An Error Correction Scheme for GCI Detection Algorithms using Pitch Smoothness Criterion

Sujith. P [1] *, Prathosh A. P.*[2]*, Ramakrishnan A. G.*[3] *and Prasanta. K. Ghosh*[3]

[1] Ittiam systems, Banglore, India.
[2]Xerox research center India, Banglore, India.
[3,4] Department of Electrical engineering, Indian Institute of science, Banglore, India.

sujith.p4@gmail.com, Prathosh.AP@xerox.com, ramkiag@ee.iisc.ernet.in, prasantg@ee.iisc.ernet.in

Abstract

Detection of error-free glottal closure instants (GCI) is a critical requirement for many applications including text-to-speech synthesis, causal anti-causal decomposition and voice morphing. Many existing GCI detection algorithms commit errors under certain conditions. In this paper, we propose a post-processing scheme for correcting errors of any GCI detection algorithm. The proposed error correction scheme works on the principle that the fundamental frequency over a voiced segment is slowly varying. The error correction is, thus, formulated as an optimization problem such that the pitch contour from the corrected GCIs has the least high frequency components. The proposed error correction scheme is experimentally evaluated on speech corpus with simultaneous EGG recordings using three state-of-the-art GCI detection algorithms viz., Dynamic Plosion Index (DPI), Zero Frequency Resonator (ZFR), and Speech Event Detection using the Residual Excitation And a Mean-based Signal (SEDREAMS). It is found that the proposed error correction scheme improves the performance of the GCI detection in clean speech as well as noisy conditions at different SNRs.

Index Terms: GCI detection, error correction, missing location estimation, epoch extraction.

1. Introduction

Pitch-synchronous analysis of the voiced speech signal is a popular technique in which the glottal closure instants (GCIs or epochs) are used to define the analysis frames. Epochs are utilized in various applications including pitch tracking, voice source estimation [1], speech synthesis [stylianou2001applying, lakkavalli2010continuity], prosody modification [2, 3, 4, 5], voiced/unvoiced boundary detection [6] and speaker identification [7]. A primary requirement for such an analysis is the knowledge of the precise locations of the GCIs. Hence, the automatic detection of the GCIs from the voiced speech signal is considered to be an important problem in speech research [8]. GCI detection algorithms that offer very high performance include the Zero Frequency Resonator-based method (ZFR) [9], the Speech Event Detection using the Residual Excitation And a Mean-based Signal (SEDREAMS) [10] and the Dynamic Plosion Index (DPI) algorithm [11]. While these techniques have high performance under clean as well as noisy conditions [12], they suffer from certain limitations which may affect the performance of the systems incorporating them. For example, if an algorithm misses a few GCIs within a vowel region, sudden glitches are perceived in a speech signal, synthesized using a method involving these GCIs. A complex cepstrum-based speech deconvoluter is another example where even a single missing GCI causes high distortions [13]. Hence, it is necessary to have a GCI detection scheme as error-free as possible. Most of the popular state-of-the-art GCI detectors commit errors under some conditions. For example, ZFR and SEDREAMS, which are based on filtering the speech signal around the fundamental frequency, may perform poorly on speech with very-low fundamental frequency content such as in telephone quality speech [11]. Further, these two algorithms depend on the a-priori estimated value of the average pitch period. Thus, if the average pitch period is estimated wrongly, these algorithms commit errors. Also, the temporal accuracy of the ZFR is observed to be low [12]. The DPI algorithm, on the other hand, is shown to offer very high-temporal accuracy and reasonable performance in the presence of noise. However, the integrated linear prediction residual (ILPR) [14] forms the basis for the DPI computation and the DPI algorithm may fail when the estimation of the ILPR is erroneous. ILPR may not be accurate for phonemes like nasals which do not adhere to the all-pole assumption of the LP-model. Thus, it is required to correct the errors committed by these algorithms to ensure better performance for systems that employ GCI detectors. In this paper, we propose an error-correction scheme for a GCI detection algorithm using a missing location estimation approach formulated as an optimization problem. We evaluate the proposed scheme on three recent and popular GCI detectors viz., DPI, ZFR, and SEDREAMS using speech with simultaneous electroglottograph (EGG) recordings in both clean and noisy conditions.

