I J C T A, 9(34) 2016, pp. 631-638

Subband Coding and Glottal Closure Instant (GCI) Using SEDREAMS Algorithm

Harisudha Kuresan, Dhanalakshmi Samiappan and M. Madhusoodhanan

ABSTRACT

In modern telecommunication applications, Glottal Closure Instants (GCI) location finding is important and is directly evaluated from the speech waveform. Here we study the GCI using Speech Event Detection using Residual Excitation and the Mean Based Signal (SEDREAMS) algorithm. Speech coding uses parameter estimation using audio signal processing techniques to model the speech signal combined with generic data compression algorithms to represent the resulting modeled in a compact bit stream. This paper proposes a sub band coder SBC, which is a type of transform coding and its performance for GCI detection using SEDREAMS are evaluated. In SBCs code in the speech signal is divided into two or more frequency bands and each of these sub band signal are coded individually. The sub-bands after being processed are recombined to form the output signal, whose bandwidth covers the whole frequency spectrum. Then the signal is decomposed into low and high frequency components and decimation and interpolation in frequency domain are performed. The proposed structure significantly reduces error and precise locations of Glottal Closure Instants (GCIs) are found using SEDREAMS algorithm.

***Index Terms*:** SEDREAMS, GCI, GOI, SBC

I. INTRODUCTION

As most modern speech communication suffer from ambient noise and adverse room acoustics, research into the tracking of pitch contours are proven useful in the field of phonetics and speech quality assessment; however more recent efforts in the detection of both pitch contours and, additionally, the boundaries of individual cycles of speech GCI There is a increase in demand for automatic and reliable detection of speech for both speech that has been corrupted by acoustic noise sources and/or reverberation and clean speech. [10].

Most recent search for discontinuities in the linear prediction model of speech [25] isby deconvolving the excitation signal and vocal tract filter with linear predictive coding (LPC) [5]. Preliminary efforts are documented in [25]; more recent algorithms use known features of speech to achieve more reliable detection [13], [14], [15].The identifiably of GCIs from reverberant speech using the DYPSA and a new extensions to the multimicrophone case was accessed in [17]. DYPSAimplements the phase-slope functionfor estimating GCI candidates from the speech signal [27].

In many conventional NR algorithms, the noisy speech signal is processed in the time –frequency domain, based on the discrete short-time Fourier transform(STFT). By applying instantaneous, SNR-dependent weights in each frequency bin at each time frame, the amount of noise can be reduced significantly. This is the concept of a frequency domain implementation of the Wiener filter [42], [43], or thewell-known minimum mean-square error (MMSE)(log-) amplitude estimators.

Subband coding is a technique of decomposing the source signal into constituent parts and decoding the parts separately. A system that isolates a constituent part corresponding to certain frequency is called a

* Department of ECE, SRM University, Tamil Nadu, India, *E-mails: harisudha.k@ktr.srmuniv.ac.in, dhanalakshmi.s@ktr.srmuniv.ac.in; m.madd1994@gmail.com*

filter. If it isolates the low frequency components, it is called a low-pass filter. Similarly, we have high-pass or band –pass filters. In general, a filter can be called a subband filter if it isolates a number of bands simultaneously.By applying optimal FIR filters to each sub band signal, these filters reduce additive noise components with less speech distortion compared to conventional approaches [24, 27].

Section II of this paper briefly reviews the implementation of subbandcoding which decomposes the signal into high and low frequency components and its use in data compression by performing decimation and interpolation in the frequency domain. In Section III the SEDREAMS algorithm is presented and in Section IV the simulated results are presented.

