[n]^2 + y_{ki}[n]^2}. \tag{5}$$

To ensure stability of the filter, the value of r is chosen slightly less than 1. The amplitude envelopes of the signal can be obtained for several frequencies, at interval of Δf. That is

$$f_k = k\Delta f, \qquad k = 1, 2, \ldots, K, \tag{6}$$

where $K = \frac{(f_s/2)}{\Delta f}$.

The computational steps involved in the SFF method are shown in the schematic block diagram in Fig. 2.

An illustration of the 3-D view of the SFF spectra (computed using $r = 0.995$) for a complete glottal cycle in a voiced speech segment is shown in Fig. 3. The SFF spectra show the harmonics due to periodic impulse-like excitation. The spectral changes within a glottal cycle are not evident due to the harmonic structure in each of the SFF spectra. The spectral changes within a glottal cycle can be seen from the SFF spectrogram shown in Fig. 4. Fig. 4(b) shows the SFF spectrogram of the voiced speech segment shown in Fig. 4(a), and the corresponding DEGG signal is shown in Fig. 4(c). The SFF spectrum is flatter around the instants of glottal closure, which can be highlighted using a spectral flatness measure as discussed in Section 3.

2.2. Zero time windowing (ZTW) method

In this method the instantaneous spectral characteristics of the time varying vocal tract response is captured. The speech signal is multiplied with a heavily decaying window at each instant, so that the samples at the beginning of the window are given more emphasis, and hence the name zero time windowing (ZTW). Group delay analysis of the windowed signal gives good spectral resolution. The method thus gives high temporal resolution, maintaining simultaneously good spectral resolution. The use of a heavily decaying window function is equivalent to integration in the frequency domain, analogous to the zero frequency filtering (ZFF) method in the time domain (Murty and Yegnanarayana, 2008). The hidden spectral features are highlighted by successively differencing the numerator of the group delay (NGD) function. Hilbert envelope (HE) of the differenced NGD function brings out even the weaker resonances as discussed in (Yegnanarayana and Gowda, 2013). The resulting spectrum is referred to as the HNGD spectrum. The following are the steps involved in extracting the instantaneous spectral characteristics using the ZTW method (Yegnanarayana and Gowda, 2013).

(a) The speech signal ($s[n]$) is pre-emphasized to reduce the effects of low frequency variations.
(b) Consider a speech segment of L *ms* (number of samples: $M = L * f_s/1000$) at each instant, where f_s is the sampling frequency. That is, $s[n]$ is defined for $n = 0, 1, \ldots, M-1$.
(c) Multiply the segment with a window $w_1^2[n]$, where

$$w_1[n] = 0, \qquad n = 0,$$
$$= \frac{1}{4\sin^2(\pi n/2N)}, \qquad n = 1, 2, \ldots, N-1. \tag{7}$$

Here N is the number of samples used for DFT computation, and $N > > M$. Multiplying $s[n]$ with the window $w_1^2[n]$ is approximately equivalent to four times integration in the frequency domain (Yegnanarayana and Gowda, 2013).
(d) Truncation of the signal in the time domain at the instant $n = M-1$, may result in a ripple effect in the frequency domain. The ripple effect is reduced by using another window $w_2[n]$, which is

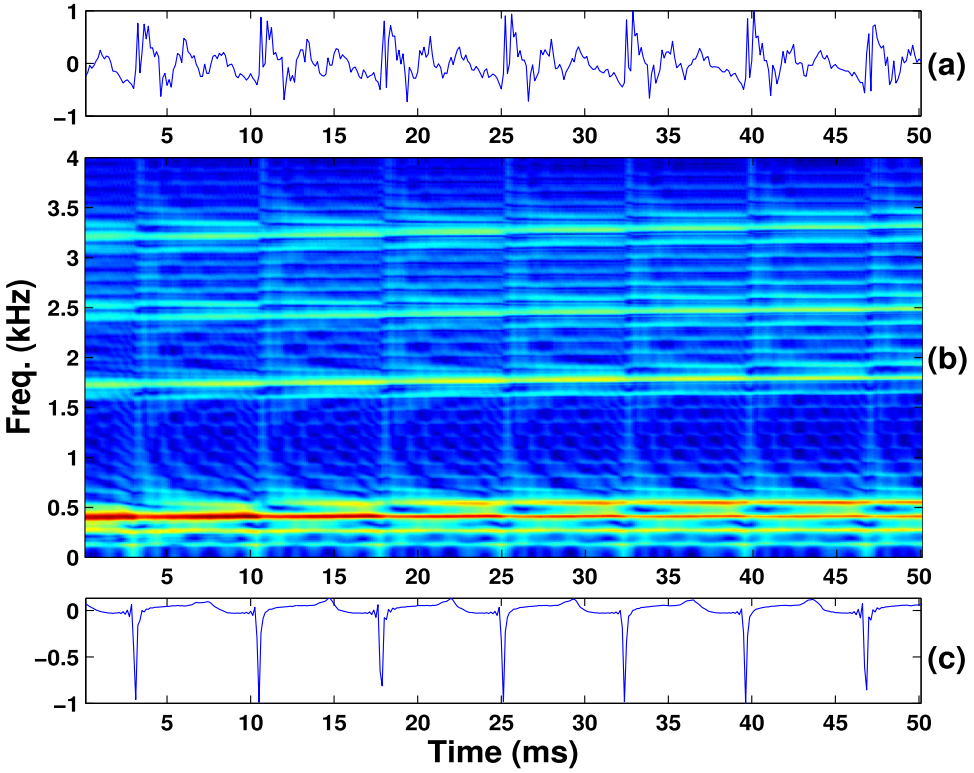

Fig. 4. An illustration of the SFF spectrogram for a voiced speech segment. (a) A segment of voiced speech. (b) SFF spectrogram. (c) DEGG signal.

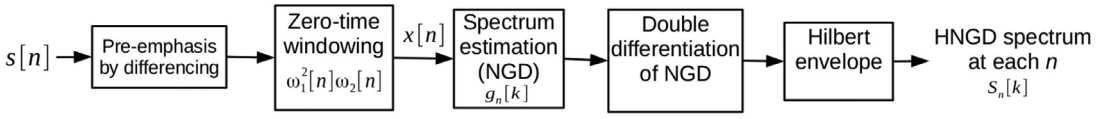

Fig. 5. Schematic block diagram of computations in the ZTW method.

square of half cosine window. That is

$$w_2[n] = 2(1 + cos(\pi\ n/M)) = \ 4\ cos^2(\pi n/2M), \quad n = 0, 1, \ldots, M-1. \tag{8}$$

(e) The NGD function ($g[k]$) of the windowed signal (i.e., $x[n] = w_1^2[n]w_2[n]s[n]$) is given by

$$g[k] = X_R[k]Y_R[k] + X_I[k]Y_I[k], \qquad k = 0, 1, 2, \ldots, N-1. \tag{9}$$

where $X_R[k]$ and $X_I[k]$ are the real and imaginary parts of the N-point DFT $X[k]$ of $x[n]$, and $Y_R[k]$ and $Y_I[k]$ are the real and imaginary parts of the N-point DFT $Y[k]$ of $y[n] = nx[n]$.

(f) The NGD function is double differenced to highlight the spectral peaks corresponding to resonances of the vocal tract. Hilbert envelope of the double differenced NGD is called the HNGD spectrum, denoted by $S_n[k]$.

The computational steps involved in the ZTW method are shown in the schematic block diagram in Fig. 5.

An illustration of the 3-dimensional (3-D) view of the HNGD spectra (computed for $L = 4$ ms) for a complete glottal cycle in a voiced speech segment is shown in Fig. 6. The spectral peaks in the HNGD spectrum correspond mostly to the resonances of vocal tract system. The initial part corresponds to the spectra in the open phase region and the later (more uniform) part corresponds to the spectra in the closed phase region. The spectral changes within a glottal cycle can be observed from the HNGD spectrogram over several glottal cycles, as illustrated in Fig. 7. Fig. 7(b) shows the HNGD spectrogram for the voiced speech segment shown in Fig. 7(a). The corresponding DEGG signal is shown in Fig. 7(c). The negative peaks in the DEGG signal correspond to the GCIs. We know that the effective length of the vocal tract system is shorter in the closed phase region and longer in the open phase region. The spectral energy is distributed more uniformly in the open phase region, compared to that in the closed phase region. A spectral flatness measure highlights this feature better as discussed in the next section.

3. Glottal activity information from SFF and HNGD Spectra

At each instant the spectrum is normalized by dividing the values with their sum across the frequency, so that the spectral sum at each instant is equal to 1. The normalized SFF spectrum $\hat{v}_k[n]$ and HNGD spectrum $\hat{S}_n[k]$ are given in Eqs. (10) and (11), respectively.

$$\hat{v}_k[n] = \frac{v_k[n]}{\sum_{k=1}^{K} v_k[n]}, \qquad k = 1, 2, \ldots, K. \tag{10}$$

$$\hat{S}_n[k] = \frac{S_n[k]}{\sum_{k=1}^{N} S_n[k]}, \qquad k = 1, 2, \ldots, N, \tag{11}$$

where K is the number of frequencies in the SFF analysis. For a sampling frequency of $f_s = 8\ kHz$, and $\Delta f = 10\ Hz$, $K = 400$. The value of $N = 2048$ is the number of DFT points used in the ZTW analysis.