2. Proposed method

2.1. Motivation

It is well known that the speech signal exhibits quasi-periodicity during the production of voiced sounds, due to the presence of periodic vibrations of the vocal folds. The frequency of this vibration is termed as the pitch or the fundamental frequency. The inverse of the interval between two successive epochs or GCIs is the pitch for the laryngeal cycle described by such successive epochs. It is noticed in the literature that the pitch frequency does not vary very rapidly, over an utterance [15]. In other words, the contour of the inter-epoch-intervals (IEI) is typically smooth, in that it is predominantly low-pass in nature. This forms the basis for formulating an optimization problem to perform GCI error correction in the IEI sequence domain as described in the following subsection.

2.2. Problem setting

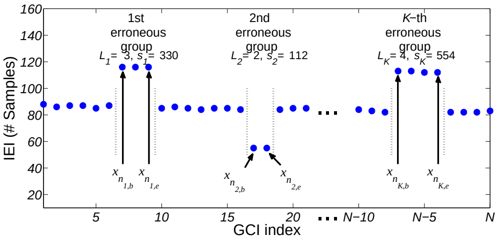

Figure 1: An illustration of the erroneous groups in an IEI sequence. Solid blue circles form the IEI sequence obtained from a GCI detection algorithm.

Let $x[n],\ 1 \le n \le N$, be an IEI sequence obtained from a GCI detection algorithm (as shown in Fig. 1) on a voiced segment of an utterance. We assume that the erroneous groups in $x[n]$ are known a priori. An erroneous group is a set of IEI indices n, where the GCI detection algorithm has made 'errors'. Let there be K such groups of erroneous IEIs in the considered voiced segment (as illustrated in Fig. 1). Let the i-th erroneous group begin at sample index $n_{i,b}$ and end at $n_{i,e}$. Also let the number of IEIs in the i-th erroneous group be L_i. Let $L = \sum_{i=1}^{K} L_i$ and $s_i = \sum_{n=n_{i,b}}^{n_{i,e}} x[n]$. We would like to estimate the IEIs in the erroneous groups such that after estimation, the entire IEI sequence is slowly varying with minimal high frequency contents. Let the estimate of the actual number of IEIs in the i-th erroneous group be $\tilde{L}_i$ and $\tilde{L} = \sum_{i=1}^{K} \tilde{L}_i$. Thus the length of the entire IEI sequence (denoted by $\tilde{x}[n]$) after error correction is $N - L + \tilde{L}$.
Let $\mathbf{x}_m^i$ be a $\tilde{L}_i$-dimensional column vector whose j-th element is denoted by $\mathbf{x}_m^{i,j}$, representing the estimate of the j-th IEI in the i-th erroneous group. Let $\mathbf{x}_m = [{\mathbf{x}_m^1}^{\mathrm{T}}, {\mathbf{x}_m^2}^{\mathrm{T}}, \ldots, \ldots, {\mathbf{x}_m^K}^{\mathrm{T}}]^{\mathrm{T}}$ be the $\tilde{L}$-dimensional vector whose elements are obtained by concatenating the estimates of all the erroneous IEIs. 'T' denotes the transpose of a vector. Thus, the problem of GCI error correction is equivalent to the estimation of $\mathbf{x}_m$.
In order to minimize the high frequency components of $\tilde{x}[n]$, we formulate the estimation of $\mathbf{x}_m$ as an optimization problem where the energy of high frequency content of $\tilde{x}[n]$ is minimized. Let $h[n]$ be the impulse response of a causal high-pass filter with cut-off frequency f_c. Thus the objective function of the proposed minimization can be written as follows:

$$f(\mathbf{x}_m) = \sum_{n=1}^{\infty} (\tilde{x}[n] \star h[n])^2, \tag{1}$$

where $\tilde{x}[n]$ is identical to $x[n]$ in the non-erroneous regions of the IEI sequence and denoted by the elements of $\mathbf{x}_m$ over the erroneous groups. In Eq. 1, $\star$ denotes the convolution operator. Thus, the objective function is the energy of the sequence obtained after high-pass filtering the re-estimated IEI.
It should be noted that the estimation of $\mathbf{x}_m$ needs to be constrained by the fact that the estimated IEIs ($\mathbf{x}_m^i$) in the i-th erroneous group should satisfy the duration constraint $\sum_{j=1}^{\tilde{L}_i} \mathbf{x}_m^{i,j} = s_i$ so that the GCIs in the non-erroneous groups remain unaltered. Also, the optimized vector $\mathbf{x}_m$ should have positive integer-valued elements since the IEIs are computed in number of samples. Thus, the proposed optimization problem has two constraints as follows:

$$\begin{aligned} \mathbf{x}_m^* &= \text{argmin } f(\mathbf{x}_m) \\ \text{subject to } &\sum_{j=1}^{\tilde{L}_i} \mathbf{x}_m^{i,j} = s_i, 1 \le i \le K \text{ and } \mathbf{x}_m \in \mathbb{Z}_+^{\tilde{L}} \end{aligned} \tag{2}$$

The objective function can be re-written in a matrix-vector form as

$$f(\mathbf{x}_m) = \frac{1}{2}\mathbf{x}_m^T \mathbf{A} \mathbf{x}_m + \mathbf{B}^T \mathbf{x}_m + const. \tag{3}$$

$\mathbf{A}$ is a symmetric $\tilde{L} \times \tilde{L}$ matrix and $\mathbf{B}$ is a $\tilde{L} \times 1$ vector. The (i, j)-the element of $\mathbf{A}$ and i-th element of $\mathbf{B}$ are given respectively by

$$\mathbf{A}_{ij} = R^h_{(i-j)} \triangleq \sum_{n=0}^{\infty} h[n]h[n-(i-j)], \text{ and}$$

$$\mathbf{B}_i = \sum_{n=i}^{\infty} \left(\sum_{k=0}^{\infty} \hat{x}[n-k]h[k] \right) h[n-i] = \hat{x}[n] \star R^h_n \text{ at } i,$$

$$\text{where, } \hat{x}[k] = \begin{cases} 0, & \text{if the } k\text{-th term is missing} \\ \tilde{x}[k], & \text{otherwise} \end{cases}$$

R^h is the autocorrelation sequence of the impulse response of the filter $h[n]$. Since $\mathbf{A}$ is an autocorrelation matrix, it is Toeplitz and positive semidefinite.
The constraint can be also written in the matrix-vector form as

$$\mathbf{C}\mathbf{x}_m = \mathbf{s} \text{ , where } \mathbf{C} \text{ is a } K \times \tilde{L} \text{ matrix and}$$

$$\mathbf{C}_{ij} = \begin{cases} 1, & \text{if } \sum_{k=1}^{i-1} \tilde{L}_k \le j \le \sum_{k=1}^{i} \tilde{L}_k \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Thus, the optimization problem now becomes

$$\begin{aligned} \mathbf{x}_m^* = \text{argmin } &\frac{1}{2}\mathbf{x}_m^T \mathbf{A} \mathbf{x}_m + \mathbf{B}^T \mathbf{x}_m + const. \\ &\text{subject to } \mathbf{C}\mathbf{x}_m = \mathbf{s},\ \mathbf{x}_m \in \mathbb{Z}_+^{\tilde{L}} \end{aligned} \tag{5}$$

This problem fits into the well-known optimization framework of Mixed Integer Quadratic Programming (MIQP) problem. Several approaches have been proposed in the literature for MIQP problem [**?**, **?**, **?**, **?**]. In this work, we use the OPTI toolbox [16] for solving the above optimization problem. Once the IEIs are estimated, they are converted to GCIs.