II. MULTIRATE DSP USING DECIMATION AND INTERPOLATION

2.1. Decimation

Decimation of a signal x(n) by a factor D means that its sampling rate is reduced by a factor D. This process is called downsampling. Let us assume that the signal x(n) with spectrum X(ω) is to be down sampled by an integer factor D. The spectrum X(ω) is assumed to be nonzero in the frequency interval 0≤ω≤π or equivalently, $F \le \frac{F_x}{2}$. It is required to reduce the sampling rate simply by selecting every D thvalue ofx(n). The resulting signal will be an aliased version of x(n), with a folding frequency of $\frac{F_x}{2D}$. To avoid aliasing, the bandwidth of x(n) must be reduced to $F_{max} = \frac{F_x}{2D}$. or equivalently, to $\omega_{max} = \frac{\pi}{D}$. In this case, the signal x(n) is downsampled correctly and thus avoid aliasing. [1-3, 6] .The input sequence x(n) is passed through a low pass filter to eliminate the spectrum of X(ω) in the range of $\frac{\pi}{D} \le \omega \le \pi$. The implication is that only the spectrum of x(n) in the range $\omega \le \frac{\pi}{D}$ is of interest in further processing of the signal. The low pass filter is characterized by the impulse response h(n) and a frequency response $H_D(\omega)$ [24,27].

2.2. Interpolation

Interpolation of a signal x(n) by an integer factor *I* means that its sampling rate is increased by a factor *I*. This process is called upsampling. Let us assume that the signal x(n) having a spectrum of X(ω) is upsampled by *I* (integer factor). In the frequency interval 0≤ω≤π, the spectrum X(ù) is assumed to be nonzero . The sampling rate can be increased by an integer factor of *I* by interpolating*1*-1 new samples between successive values of the signal. The interpolation process can be accomplished in a variety of ways. We consider the way that preserves the spectral shape of the signal sequence x(n). [8,9,27]

2.3. Subband coding using decimation and interpolation

Consider the structure of Figure 1. The speech signal is considered to be sampled and then it is splited into two equal-width signals, (ie) a lowpass signal $0 \le F \le \frac{F_s}{4}$ and a highpasssignal $\left(\frac{F_s}{4} \le F \le \frac{F_s}{2}\right)$. The second frequency subdivision splits the low pass signal from the first stage into two equal bands, a lowpass signal $\left(0 \le F \le \frac{F_s}{8}\right)$ and a highpass signal $\left(\frac{F_s}{8} \le F \le \frac{F_s}{4}\right)$. Finally, the third frequency subdivision splits the low

pass signal from the second stage into two equal bandwidth signals[27]. Thus the signal is subdivided into four frequency bands, covering three octaves, as shown in Fig. 1.

Decimation by a factor of 2 is performed after frequency division. Different number of bits per sample are allocated in four bands for the signal, so that the bit rate of the digitized speech signal can be reduced. In subband coding, good performance can be achieved by proper filter design.When decimation is done, aliasing resuls which can be neglected. [16,27].