3.1. Spectral flatness measure

The spectral flatness is calculated by dividing the geometric mean of the spectral values with the arithmetic mean. A high spectral flatness indicates that the spectral values are more uniformly spread out. A low spectral flatness indicates that the spectral values are more unevenly

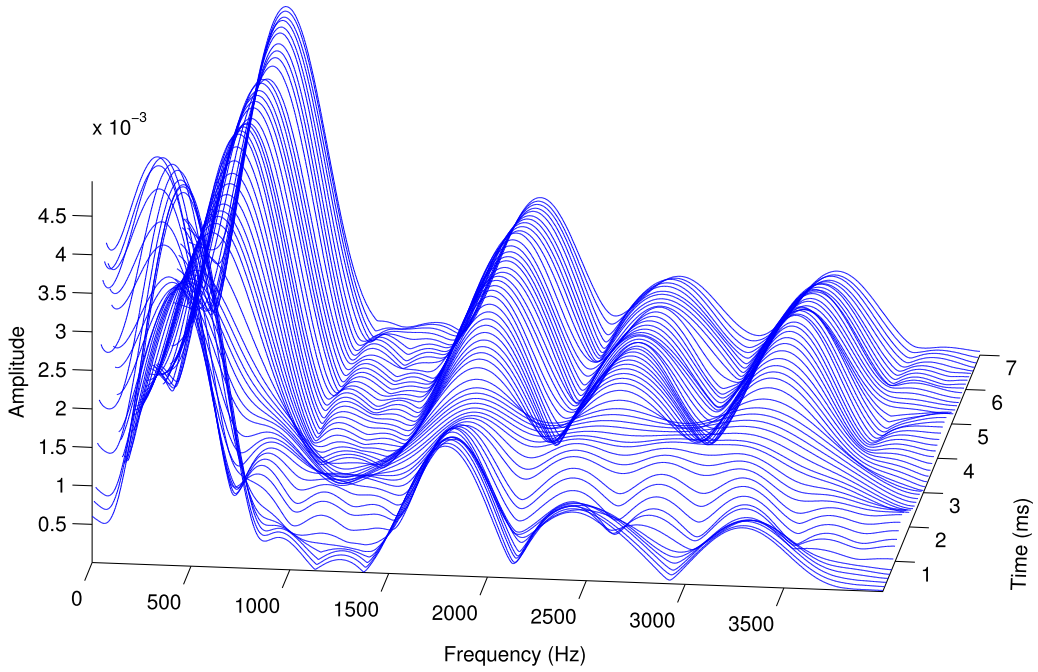

Fig. 6. Illustration of 3-D view of HNGD spectra for a complete glottal cycle, starting from glottal open region. The effect of the initial glottal open region on the HNGD spectrum can be seen clearly in comparison with the HNGD spectra in the glottal closed region.

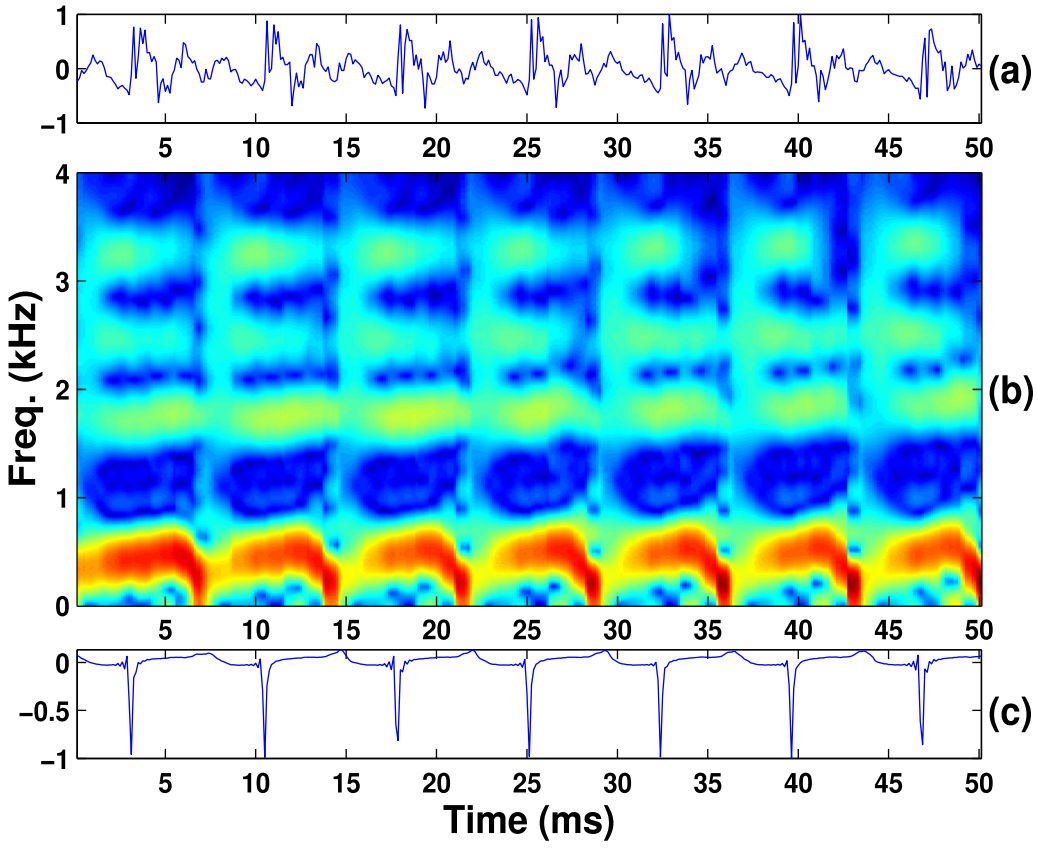

Fig. 7. An illustration of the HNGD spectrogram for a voiced speech segment. (a) A segment of voiced speech. (b) HNGD spectrogram. (c) DEGG signal.

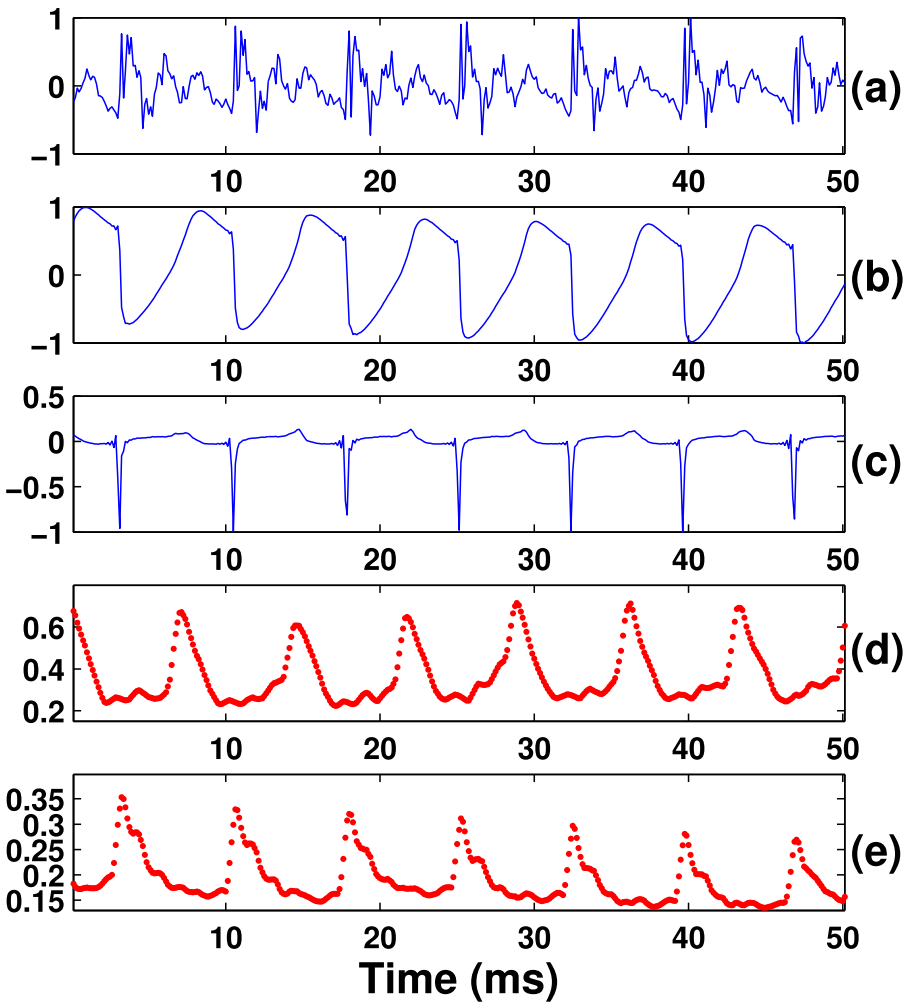

Fig. 8. An illustration of spectral flatness contours derived from ZTW and SFF methods. (a) A segment of voiced speech. (b) EGG signal. (c) DEGG signal. (d) Spectral flatness contour derived from HNGD spectrum for a window length of 4 *ms*. (e) Spectral flatness contour derived from SFF spectrum for $r = 0.995$.