2.3. Implementation details

In the proposed GCI error correction scheme, it is assumed that the erroneous groups are known a priori. Also, the average pitch period (APP) is required to be known for estimating the actual number of IEIs in each erroneous region. If a GCI detection algorithm uses the APP a priori (for example, ZFR and SEDREAMS), it is also used in the GCI correction scheme; APP is estimated using the ground truth GCIs as explained in section III. If the GCI detection algorithm does not require the APP (for example DPI), it is estimated from the GCIs given by the algorithm. This is done by first computing IEIs from the GCIs and then finding the mode of IEIs over an utterance. The APP derived in this manner is assumed to be a good approximation to the actual APP. That is, if θ is the APP in an utterance, a sample in the IEI sequence is considered to be erroneous if it is δ away from θ. For the experiments considered here, the δ is taken to be 20% of θ, based on the assumption that the pitch of an utterance can vary between $\pm 20\%$ of the APP [1]. A set of contiguous erroneous IEIs is treated as an 'erroneous group'. The IEIs in all the erroneous groups are re-estimated using Eq. 5.

We estimate the number of missing samples ($\tilde{L}_i$) in each erroneous segment as follows: an initial estimate is computed as the ratio of s_i to the APP (θ). Then the optimal number of missing samples in each segment is estimated without altering the estimated values of the remaining segments. In order to find the optimal number of missing samples for the i-th segment, we search from $\lfloor\{s_i/(\theta+\delta)\}\rfloor$ to $\lceil\{s_i/(\theta-\delta)\}\rceil$ with a step size of 1 and $\tilde{L}_i$ is determined to be the one where the objective function achieves the minimum value.

The high-pass filter $h[n]$ is chosen to be an FIR filter of order 200. Figure 2 shows an illustrative example of how the proposed GCI error correction scheme works on a voiced speech segment at 5 dB additive noise, when DPI is used for GCI detection. The true IEIs obtained from the ground truth are shown in Fig. 2(a). The IEIs estimated by the DPI algorithm and those obtained after applying the proposed post-processing technique are shown in Figs. 2 (b) and (c), respectively. It is seen that the errors in GCIs caused by the GCI detection algorithm are corrected by the proposed post-processing technique.

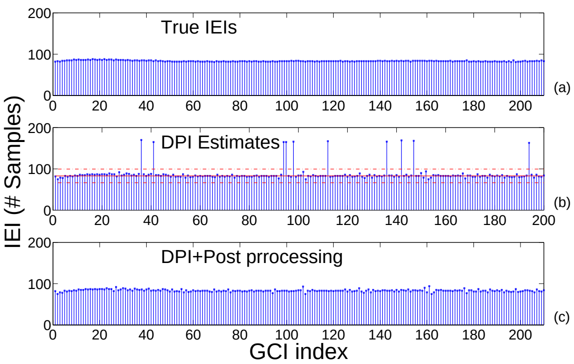

Figure 2: Illustration of the results of error correction scheme on a voiced segment of an utterance. IEIs obtained by the ground truth, DPI algorithm before and after pre-processing are shown in traces (a), (b) and (c), respectively. The horizontal dot-dash lines in Fig. (b), represents the $\pm$ 20% values around the average pitch period

3. Experiments and results

3.1. Database, experimental setup and performance measures

The dataset provided with the book 'Speech processing and Synthesis toolboxes' by D. G. Childers [17] is used for evaluation in this work (henceforth referred to as the Childers' data). Childers' data includes speech and simultaneously recorded EGG signals from 52 speakers (25 male and 27 female, with ages ranging from 20 to 80 years) in a single-wall sound room. The speech material consists of utterances of 12 sustained vowels, 16 sustained fricatives, an utterance counting one to ten (one utterance) with comfortable pitch, one counting one to ten with a progressive increase in loudness, singing the musical scale using 'la' and three sentences. In this study, fricative stimuli are not used. We use a negative threshold (1/6 of the maximum value [18]) on the differentiated EGG (DEGG) signal to distinguish the voiced from the unvoiced speech. The negative peaks of DEGG provide the ground truth GCIs for validation. A total of 392015 true epochs were obtained by this procedure.