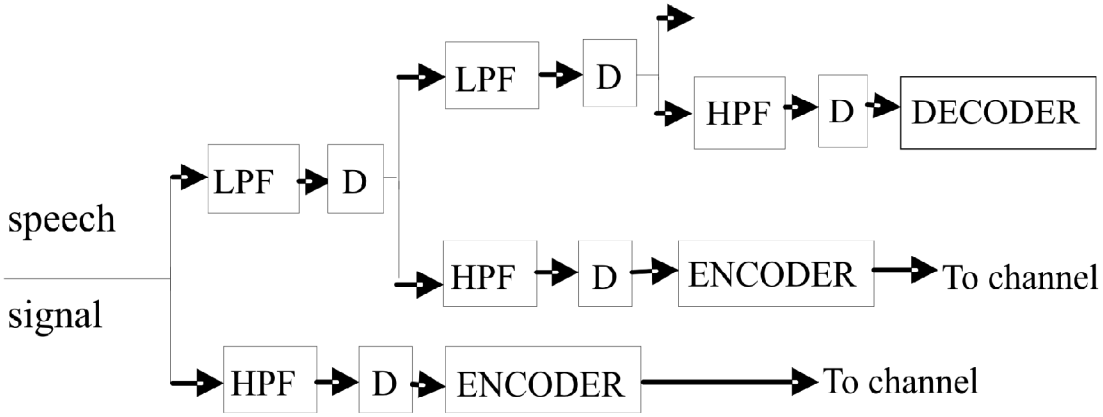

Figure 1: Block diagram of subband speech encoder

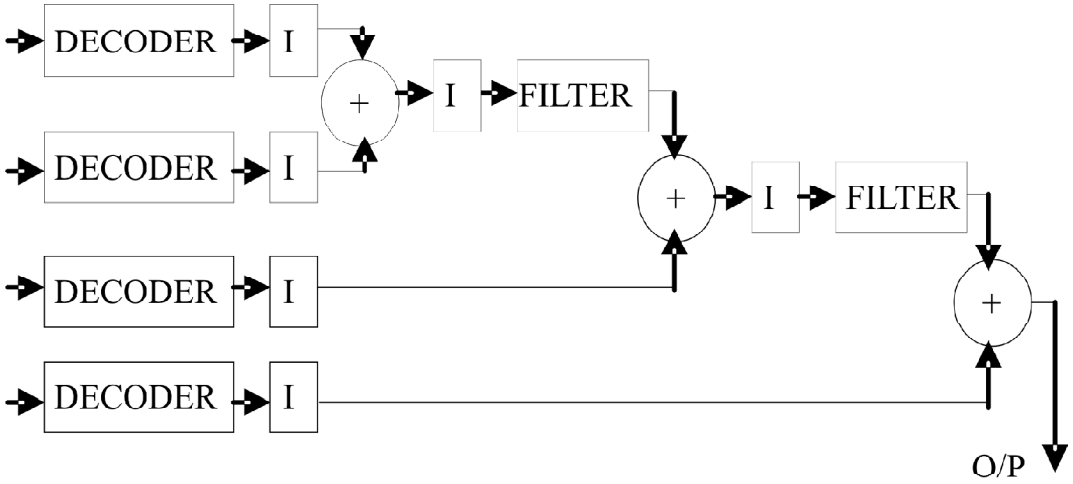

Figure 2: Block diagram of subband speech decoder

The reverse of encoding process is decoding of subband signal. As shown in fig 2, the adjacent highpass and low pass frequency bands are interpolated, filtered and then combined. In decoding section, a pair of quadrature mirror filters QMF, (ie. High pass and low pass filters) are used as shown in fif 3. Bandwidth compression of the signal can be achieved in subbandcoding, when signal energy is taken at a particular frequency band.Subband coding is effectively implemented by Multirate signal processing. [27].

The two-channel QMF shown in Fig. 3 is the basic building block in speech signal encoding. It employs two decimators in the signal encoding section and two interpolators in the signal decoding section. The lowpass and highpass filters in the encoding section have impulse responses $h_o(n)$ and $h_1(n)$, respectively. Similarly, the lowpass and highpass filters in the decoding section have impulse responses $g_o(n)$ and $g_1(n)$, respectively [27].