distributed. The spectral flatness measure for spectra from SFF and ZTW methods are given in Eqs. (12) and (13), respectively.

$$\delta_{SFF}[n] = \frac{\sqrt[K]{\prod_{k=1}^{K} \hat{v}_k[n]}}{\frac{1}{K}\sum_{k=1}^{K} \hat{v}_k[n]}. \tag{12}$$

$$\delta_{ZTW}[n] = \frac{\sqrt[N]{\prod_{k=1}^{N} \hat{S}_n[k]}}{\frac{1}{N}\sum_{k=1}^{N} \hat{S}_n[k]}. \tag{13}$$

3.2. Glottal activity information from SFF method

The effect of the impulse-like excitation is observed in the amplitude envelope extracted at each frequency in the SFF method. While there is temporal smearing due to IIR nature of the filter response, the effect of the impulse-like excitation is preserved at the respective time instants in all the frequencies. The spectral flatness computed from the SFF spectrum provides a one dimensional representation of the excitation source features. The spectral flatness contour derived using the SFF envelopes for $r = 0.995$ is shown in Fig. 8(e). It can be seen that the SFF spectral flatness contour has higher values in the regions where the EGG values are low. The values of the SFF spectral flatness are low in the open phase region where the EGG values are high. The sudden changes in the spectral flatness occur at the GCI locations, which are well matched with the negative peaks in the DEGG signals. The abrupt transition of the SFF spectral flatness values from low to high can be attributed to the impulse-like excitation of the vocal tract system. The spectral flatness contour gradually moves to lower values at other regions due to temporal smearing of the SFF output.

The SFF spectra show the harmonics due to sequence of impulse-like excitation. Fig. 9 shows the SFF spectra for $r = 0.995$ at ten locations within a glottal cycle. Fig. 9(a) shows the segment of voiced speech, Fig. 9(b) shows the corresponding SFF spectral flatness contour. Fig. 9(c1)–(c10) show the normalized SFF spectra at time instants 1 to 10, marked in the speech signal (Fig. 9(a)) and in the spectral flatness contour (Fig. 9(b)). In Figs. 9(c1)–(c4), the spectral tilt is higher (although it is difficult to visualize due to presence of the harmonic structure), as it is located in the open phase region. Near the glottal closure (Fig. 9(c5)), the spectral tilt is lower (i.e., flatter spectrum). The flatness of the spectrum is evident, if we see carefully in the regions 1000–2000 Hz and 2000–3000 Hz in (c5) compared to (c1-c4) and (c6-c10). There is a sudden rise in the SFF spectral flatness value (see Fig. 9(b) before the instant 5) in this region. This sudden change in the spectral tilt can be attributed to the sharp closure of the glottis. The spectral tilt increases slowly in the open phase region, although it cannot be seen in the Figs. 9(c6)–(c10). The spectral flatness goes to lower values (see Fig. 9(b) at the instants 7 to 10).

3.3. Glottal open region information from ZTW method

Changes in the length of the vocal tract during opening and closed phases are reflected in the spectral characteristics derived from the HNGD spectra at each instant of time. In particular, the spectral flatness computed from the HNGD spectrum provides a one dimensional repre-

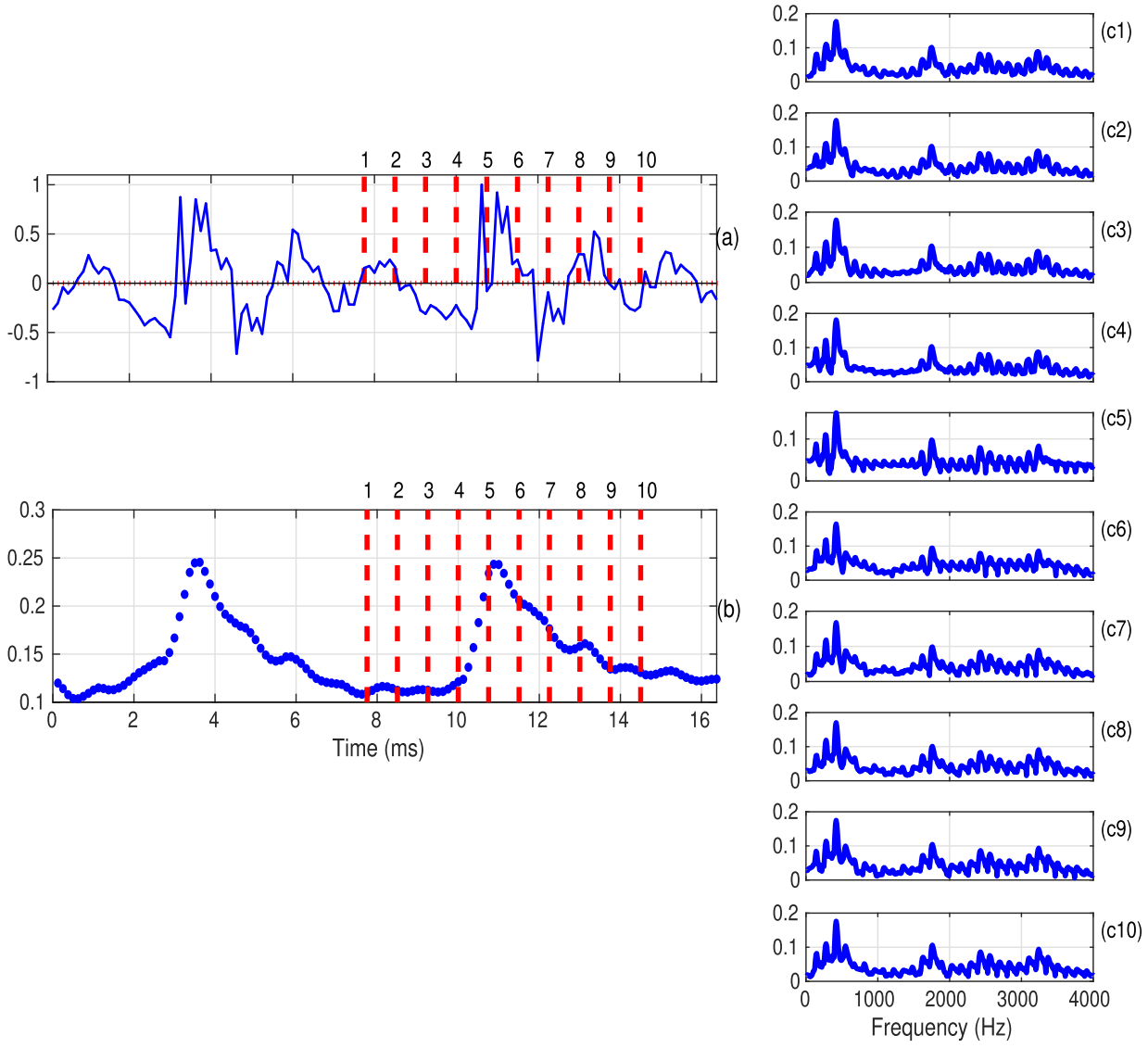

Fig. 9. An illustration of SFF spectral structure within a glottal cycle at ten instants. (a) A segment of voiced speech (along with ten instants marked by dotted lines). (b) Spectral flatness contour derived from SFF spectrum using $r = 0.995$. (c1)-(c10) are the SFF spectra at the marked instants 1 to 10 in (a) and (b).

sentation of the dynamics of the vocal tract system. Fig. 8(a) shows a segment of voiced speech, and Fig. 8(b) and (c) show the corresponding EGG and DEGG signals. The spectral flatness contour derived using a $L = 4$ ms window in the ZTW analysis is shown in Fig. 8(d). It can be seen that the spectral flatness contour has higher values in the regions where the EGG values are high, and the values of the spectral flatness are low in the closed phase region where the EGG values are low.

The transition of the spectral flatness value from low to high can be attributed to the coupling of the subglottal system with the supraglottal system. However, the change is gradual, and it also depends on the size of the analysis window in relation to the pitch period.

Fig. 10 illustrates the HNGD spectra obtained using a 4 *ms* window at ten equidistant locations within a glottal cycle. Fig. 10(a) shows the segment of voiced speech, Fig. 10(b) shows the corresponding spectral flatness contour. Fig. 10(c1)–(c10) show the normalized HNGD spectra for the locations at the time instants 1 to 10, marked in the speech signal and in the spectral flatness contour.

In Fig. 10(c1) and (c2), the spectral amplitude is distributed around a few resonances, as the analysis window is located in the closed phase region. As the analysis window approaches the glottal opening (Fig. 10(c3) and (c4)), the spectral amplitude is getting distributed more uniformly. It can also be observed that the spectral amplitude is tilting towards low frequency due to coupling of the subglottal system. We observe a sudden rise of the spectral flatness value in this region. As the analysis window approaches the instant of glottal closure, the spectral amplitude is higher at the prominent spectral peak, as can be seen in Fig. 10(c5) and (c6). Thus there is a gradual change in the spectral flatness from a higher value in the open phase region to a lower value in the closed phase region. Transition in the spectral flatness value from open to closed phase regions appears gradual.