We consider three state-of-the-art GCI detection algorithms viz. DPI, ZFR and SEDREAMS for our experiments. For this study, we use the standard performance measures namely identification rate (IDR), miss rate (MR), false alarm rate (FAR) and identification accuracy (IDA) or the standard deviation of error (SDE) 9, 10, 11], which are illustrated in Fig. 6 of [12]. The cutoff frequency for the high-pass filter is optimized for each of the GCI detection algorithms separately by performing a grid search from 2 Hz to 12 Hz with a step size of 1 Hz. The optimal cut-off frequencies corresponding to the highest IDR are found to be 3, 4 and 10 Hz for DPI, ZFR and SEDREAMS respectively, by optimizing on a development set comprising four speakers from the Childers' data. The remaining data from 48 speakers is used for validation in clean as well as in noisy conditions (white and babble). Noise samples are taken from Noisex-92 database [19] and SNRs of -20 to 20 dB are considered. DEGG signal is used for estimating the APP for ZFR and SEDREAMS algorithms.

We also compare our results with two naive methods of correcting the GCI errors described as follows - (i) Once the erroneous groups are found out by the method described earlier, the erroneous GCIs are corrected by forcing them to follow a linear contour obtained by fitting a line to the pitch-contour in the neighborhood of the erroneous group. (ii) smoothing the erroneous GCI contours by the median of the GCI values within the entire erroneous GCI group. Note that neither of these process involve optimization as in the case of the proposed method.

3.2. Results and discussion

Figure 3 compares the performances of the GCI detection algorithms with (involving both the proposed method and the naive schemes) and without the error correction scheme. SEDREMS is abbreviated as SED in the figure and hereafter. FAR is not shown in the figure since it can be obtained by calculating $1-IDR-MR$. In the clean condition, the improvement in the IDR, after post-processing is about 1, 4 and 0.1 % (absolute) for DPI, ZFR and SED respectively. This is expected because, in clean speech, the IDRs are already on the higher side (more than 97%), except for ZFR, for which IDR increases from 93% to 97%. In the case of additive white noise, IDR increases consistently with post-processing for DPI, ZFR, and SED at all SNRs. The scenario is similar in the case of the babble noise, but the improvement in IDR is less than that in the case of white noise. It is observed that the improvement is more at lower SNRs for both noises. It is noteworthy that, the proposed post-processing technique improves the IDR while not altering the SDE significantly. This implies that the post-processing not only rectifies the erroneous GCIs but also places the rectified GCIs closer to the ground truth. Note that the MR values in the case of DPI and SED (a range of 3 to 38 %), in additive white noise are higher than that (a range of 1.2 to 1.6 %) in the case of ZFR. However, IDR values for DPI, SED and ZFR are similar; this suggests that ZFR commits more insertion errors than DPI and SED. Following the error correction scheme, the MR drops largely in the case of DPI and SED (26 and 24% respectively) leading to improvement in their IDRs. But there is no such scope

[1] This choice is made empirically for the databases considered. However, the threshold for deciding the erroneous group can be set depending upon the data by considering the maximum extent of pitch variation in a given utterance.