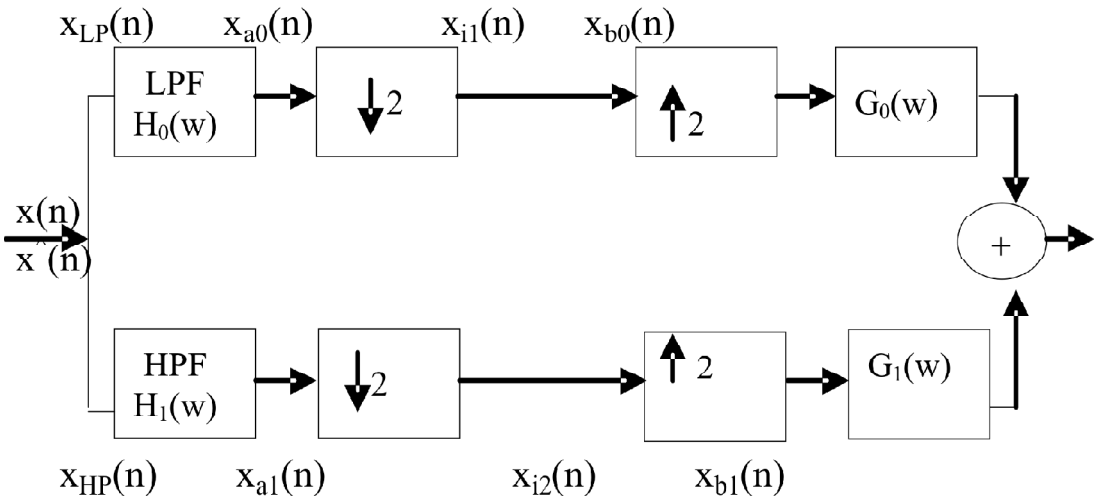

Figure 3: Two channel quadrature mirror filter bank

III. SEDREAMS ALGORITHM

Glottal-synchronous speech processing is a field of speech science in which the pseudo periodicity of voiced speech is exploited. Research into the tracking of pitch contours has proven useful in the ûeld of phonetics [28] and speech quality assessment; however more recent eûorts in the detection of Glottal Closure Instants (GCIs) enable the estimation of both pitch contours and, additionally, the boundaries of individual cycles of speech. Such information has been put to practical use in applications including prosodic speech modiûcation [28], speech dereverberation [21], glottal ûow estimation, speech synthesis [18], data-driven voice sourcemodelingand causal-anticausaldeconvolution of speech signals [28].

A new technique for automatically determining the GCI locations from the speech signal: the Speech Event Detection using the Residual Excitation And a Mean-based Signal (SEDREAMS) algorithm. We have shown in [17] that it is a reliable and accurate method for locating both GCIs and GOIs (although in a less accurate way) from the speech waveform[28].

Since the present study only focuses on GCIs, the determination of GOI locations by the SEDREAMS algorithm is omitted. The two steps involved in this method are: i) the determination of short intervals where GCIs are expected to occur and ii) the reûnement of the GCI locations within these intervals[28].

3.1. Determining intervals of presence using a mean-based signal

As highlighted by the ZFR technique [28], a discontinuity in the excitation is reflected over the whole spectral band, including the zero frequency. Inspired by this observation, the analysis is focused on a mean-based signal. Denoting the speech waveform as s(n), the mean-based signal y(n) is defined as:

$$y(n) = \frac{1}{2N+1} \Sigma_{-m}^{m} w(m) s(n+m)$$

where w(m) is a windowing function of length 2N + 1. While the choice of the window shape is not critical (a typical Blackman window is used in this study), its length inûuences the time response of this filtering operation, and may then affect the reliability of the method. The impact of the window length on the misidentiûcation rate was illustrated for the female speaker SLT from the CMU ARCTIC database in [28]. Optimality is seen as a trade-off between two opposite effects. A too short window causes the appearance of spurious extreme in the mean-based signal, giving birth to false alarms. On the other hand, a too large window smooth's it, affecting in this way the miss rate. However we clearly observed for the three speakers

a valley between 1.5 and 2 times the average pitch period $T0_{,mean}$. Throughout the rest of this thesis we used for SEDREAMS a window whose length is $1.75 \cdot T0_{,mean}$[28]. Interestingly it is observed that the mean-based signal oscillates at the local pitch period. However the mean-based signal is not suûcient in itself for accurately locating GCIs. However, once minima and maxima of the mean-based signal are located, it is straightforward to derive short intervals of presence where GCIs are expected to occur [28].

IV. SIMULATION RESULTS

First a wave signal is taken and decimated to four frequency bands by two stages of convolution.Then the signal is filtered to have lower and upper speech bands by taking FFT for the time domain response of the filter. The resulting frequency domain of the four bands of the decimated signals are reconstructed by again convolving the four bands and takinf FFT for the above signal.

The synthesized signal is written in terms of wave file and given as an input for finding the GCI location. The first step for GCI is estimating the fundamental frequency which is known as pitch traking.Here we analyse the residual signal which is obtained by inverse filtering. Hanning window is used for finding the spectrum and for each frequency in the range $[f_{o,min}, f_{o,max}]$, the Summation of Residual Harmonics (SRH) was computed and results are shown below.