In studies (Barney et al., 2007; Childers and Wong, 1994; Yegnanarayana and Veldhuis, 1998; Prasad and Yegnanarayana, 2016), the changes in the spectral characteristics of the vocal tract system during the open phase region within a glottal cycle were reported. The dissipation of the first formant energy was observed during the open phase of the glottal cycle, and it was termed as glottal damping. The transition of the spectral flatness to a higher value also takes place at the onset of glottal damping due to glottal opening, which is also in conformity with the studies reported in (Barney et al., 2007; Childers and Wong, 1994; Yegnanarayana and Veldhuis, 1998; Prasad and Yegnanarayana, 2016). It was also found that during the open phase there exists a shift

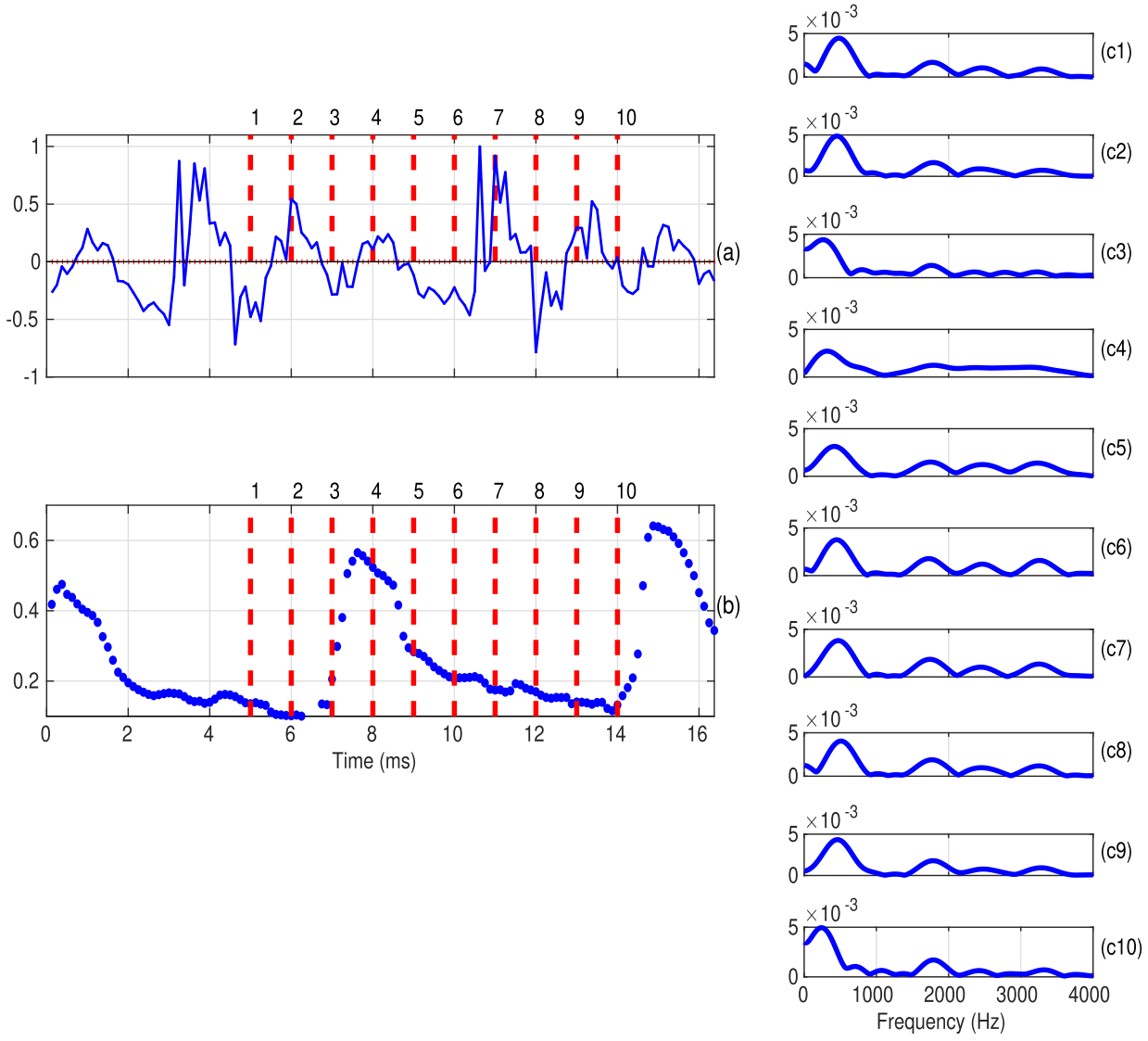

Fig. 10. An illustration of HNGD spectral structure within a glottal cycle at ten instants. (a) A segment of voiced speech (along with ten instants marked by dotted lines). (b) Spectral flatness contour derived from HNGD spectrum for a window length of 4 *ms*. (c1)-(c10) are the HNGD spectra at the marked instants 1 to 10 in (a) and (b).

of the formant locations and increase in the bandwidths (Barney et al., 2007; Childers and Wong, 1994). The HNGD spectral characteristics in the voiced speech also confirms this observation.

4. Effect of parameters used in SFF and ZTW methods on spectral flatness

The effect of r value in the SFF method is illustrated in Fig. 11. For higher r value (i.e., closer to one), the SFF provides good spectral resolution and lower temporal resolution due to smearing of certain features. For lower r value, the SFF provides good temporal resolution but lower spectral resolution of certain features. A segment of voiced speech and the corresponding DEGG signal are shown in Fig. 11(a) and (b). Fig. 11(c), (e), (g) and (i) show the SFF spectrograms and Fig. 11(d), (f), (h) and (j) show the corresponding spectral flatness contours for $r = 0.95, 0.97, 0.99$, and 0.995, respectively. From the SFF spectrograms, it can be clearly seen that the SFF spectrum appears to be flatter around the GCIs. Even though the SFF spectral flatness seems to be somewhat broader for higher r values, the increase in the SFF spectral flatness from lower to higher value occurs at the same location of GCI for different values of r.

The effect of the length of the analysis window in the ZTW method is illustrated in Fig. 12, which shows that the HNGD spectrograms and spectral flatness contours for different lengths of the analysis window varying from 3 to 6 ms. A segment of voiced speech and the corresponding DEGG signal are shown in Fig. 12(a) and (b). Fig. 12(c), (e), (g) and (i) show the HNGD spectrograms and Fig. 12(d), (f), (h) and (j) show the corresponding spectral flatness contours for $L = 3, 4, 5$, and 6 ms, respectively. From the HNGD spectrograms, it can be seen that the spectrum appears to be flatter in some regions between two successive GCIs. A longer (6 ms) analysis window gives a smoother spectral flatness contour compared to a smaller (3 ms) window. But the effects of the glottal opening region can still be observed in the spectral flatness contour. Hence, the spectral flatness contour obtained from the HNGD spectra using a suitable window size is useful to study the changes in the vocal tract during the opening phase.

For smaller (3 ms) window length, the spectral flatness contour has fluctuations, mostly due to window effects. As the window length is increased (from 3 ms to 6 ms), the temporal resolution in the spectral flatness contour is decreased due to the combined effect of closed and open regions. A window length in the range of 4–5 ms seems to be use-