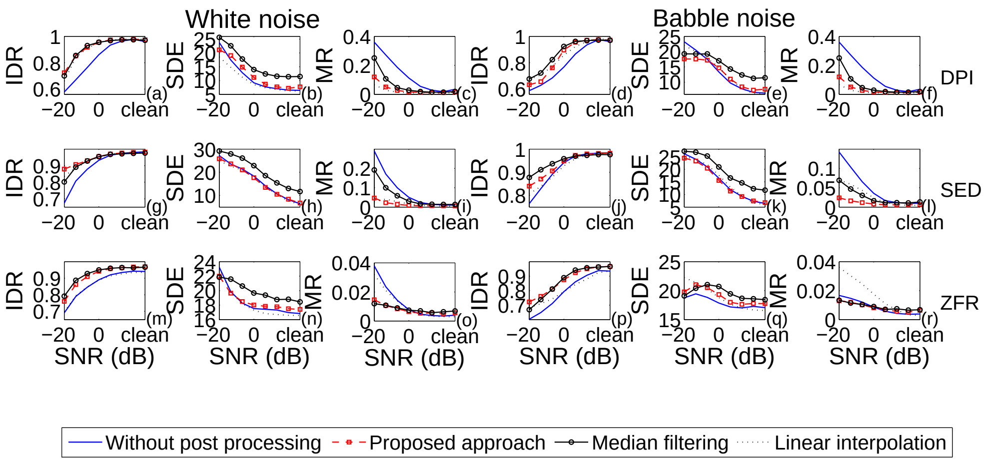

Figure 3: Results of the proposed post-processing technique on DPI, SEDREAMS and ZFR at two different noisy conditions under several SNRs. It is seen that the postprocessing always improves the IDR while SDE is maintained close to the original value. The solid blue and the red box lines in the figures denote the results without and with postprocessing, respectively. The dotted line represents the results with the GCI correction using the naive linear interpolation method. The solid black line with round beads represent the GCI correction using median filtering approach.

in the case of ZFR. The improvement in IDR for ZFR comes mainly from correcting the insertion errors. Lower improvement in IDR for ZFR than for DPI and SED at low SNR could be because the post-processing scheme rectifies more miss error than insertion error. Careful examination reveals that many of the insertion errors are not detected in the first place, while that is not the case with miss errors. This is because, in this work we are using a simple detector which detects error if an IEI is outside the 20% frequency band (a factor of 0.8 to 1.2) around the APP. Such a detector detects a miss more reliably than an insertion. For example, suppose an erroneous group of s sample duration has M correct GCIs. Thus the APP is s/M. If a GCI detection algorithm makes an insertion error the average IEI becomes $s/(M+1)$. While in the case of miss error it is $s/(M-1)$. Thus, on average, the IEI due to insertion error is less than s/M by a factor of $1/(M+1)$, while that is $1/(M-1)$ for miss error. Since $1/(M-1) > 1/(M+1)$, factor for miss error will have higher chance of being greater than 1.2 than that for insertion error being lower than 0.8. Assuming equal number of insertion and miss errors, the simple GCI error detector would fail to detect insertion errors more than miss errors. Figure 3 also depicts the results for GCI correction using naive linear interpolation method described earlier. It can be seen that for the DPI algorithm, the proposed method is marginally better than the linear interpolation method whereas for the ZFR and SEDREAMS algorithms, the improvement in IDR with the proposed method is better than that with the linear interpolation. This is may be explained by the following empirical observation: It is seen that the proposed method of GCI correction works better when the erroneous groups contain significant number of GCI errors in series (typically of order of 5-10) whereas it is comparable to the linear interpolation when the erroneous groups comprises fewer GCI errors. It is also observed that the erroneous groups arising out of errors committed by the DPI algorithm tend to contain lesser number of serial erroneous GCI locations compared to the other two algorithms for which erroneous groups contain higher number of serial GCI locations. Thus, it may be concluded that the proposed algorithm is favored over the linear interpolation when the GCI detector imposes significant number series of GCI errors. In the case of median filtering approach, the IDR is comparable or better than the proposed method for some algorithms at some SNRs. However, the SDR performance measure is consistently lower than that for the proposed method. This implies that even though a median filtering approach corrects erroneous GCIs, they offer a poorer accuracy compared to the proposed method.

It is also important to note that the genuine pitch doubling and halving may be treated as errors by the simple error detection scheme used in this work. Thus, the GCIs obtained using the error correction scheme would be erroneous in regions which contain genuine pitch doubling and halving while the original GCI detection algorithm may not incur any error. Thus, the effectiveness of the proposed error correction scheme depends on the accuracy of the erroneous IEI detection as well as the accuracy of the estimate of APP. However, if a fairly accurate estimate of the APP is known a priori using technique other than the GCI estimates, the proposed post-processing technique could have a better performance.

4. Conclusion

In this paper, we have proposed an error correction scheme for any GCI detection algorithm. The GCIs errors within erroneous groups are estimated by minimizing the high-pass energy in the corrected IEI contour. Experiments are carried out on the speech data from 52 speakers, comprising simultaneous EGG recordings, using three state-of-the-art GCI detection algorithms. It is found that the proposed post-processing technique improves the performance of the GCI detection algorithms compared to naive methods of linear interpolation of the erroneous GCIs and median filtering. The proposed error correction scheme could be made more robust by designing a better detector of the erroneous IEI regions as well as a better estimate of APP particularly at low SNRs.