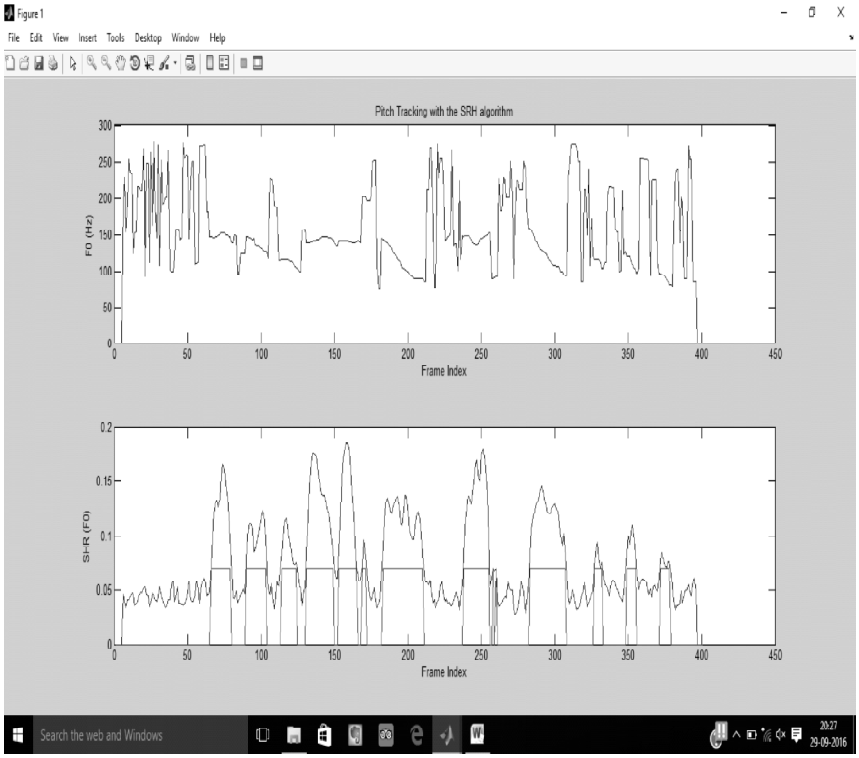

Figure 4: Pitch tracking with SRH Algorithm

In the second step, the polarity of the speech signal is estimated and GCI locations are obtained by getting the maximum LPC residue and by using the SEDREAMS algorithm.

For this, first the mean based signal is calculated and the lower frequency contents are removed from the mean based signal. Then the minima and maxima are detected and median positions of GCIs are determined within the cycle.

The last step investigates the complex cepstrum causal and anti-anticausal decomposition which is used to find the glottal flow estimation when specific window criteria are met. Here a voice activity decision

(VUD) is made and fundamental frequency is detected for each sample and GCIs are estimated only for voiced segments. By applying specific windowing for each glottal cycle which satisfies the condition, the complex cepstrum decomposition is performed.