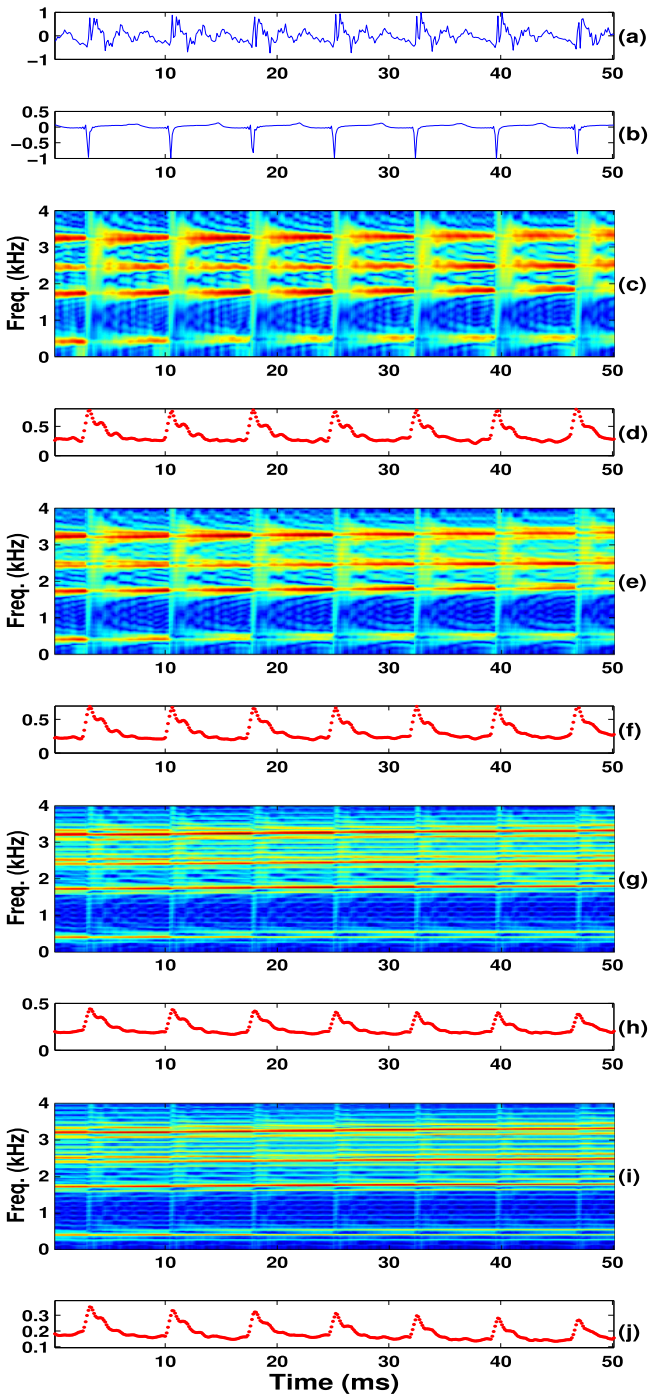

Fig. 11. An illustration of SFF spectrograms and spectral flatness contours derived from SFF method using different r values. (a) A segment of voiced speech. (b) DEGG signal. (c), (e), (g) and (i) are the SFF spectrograms, (d), (f), (h) and (j) are the corresponding spectral flatness contours, computed with r values of $r = 0.95, 0.97, 0.99$, and 0.995, respectively.

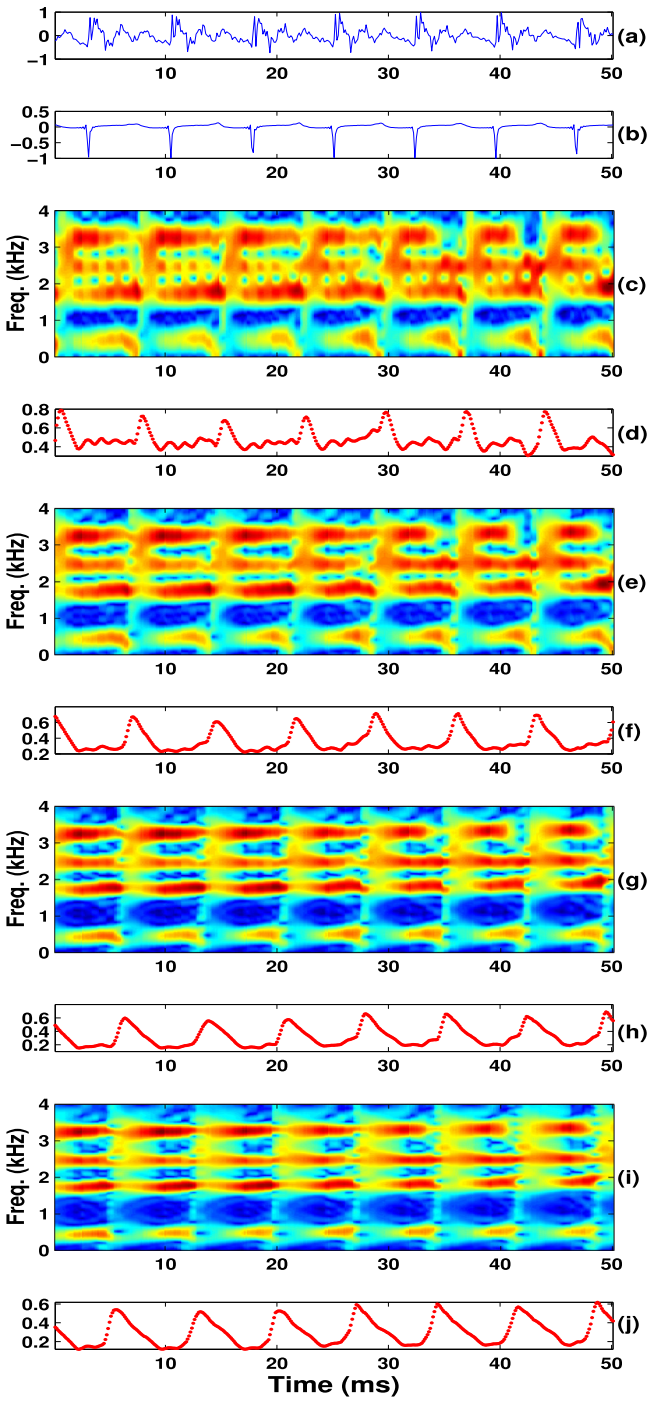

Fig. 12. An illustration of HNGD spectrograms and spectral flatness contours derived from ZTW method using different lengths of analysis window. (a) A segment of voiced speech. (b) DEGG signal. (c), (e), (g) and (i) are the HNGD spectrograms, (d), (f), (h) and (j) are the corresponding spectral flatness contours, computed with window size of 3, 4, 5 and 6 *ms*, respectively.

ful for identification of the glottal open region from the HNGD spectra, although the choice also depends on the pitch period.

5. Methods for detection of GCI and GOR

This section presents the methods for detection of GCI and GOR from speech signals.

5.1. Proposed method of GCI detection

The proposed method of GCI detection uses the ZFF method (Murty and Yegnanarayana, 2008). The ZFF method captures the major impulse-like discontinuity present in speech signal. In this study, we use the spectral flatness contour obtained using the SFF method as input to the ZFF method in order to detect the region of presence of GCIs. The

precise GCI location is obtained by searching for the spectral flatness peak in that region.

The following are the steps involved in the proposed method of GCI detection.

(a) Obtain the spectral envelope ($v_k[n]$) of the signal $s[n]$ using the SFF method as in Eq. (5) for $r = 0.995$.
(b) Derive the normalized spectral envelope ($\hat{v}_k[n]$) (Eq. (10)) and compute the spectral flatness $\delta_{SFF}[n]$ of $\hat{v}_k[n]$ across the frequency at each instant, as in Eq. (12).
(c) Spectral flatness contour is given as input to the ZFF method. The $\pm$ 1ms region around the positive-to-negative zero crossings of filtered output signal is hypothesized as the presence of GCIs.
(d) The location of the spectral flatness peak in the hypothesized GCIs region is marked as accurate GCI. The proposed method is referred as $SFF - SF$.

5.2. Proposed method for GOR detection

In this study, we propose to derive the open phase region using the spectral flatness obtained from the HNGD spectrum. The region between the spectral flatness peak to the following GCI in a glottal cycle is hypothesized as the open phase region. The following are the steps involved in deriving the open phase region within each glottal cycle.

(a) Compute the HNGD spectrum ($S_n[k]$) at each instant for a window length of 4 ms as described in Section 2.2.
(b) Derive the normalized spectral envelope ($\hat{S}_n[k]$) (Eq. (11)) and compute the spectral flatness $\delta_{ZTW}[n]$ of $\hat{S}_n[k]$ across the frequency at each instant, as in Eq. (13).
(c) Smooth the spectral flatness contour using a 5–point median filter to remove outliers, if any.
(d) Obtain the GCI locations for voiced segments as described in the above Section 5.1, and identify the region between two successive GCIs as the glottal cycle.
(e) The region between the spectral flatness peak in a glottal cycle to the following GCI region is identified as the open phase region. The rest of the glottal cycle is identified as the closed phase region (i.e., region between GCI to spectral flatness peak). The proposed method is referred as $ZTW - SF$.

6. Results and discussion

6.1. Experimental protocol

The CMU-Arctic databases (Kominek and Black, 2004) were considered to evaluate the proposed methods and to compare the results with the existing methods. The database consists of sentences spoken by two male (BDL and JMK) and one female (SLT) speakers. The database contains the simultaneous recordings of the EGG signals. Reference locations of the GCIs were extracted by finding peaks in the differenced EGG signal. For each sentence, 1/8 times of the maximum negative value of dEGG was used as a threshold (Murty and Yegnanarayana, 2008; Prathosh et al., 2013; Legát et al., 2011). The EGG and speech signals of each speaker were time aligned to compensate for the delay, which is approximately 0.7 ms (Murty and Yegnanarayana, 2008). A set of 100 utterances are considered in this study.

The proposed method of GCI detection is compared with the five state-of-art GCI detection methods. They are: ZFF (Murty and Yegnanarayana, 2008), SEDREAMS (Drugman et al., 2012b), MSF (Khanagha et al., 2014), DYPSA (Naylor et al., 2007), and YAGA (Thomas et al., 2012). The ZFF and YAGA methods are obtained from the authors, SEDREAMS is available in (cov, 0000), MSF is available in (msf, 0000), and DYPSA is available in (Brookes, 0000). The ZFF method exploits the nature of the impulse-like excitation by filtering the speech signal around 0 Hz. The SEDREAMS method uses the mean based signal for finding the possible GCI locations in an interval and then an LP residual is inspected in that to detect accurate location. The MSF method is based on the approach of microcanonical multi-scale formalism and relies on the precise estimation of multi-scale parameter called as singularity exponent at each sampling instant. The DYPSA method uses the zero crossings of the phase slope function calculated on the LP residual along with dynamic programming. The YAGA method uses the voice source signal (estimated using inverse filtering), group delay function, wavelet transform and M-best dynamic programming for identifying the glottal closure and opening instants.