5. References

[1] D. Wong, J. Markel, and A. Gray Jr, "Least squares glottal inverse filtering from the acoustic speech waveform," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, no. 4, pp. 350–355, 1979.

[2] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech communication*, vol. 9, no. 5, pp. 453–467, 1990.

[3] M. R. Shanker, R. Muralishankar, and A. G. Ramakrishnan, "Bauer method of mvdr spectral factorization for pitch modification in the source domain," in *accepted, IEEE Workshop on ASPAA07*, 2007, pp. 21–24.

[4] R. Muralishankar, M. Ravi Shanker, and A. G. Ramakrishnan, "Perceptual-mvdr based analysis-synthesis of pitch synchronous frames for pitch modification," in *Multimedia and Expo, 2008 IEEE International Conference on*. IEEE, 2008, pp. 81–84.

[5] R. Muralishankar, A. G. Ramakrishnan, and P. Prathibha, "Modification of pitch using dct in the source domain," *Speech Communication*, vol. 42, no. 2, pp. 143–154, 2004.

[6] T. V. Ananthapadmanabha, A. P. Prathosh, and A. G. Ramakrishnan, "Detection of the closure-burst transitions of stops and affricates in continuous speech using the plosion index," *J. Acoust. Soc. Am.*, vol. 135, no. 1, pp. 460–471, 2014.

[7] M. D. Plumpe, T. F. Quatieri, and D. A. Reynolds, "Modeling of the glottal flow derivative waveform with application to speaker identification," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 5, pp. 569–586, 1999.

[8] B. Yegnanarayana and S. Gangashetty, "Epoch-based analysis of speech signals," *Sadhana*, vol. 36, part 5, pp. 651–697, Oct. 2011.

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

[10] T. Drugman and T. Dutoit, "Glottal closure and opening instant detection from speech signals,," in *Proc. Interspeech Conf.*, 2009.

[11] A. P. Prathosh, T. V. Ananthapadmanabha, and A. G. Ramakrishnan, "Epoch extraction based on integrated linear prediction residual using plosion index," *IEEE Trans. on Audio, Speech, and Lang. Process.*, vol. 21, no-12, pp. 2471–2480, Dec. 2013.

[12] T. Drugman, M. Thomas, J. Gudnason, P. Naylor, and T. Dutoit, "Detection of glottal closure instants from speech signals: A quantitative review," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 3, pp. 994–1006, Mar. 2012.

[13] T. Drugman, B. Bozkurt, and T. Dutoit, "Causal–anticausal decomposition of speech using complex cepstrum for glottal source estimation," *Speech Communication*, vol. 53, no. 6, pp. 855–866, 2011.

[14] T. V. Ananthapadmanabha, "Acosutic factors determining perceived voice quality," in *Vocal fold Physiology - Voice quality control*, O.Fujimura and M. Hirano, Eds. San Diego,Cal.: Singualr publishing group, 1995, ch. 7, pp. 113–126.

[15] M. N. Vieira, J. E. de C Silva, and H. C. Yehia, "Vibrato and tremor extent spectrum: Algorithm and applications," *J. Acoust. Soc. Am.*, vol. 130, pp. EL1–EL7, 2011.

[16] J. Currie and D. I. Wilson, "OPTI: lowering the barrier between open source optimizers and the industrial matlab user," *Foundations of Computer-Aided Process Operations, Savannah, Georgia, USA*, pp. 8–11, 2012.

[17] D. G. Childers, *Speech Processing and Synthesis Toolboxes*. Wiley, Newyork, 2000.

[18] D. G. Childers and A. K. Krishnamurthy, "A critical review of electroglottography," *CRC Crit. Rev. Bioeng.*, vol. 12, pp. 131–164, 1985.

[19] Noisex-92. [Online]. Available: http://www.speech.cs.cmu.edu/comp.speech/Sectionl/Data/noisex.html