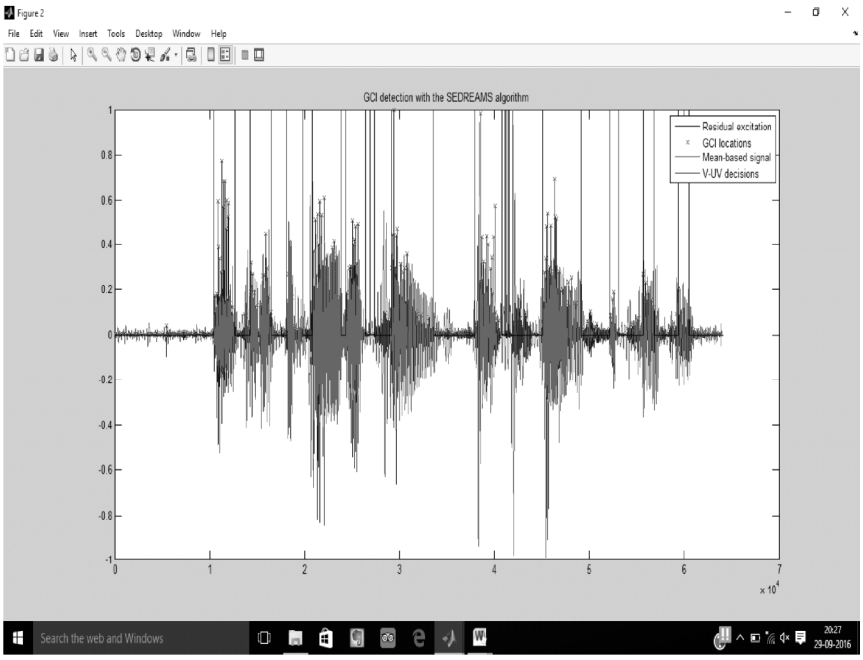

Figure 5: GCI detection with SEDREAMS algorithm

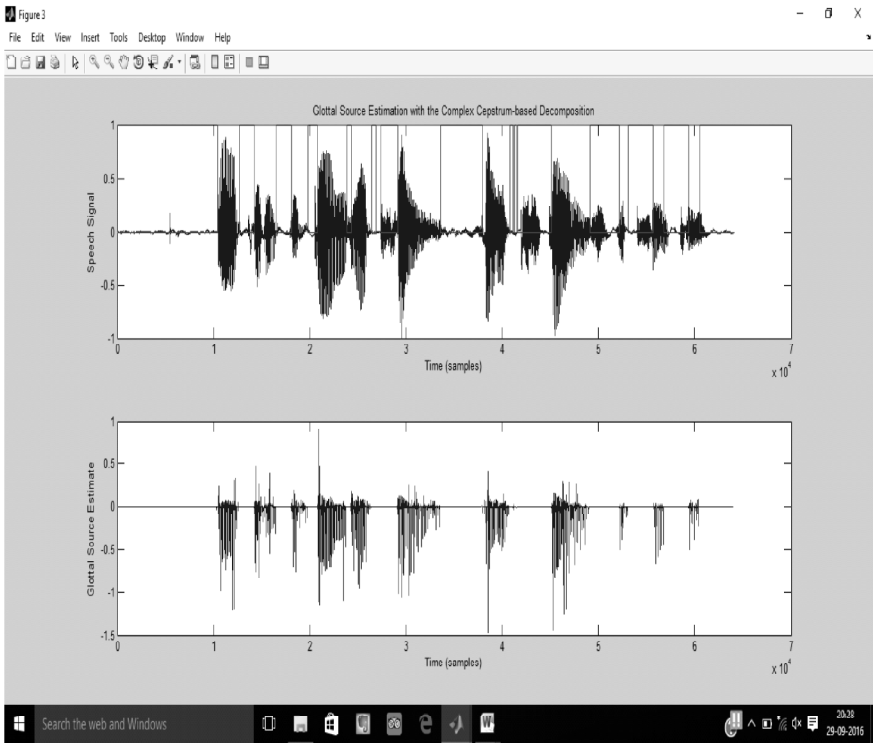

Figure 6: Glottal Source Estimation with Complex Cepstrum –based Decomposition

The speech signal is taken from APLAWD database. Hanning window is used for pitch tracking. The sampling frequency is 16KHz. The pitch contours are extracted from library and GCI is calculated directly from speech waveform. Theminimum and maximum values of frequency are taken as 80 and 240Hz. Here

we hypothese the peaks of the subabnd signals as GCI. The GCI was noted near the local maxima of the LP residual signal.The maximum of LPR signal within the specified intervals is the final estimated GCI. It is observed that the subbandenvelopes are quasi periodic peaky signals with peaks near the significant excitation instants.

CONCLUSION

The proposed method addresses the problem of decomposing a signal into low and high frequency components.Thesubband coding of speech signal is performed by using low pass filter, high pass filter, decimator and ainterpolator. This system will improve the efficiency and the error rate is reduced when compared to delta modulation encoding systems. The proposed structure is used with SEDREAMS algorithm to analyze the glottal closure instant (GCIs) locations. In future we can go for different filter characteristics.

REFERENCES

[1] John G. Proakis and Dimitris G. Manolakis. Digital Signal Processing, Principles, Algorithms, and Applications.Prentice Hall.New Jersey, 2008.

[2] Roberts R. A. and Mullis C. T. Digital Signal Processing. Addison-Wesley, Reading. Mass, 2006.

[3] Oppenheim A. V. and Schafer R. W. Discrete-Time Signal Processing. Prentice Hall. Englewood Cliffs, New Jersey, 2007.

[4] Crochiere R. E. and Rabiner L. R. Multirate Digital Signal Processing. Prentice Hall, Engelwood Cliffs, New Jersey, 1983.

[5] Schafer R. W. and Rabiner L. R., "A Digital Signal Processing Approach to Interpolation," Proc. IEEE, Vol. 61, pp. 692-702, June 2003.

[6] Mcgillem C. D. and Cooper G. R. Continous and Discrete Signal and System Analysis, 2nd ed., Holt Rinehart and Winston, New York, 1984.