Table 1
Performance comparison of the GCI detection methods for clean speech. IDR - Identification rate, MR - Miss rate, FAR - False alarm rate, IDA - Identification accuracy.

Database	Method	IDR%	MR %	FAR %	IDA (ms)
BDL	SFF-SF	97.58	1.20	1.22	0.35
	ZFF	97.87	0.20	1.93	0.29
	SEDREAMS	98.58	0.67	0.75	0.30
	MSF	95.26	2.30	2.44	0.33
	DYPSA	94.44	1.38	4.18	0.35
	YAGA	97.45	0.47	2.08	0.28
JMK	SFF-SF	95.44	4.10	0.46	0.44
	ZFF	97.59	2.25	0.16	0.42
	SEDREAMS	98.72	1.01	0.27	0.39
	MSF	95.45	1.36	3.19	0.41
	DYPSA	96.22	2.67	1.11	0.40
	YAGA	98.53	0.55	0.92	0.38
SLT	SFF-SF	95.91	3.25	0.84	0.34
	ZFF	99.01	0.65	0.34	0.24
	SEDREAMS	97.25	1.97	1.78	0.29
	MSF	96.39	3.26	0.35	0.35
	DYPSA	96.81	2.23	0.96	0.38
	YAGA	98.06	1.11	0.83	0.27

The glottal open phase regions obtained using the proposed method are compared with the open phase regions derived using three existing methods, namely SIGMA (Thomas and Naylor, 2009), YAGA (Thomas et al., 2012) and dominant resonance frequency (DRF) (Prasad and Yegnanarayana, 2016) methods. The SIGMA algorithm uses multiscale analysis and group delay function over EGG signals to compute the glottal closing and opening instants. In the SIGMA and YAGA methods, glottal open phase region is defined as the region between GOI and the following GCI. In DRF method, the glottal open phase region is derived using a threshold value of 0.5 over the normalized DRF contour. The region below this threshold is identified as the open phase region and the remaining region of the glottal cycle is identified as the closed phase region.

Performance of the GCI detection methods was evaluated using the measures defined in (Naylor et al., 2007). They are: identification rate (IDR), miss rate (MR) and false alarm rate (FAR) and the identification accuracy (IDA). For a good GCI method, IDR should be high (i.e., MR and FAR should be low) and IDA should be low. Performance of the GOR detection methods was evaluated using the scatter plots for the durations of open phase and closed phase as in (Prasad and Yegnanarayana, 2016).

6.2. Results of GCI detection methods

The GCI detection methods are evaluated in clean speech and telephone quality speech. Table 1 shows the results of GCI detection methods on the three databases for clean speech. From the table, it can be observed that the proposed method (SFF-SF) performs reasonably well and comparable to the existing methods for all the three databases. In some cases such as for BDL database, the IDR of the proposed method is better than the methods such as DYPSA and MSF, and comparable to ZFF, SEDREAMS and YAGA methods. In terms of IDA measure, methods such as YAGA, SEDREAMS and ZFF are performing better than the DYPSA, MSF and proposed method. The IDA of the proposed method is comparable to DYPSA and MSF in all the three databases.

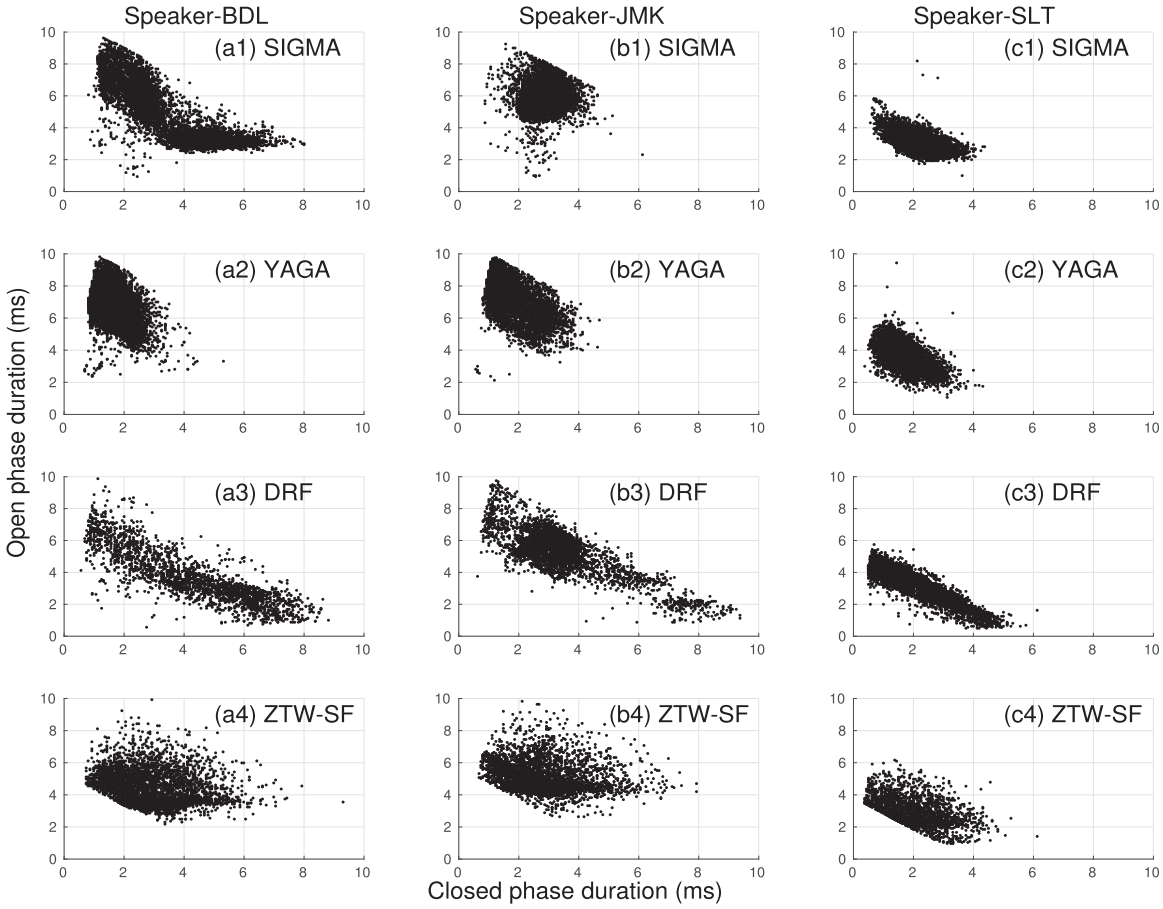

Fig. 13. Scatter plots of open *vs* closed phase duration points obtained for three speakers in the CMU-Arctic database. (a1), (b1) and (c1) are obtained using SIGMA method on EGG signals. (a2), (b2) and (c2) are obtained using YAGA method on speech signals. (a3), (b3) and (c3) are obtained using DRF method on speech signals. (a4), (b4) and (c4) are obtained using the proposed method (ZTW-SF) for window length of 4 ms.

Table 2
Performance comparison of the GCI detection methods for telephone quality speech. IDR - Identification rate, MR - Miss rate, FAR - False alarm rate, IDA - Identification accuracy.

Database	Method	IDR%	MR %	FAR %	IDA (ms)
BDL	SFF-SF	96.82	0.61	2.57	0.61
	ZFF	46.19	1.19	52.62	0.88
	SEDREAMS	45.97	2.02	52.01	0.84
	MSF	90.19	4.17	5.64	0.66
	DYPSA	87.08	3.94	8.98	0.65
	YAGA	34.06	1.11	64.83	0.93
JMK	SFF-SF	97.48	0.97	1.55	0.69
	ZFF	33.18	2.38	64.44	0.89
	SEDREAMS	31.01	0.84	68.15	1.21
	MSF	91.51	3.41	5.08	0.70
	DYPSA	86.71	2.69	10.60	0.72
	YAGA	32.88	2.11	65.01	0.93
SLT	SFF-SF	96.58	1.82	1.60	0.62
	ZFF	64.72	0.95	34.33	0.74
	SEDREAMS	61.92	3.04	35.04	0.69
	MSF	87.19	8.98	3.83	0.63
	DYPSA	71.63	2.89	25.48	0.64
	YAGA	70.29	1.77	27.94	0.75

The results of GCI detection methods for telephone quality speech are given in Table 2. Telephone quality speech is simulated using G.191 software (ITU-T, Recommendation, 2005). From the table, it can be observed that the performance of the existing methods for all the three databases is severely degraded for telephone quality speech compared to clean speech (Table 1). On the other hand, the performance of the proposed method is significantly better than all the existing methods for all the three databases in terms of both IDR and IDA measures. The degradation in performance of the existing methods for telephone quality speech is mainly due to their inability to capture the impulse-like discontinuity corresponding to GCI from the speech signal (as in ZFF, SEDREAMS and MSF) or derived glottal source signal (as in DYPSA and YAGA). Among the existing methods, the performance of the DYPSA and MSF are relatively higher than ZFF, SEDREAMS and YAGA methods. The performance of DYPSA is better than YAGA because of the use of LP residual signal directly, without estimation of glottal source signal as in YAGA. It is known that estimation of glottal source signal from telephone speech is a difficult task. Overall, the performance of the proposed method is better in terms of both IDR and IDA measures compared to all the existing methods due to exploitation of impulse-like discontinuity using spectral flatness parameter.