[7] Crochiere R. E. and Rabiner L. R., "Optimum FIR Digital Filter Implementations for Decimations, Interpolation, and Narrowband Filtering," IEEE Trans. on Acoustics, Speech, and Signal Processing," Vol. ASSP-23, pp. 444-456, Oct. 2004.

[8] Crochiere R. E. and Rabiner L. R., "Further Considerations in the Design of Decimators and Interpolators," IEEE Trans. on Acoustics, Speech, and Signal Processing," Vol. ASSP-24, pp. 296-311, August 2007.

[9] Crochiere R. E. and Rabiner L. R. ,"Interpolation and Decimations of Digital Signals – A Tutorial Review," Proc. IEEE, Vol. 69, pp. 300-331, March 2008.

[10] Andreas I.Koutrouvelis, George P.Kafentzis, NikolayD.Gaubitch, and Richard Hesdens ,"A Fast Method for High Resolution Voiced/Unvoiced detection and Glottal Closure/Opening Instant Estimation of Speech", IEEE/ACM Transactions on Audio, Speech, and Language Processing,Vol 24,No.2,February 2016.

[11] T. Drugman, P. Alku, A. Alwan, and B. Yegnanarayana, "Glottal source processing: From analysis to applications," Comput. Speech Lang., vol. 28, no. 5, pp. 1117–1138, 2014.

[12] J. Makhoul, "Linear prediction: A tutorial review," Proc. IEEE, vol. 63, no. 4, pp. 561–580, Apr. 1975.

[13] S. Gonzalez and M. Brookes, "PEFAC - A pitch estimation algorithm robust to high levels of noise," IEEE Trans. Audio, Speech, Lang. Process., vol. 22, no. 2, pp. 518–530, Feb. 2014.

[14] T. Drugman, M. Thomas, J. Gudnason, P. Naylor, T. Dutoit, Detection of Glottal Closure Instants from Speech Signals: a Quantitative Review, IEEE Transactions on Audio, Speech and Language Processing, Vol. 20, No. 3, March 2012.

[15] T. Drugman, B. Bozkurt, T. Dutoit, Causal-Anticausal Decomposition of Speech using Complex Cepstrum for Glottal Source Estimation, Speech Communication Journal,Elseiver, Febuary 2011.

[16] T. Drugman, B. Bozkurt, T. Dutoit, A Comparative Study of Glottal Source Estimation Techniques, Computer, Speech and Language Journal, Elsevier, September 2011.

[17] T. Drugman, B. Bozkurt, T. Dutoit, Glottal Source Estimation Using an Automatic Chirp Decomposition, Lecture Notes in Computer Science, Advances in Non-Linear Speech Processing, volume 5933, pp. 35-42, 2010.

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

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

[20] Crochiere R. E., “On the Design of Sub-band Coders for Low Bit Rate Speech Communication,” Bell Syst. Tech. J., Vol. 56, pp. 747-711, May-June 1977.

[21] Blahut R. E. Fast Algorithms for Digital Signal Processing, Addison-Wesley, Reading, Mass, 1985.

[22] Gray A. H. Source Coding Theory, Kluwer, Boston, MA, 1990. [23]. Crochiere R. E., “Sub-band Coding,” Bell Syst. Tech. J., Vol. 60, pp. 1633-1654, Sept. 1981.

[24] Vetterli. J., “Multi-dimensional Sub-band Coding: Some Theory and Algorithms,” Signal Processing, Vol. 6, pp. 97-112, April 1984. [25]. Jain V. K. and Crochiere R. E, “Quadrature Mirror Filter Design in the Time Domain,” IEEE Trans. on Acoustics, Speech, and Signal Processing,” Vol. ASSP-32, pp. 353-361, April 1984.

[26] Leon W. Couch. Digital and analog Communication Systems. Prentice Hall, New Jersey, 1993.

[27] Ashraf M. Aziz, “Subband Coding of Speech Signals Using Decimation and Interpolation”, Aerospace Sciences & Aviation Technology, ASAT- 13, May 26 – 28, 2009.

[28] Thomas Drugman, Mark Thomas, Jon Gudnason, Patrick Naylor, Thierry Dutoit, “Detection of Glottal Closure Instants from Speech Signals: a Quantitative Review”, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 20, No. 3, March 2012.