6.3. Results of GOR detection methods

Fig. 13 shows the scatter plots for the durations of open phase and closed phase for SIGMA, YAGA, DRF and the proposed method. The three columns in the figure correspond to three different speakers, namely BDL, JMK and SLT. The results using SIGMA method are shown in Figs. 13(a1), (b1) and (c1), YAGA method are shown in Figs. 13(a2), (b2) and (c2), DRF method are shown in Figs. 13(a3), (b3) and (c3). Finally, Figs. 13(a4), (b4) and (c4) show the results using the proposed method (ZTW-SF) for window length 4 ms. From the figures, it can be observed that the scatter plots of the open *vs* closed phase points appear

in a relatively tight cluster in the case of SIGMA and YAGA methods. This is mostly because of the fact that the GOIs obtained from the EGG/voice source signal are post-processed using dynamic programming to ensure the best alignment of open quotient values over successive glottal cycles. This constraints the locations of GOIs within close bounds, resulting in a tight cluster.

The scatter plots obtained using DRF method exhibits more spread in the cluster compared to the results obtained using SIGMA and YAGA methods. Also it can be observed that, there are some points occurring in the extreme values of open and closed duration values. It was found that these regions occur during the onset and offset of voiced segments due to gradual build up of the glottal source characteristics and also depends the thresholds (Prasad and Yegnanarayana, 2016). The scatter plots obtained using the proposed method exhibits relatively less spread in the cluster compared to DRF method and relatively more spread in the cluster compared to the results of SIGMA and YAGA methods. This is mainly because the proposed method does not use thresholds or dynamic programming constraints. Results obtained from the proposed method appear closer to those obtained using the SIGMA and YAGA methods. Different post-processing techniques can be used to refine the open phase regions obtained by the proposed method.

7. Summary and conclusion

In this paper, the problem of detecting some of the features of glottal activity, namely, glottal closure instant (GCI) and glottal open region (GOR) from the speech signal is addressed. The changes in the spectral characteristics within a glottal cycle are exploited using two recently proposed signal processing methods, namely, single frequency filtering (SFF) and zero time windowing (ZTW). A spectral flatness parameter derived from the SFF spectra highlights the impulse-like characteristics at the GCI. The spectral flatness derived from the HNGD spectra highlights the GOR, as there are significant changes in the effective vocal tract length in the glottal opening region. The spectral flatness representation of the dynamic vocal tract characteristics may provide complementary information with other types of source representations, and thus may help to improve our understanding of the glottal source.

The proposed methods for the detection of GCI and GOR do not use any knowledge of the periodicity information of the excitation source or of the resonance characteristics of the vocal tract system. The proposed approach may be useful to analyze changes in the glottal source characteristics for different types of voices, like different phonations, emotions, singing, laughter, pathological voices (Titze, 2008; Laver, 1980; Mittal and Yegnanarayana, 2013; Kadiri et al., 2015; Henrich et al., 2011; Thati et al., 2013; Moore et al., 2008), etc. Preliminary experiments in (Kadiri and Yegnanarayana, 2018b; 2018a; Kadiri, 2018) showed the usefulness of these methods for the analysis and detection of phonation types in speech and singing. Detailed analysis of the SFF and HNGD spectral flatness contours may help in bringing out the distinctive characteristics of different voice qualities caused by different types of glottal excitations.

Since the ZTW method provides information at each sampling instant with good spectral resolution, these studies can be explored for analysis of subglottal effect on the supraglottal system (Lulich et al., 2009), and also for analysis of source-tract interaction within a glottal cycle (Chi and Sonderegger, 2007; Titze, 2004; Guerin et al., 1976; Rothenberg, 1981).

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgements

The first author would like to thank the Academy of Finland (project no. 312490) for supporting his stay in Finland as a Postdoctoral Researcher. The third author would like to thank the Indian National Science Academy (INSA) for their support.

References

Abberton, E.R.M., Howard, D.M., Fourcin, A.J., 1989. Laryngographic assessment of normal voice: a tutorial. Clin. Linguist. Phonet. 3 (3), 263–296.
Airaksinen, M., Raitio, T., Story, B., Alku, P., 2014. Quasi closed phase glottal inverse filtering analysis with weighted linear prediction. IEEE/ACM Trans. Audio, Speech Lang. Process. 22 (3), 596–607.
Alku, P., 1992. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. Speech Commun. 11, 109–118.
Alku, P., 2011. Glottal inverse filtering analysis of human voice production - a review of estimation and parameterization methods of the glottal excitation and their applications. Sadhana 36 (5), 623–650.
Alku, P., Magi, C., Yrttiaho, S., Backstrom, T., Story, B., 2009. Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering. J. Acoust. Soc. Am. 120, 3289–3305.
Ananthapadmanabha, T., Yegnanarayana, B., 1979. Epoch extraction from linear prediction residual for identification of closed glottis interval. IEEE Trans. Speech Audio Process. 27, 309–319.
Aneeja, G., Yegnanarayana, B., 2015. Single frequency filtering approach for discriminating speech and nonspeech. IEEE/ACM Trans. Audio, Speech Lang. Process. 23 (4), 705–717.
Barney, A., De Stefano, A., Henrich, N., 2007. The effect of glottal opening on the acoustic response of the vocal tract. Acta Acustica united with Acustica 93 (6), 1046–1056.
Bouzid, A., Ellouze, N., 2009. Voice source parameter measurement based on multi-scale analysis of electroglottographic signal. Speech Commun. 51 (9), 782–792.
Brookes, M.,. Voicebox: speech processing toolbox for matlab. Source: https://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html.
Chi, X., Sonderegger, M., 2007. Subglottal coupling and its influence on vowel formants. J. Acoust. Soc. Am. 122 (3), 1735–1745.
Childers, D.G., Krishnamurthy, A.K., 1984. A critical review of electroglottography. Crit. Rev. Biomed. Eng. 12 (2), 131–161.
Childers, D.G., Wong, C.-F., 1994. Measuring and modeling vocal source-tract interaction. IEEE Trans. Biomed. Eng. 41 (7), 663–671.
D Alessandro, C., Sturmel, N., 2011. Glottal closure instant and voice source analysis using time-scale lines of maximum amplitude. Sadhana 36 (5), 601–622.
Degottex, G., Roebel, A., Rodet, X., 2009. Glottal closure instant detection from a glottal shape estimate. In: 13th International Conference on Speech and Computer (SPECOM), St-Petersburg, Russia, pp. 226–231.
Degottex, G., Roebel, A., Rodet, X., 2010. Joint estimate of shape and time-synchronization of a glottal source model by phase flatness. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Dallas, USA, pp. 5058–5061.
Degottex, G., Roebel, A., Rodet, X., 2011. Function of phase-distortion for glottal model estimation. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Prague, Czeck Republic, pp. 4608–4611.
Degottex, G., Roebel, A., Rodet, X., 2011. Phase minimization for glottal model estimation. IEEE Transactions on Acoustics, Speech and Language Processing 19 (5), 1080–1090.
Drugman, T., Alku, P., Alwan, A., Yegnanarayana, B., 2014. Glottal source processing: from analysis to applications. Comput. Speech Lang. 28 (5), 1117–1138.
Drugman, T., Bozkurt, B., Dutoit, T., 2011. Causal-anticausal decomposition of speech using complex cepstrum for glottal source estimation. Speech Commun. 53, 855–866.
Drugman, T., Bozkurt, B., Dutoit, T., 2012. A comparative study of glottal source estimation techniques. Comput. Speech Lang. 26, 20–34.
Drugman, T., Thomas, M., Gudnason, J., Naylor, P., Dutoit, T., 2012. Detection of glottal closure instants from speech signals: a quantitative review. IEEE Trans. Audio Speech Lang. Process. 20 (3), 994–1006.
Fant, G., 1995. The LF-model revisited. transformations and frequency domain analysis. Speech Transm. Lab. Q. Progr.Status Report 36, 119–156.
Fu, Q., Murphy, P., 2006. Robust glottal source estimation based on joint source-filter model optimization. IEEE Trans. Audio Speech Lang.Process. 14, 492–501.
Guerin, B., Mrayati, M., Carre, R., 1976. A voice source taking account of coupling with the supraglottal cavities. In: ICASSP, 1, pp. 47–50.
Henrich, M.K.N., Herzel, H., Howard, D., Tokuda, I., Wolfe, J., 2011. Analysing and understanding the singing voice : recent progress and open questions. Curr. Bioinform. 6 (3), 362–374.
Henrich, N., D Alessandro, C., Doval, B., Castellengo, M., 2004. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. J. Acoust. Soc. Am. 115 (3), 1321–1332.
Herbst, C.T., Lohscheller, J., Švec, J.G., Henrich, N., Weissengruber, G., Fitch, W.T., 2014. Glottal opening and closing events investigated by electroglottography and super-high-speed video recordings. J. Exper. Biol. 217 (6), 955–963.
ITU-T, Recommendation, 2005. G.191, software tools for speech and audio coding standardization. Source: http://www.itu.int/rec/T-REC-G.191-200509-I/en.
Jain, P., Pachori, R.B., 2012. Time-order representation based method for epoch detection from speech signals. J. Intell. Syst. 21 (1), 79–95.
Jain, P., Pachori, R.B., 2013. Gci identification from voiced speech using the eigen value decomposition of Hankel matrix. In: International Symposium on Image and Signal Processing and Analysis (ISPA), pp. 371–376.
Jain, P., Pachori, R.B., 2014. Event-based method for instantaneous fundamental frequency estimation from voiced speech based on eigenvalue decomposition of the hankel matrix. IEEE/ACM Trans. Audio SpeechLang. Process. 22 (10), 1467–1482.
Kadiri, S.R., 2018. Analysis of Excitation Information in Expressive Speech. Speech Processing Laboratory, IIIT Hyderabad Ph.D. thesis..

Kadiri, S.R., Gangamohan, P., Gangashetty, S.V., Yegnanarayana, B., 2015. Analysis of excitation source features of speech for emotion recognition. In: INTERSPEECH, pp. 1324–1328.
Kadiri, S.R., Yegnanarayana, B., 2017. Epoch extraction from emotional speech using single frequency filtering approach. Speech Commun. 86, 52–63.
Kadiri, S.R., Yegnanarayana, B., 2018. Analysis and detection of phonation modes in singing voice using excitation source features and single frequency filtering cepstral coefficients (SFFCC). In: INTERSPEECH, pp. 441–445.
Kadiri, S.R., Yegnanarayana, B., 2018. Breathy to tense voice discrimination using zero–time windowing cepstral coefficients (ZTWCCs). In: INTERSPEECH, pp. 232–236.
Kafentzis, G., Stylianou, Y., Alku, P., 2011. Glottal inverse filtering using stabilised weighted linear prediction. In: ICASSP, pp. 5408–5411.
Khanagha, V., Daoudi, K., Yahia, H., 2014. Detection of glottal closure instants based on the microcanonical multiscale formalism. IEEE/ACM Trans. Audio, Speech Lang. Process. 22 (12), 1941–1950.
Kominek, J., Black, A., 2004. The CMU Arctic speech databases. In: 5th ISCA Speech Synthesis Workshop, pp. 223–224.
Krishnamurthy, A., Childers, D., 1986. Two-channel speech analysis. IEEE Trans. Audio Speech Signal Process. 34, 730–743.
Larsson, H., Hertegård, S., Lindestad, P.-Å., Hammarberg, B., 2000. Vocal fold vibrations: high-speed imaging, kymography and acoustic analysis: a preliminary report. Laryngoscope 110, 2117–2122.
Laver, J., 1980. The Phonetic Description of Voice Quality. Cambridge University Press.
Legát, M., Matoušek, J., Tihelka, D., 2011. On the detection of pitch marks using a robust multi-phase algorithm. Speech Commun. 53 (4), 552–566.
Lieberman, P., 1963. Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. J. Acoust. Soc. Am. 35, 344–353.
Lohscheller, J., Eysholdt, U., Toy, H., Dollinger, M., 2008. Phonovibrography: mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing and analyzing the underlying laryngeal dynamics. IEEE Trans. Med. Imag. 27 (3), 300–309.
Lulich, S.M., Zanartu, M., Mehta, D.D., Hillman, R.E., 2009. Source-filter interaction in the opposite direction: subglottal coupling and the influence of vocal fold mechanics on vowel spectra during the closed phase. In: Proceedings of Meetings on Acoustics, 6.
Ma, Y.K.C., Willems, L.F., 1994. A frobenius norm approach to glottal closure detection from the speech signal. IEEE Trans. Speech Audio Process. 2, 258–265.
Mehta, D., Deliyski, D., Quatieri, T., Hillman, R., 2011. Automated measurement of vocal fold vibratory asymmetry from high-speed videoendoscopy recordings. J. Speech Lang. Hear. Res. 54, 47–54.
Mittal, V.K., Yegnanarayana, B., 2013. Effect of glottal dynamics in the production of shouted speech. J. Acoust. Soc. Am. 133 (5), 3050–3061.
Moore, E., Clements, M., Peifer, J., Weisser, L., 2008. Critical analysis of the impact of glottal features in the classification of clinical depression in speech. IEEE Trans. Biomed. Eng. 55 (1), 96–107.
Moulines, E., Di Francesco, R., 1990. Detection of the glottal closure by jumps in the statistical properties of the speech signal. Speech Commun. 9 (5), 401–418.
Murty, K.S.R., Yegnanarayana, B., 2008. Epoch extraction from speech signals. IEEE Trans. Audio Speech Lang. Process. 16 (8), 1602–1613.
Naylor, P.A., Kounoudes, A., Gudnason, J., Brookes, M., 2007. Estimation of glottal closure instants in voiced speech using the DYPSA algorithm. IEEE Trans. Audio Speech Lang. Process. 15 (1), 34–43.
Prasad, R.S., Yegnanarayana, B., 2016. Determination of glottal open regions by exploiting changes in the vocal tract system characteristics. J. Acoust. Soc. Am. 140 (1), 666–677.
Prathosh, A., Ananthapadmanabha, T., Ramakrishnan, A., 2013. Epoch extraction based on integrated linear prediction residual using plosion index. IEEE Trans. Audio Speech Lang. Process. 21 (12), 2471–2480.
Ramesh, K., Prasanna, S.R.M., Govind, D., 2013. Detection of glottal opening instants using Hilbert envelope. In: INTERSPEECH, pp. 44–48.
Rao, K.S., Prasanna, S.R.M., Yegnanarayana, B., 2007. Determination of instants of significant excitation in speech using hilbert envelope and group-delay function. IEEE Signal Process. Letters 14 (10), 762–765.
Rothenberg, M., 1981. Acoustic interaction between the glottal source and the vocal tract. Vocal Fold Physiol. 305–328.
Rothenberg, M., Mahshie, J.J., 1988. Monitoring vocal fold abduction through vocal fold contact area. J. Speech Hear. Res. 31, 338–351.
Schleusing, O., Kinnunen, T., Story, B.H., Vesin, J., 2013. Joint source-filter optimization for accurate vocal tract estimation using differential evolution. IEEE Trans. Audio Speech Lang. Process. 21 (8), 1560–1572.
Silva, D., Oliveira, L., Andrea, M., 2009. Jitter estimation algorithms for detection of pathological voices. EURASIP J. Adv. Signal Process..
Source: https://covarep.github.io/covarep/.
Source: https://geostat.bordeaux.inria.fr/index.php/downloads.html.
Stevens, K.N., 1977. Physics of laryngeal behavior and larynx models. Phonetica 34, 264–279.
Thati, S.A., Kumar K, S., Yegnanarayana, B., 2013. Synthesis of laughter by modifying excitation characteristics. J. Acoust. Soc. Am. 133 (5), 3072–3082.
Thomas, M.R.P., Gudnason, J., Naylor, P.A., 2012. Estimation of glottal closing and opening instants in voiced speech using the yaga algorithm. IEEE Trans. Audio Speech Lang. Process. 20 (1), 82–91.
Thomas, M.R.P., Naylor, P., 2009. The sigma algorithm: a glottal activity detector for electroglottographic signals. IEEE Trans. Audio Speech Lang. Process. 17 (8), 1557–1566.
Titze, I., 2004. Theory of glottal airflow and source-filter interaction in speaking and singing. Acta Acustica united with Acustica 90 (4), 641–648.
Titze, I.R., 2008. Nonlinear source filter coupling in phonation: theory. J. Acoust. Soc. Am. 123 (5), 2733–2749.
Veldhuis, R., 1998. A computationally efficient alternative for the liljencrants-Fant model and its perceptual evaluation. J. Acoust. Soc. Am. 103 (1), 566–571.
Walker, J., Murphy, P., 2007. A review of glottal waveform analysis. Springer Lecture Notes Comput. Sci. (LNCS) 4391, 1–21.
Wong, D., Markel, J., Gray, A., 1979. Least squares glottal inverse filtering from the acoustic speech waveform. IEEE Trans. Audio Speech Signal Process. 27, 350–355.
Yan, Y., Chen, X., Bless, D., 2006. Automatic tracing of vocal-fold motion from high-speed digital images. IEEE Trans. Biomed. Eng. 53 (7), 1394–1400.
Yegnanarayana, B., Gangashetty, S.V., 2011. Epoch-based analysis of speech signals. Sadhana 36 (5), 651–697.
Yegnanarayana, B., Gowda, D.N., 2013. Spectro-temporal analysis of speech signals using zero-time windowing and group delay function. Speech Commun. 55 (6), 782–795. doi:10.1016/j.specom.2013.02.007.
Yegnanarayana, B., Veldhuis, N., 1998. Extraction of vocal-tract system characteristics from speech signals. IEEE TASP 6, 313–327.