iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 Issue: VIII Month of publication: August 2017

DOI: http://doi.org/10.22214/ijraset.2017.8337

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, August 2017- Available at www.ijraset.com

Glottal Analysis Using Speech signals

Korutla Sudhir Sai[1], .Polasi Phani Kumar[2]
[1](Studying M.Tech. in TELEMATICS specialization, Velagapudi Ramakrishna Siddhartha Engineering College, Jawaharlal Nehru Technological University -Kakinada, Kanuru, Vijayawada-7, Andhra Pradesh)
[2](Associate Professor, Electronics and Communication Engineering department, Velagapudi Ramakrishna Siddhartha Engineering College, Jawaharlal Nehru Technological University -Kakinada, Kanuru, Vijayawada-7, Andhra Pradesh)

Abstract: Speech processing applications became the part in our daily life. By offering solutions to companies seeking for efficiency enhancement in the field of speech technology. This work involves the advancement in glottal analysis in order to bring down new techniques within the speech processing advancements. While current systems are depends on information related to the vocal tract configuration, the air flowing through the vocal tracts, are called glottal flow, is expected to exhibit a relevant similarity. Unfortunately, glottal analysis from speech recordings requires specific complex processing operations, this explains that it has been generally prevented. This to provide new advancement in glottal analysis so as to popularize it in speech processing. First, new techniques for glottal excitation estimation and modeling are proposed and shown to excite other state-of-the-art approaches on large state of real speech. Moreover, designed methods are integrated within several speech processing applications: speech synthesis, voice pathology detection, speaker recognition and expressive speech analysis. They are shown to lead to a substantial improvement when compared to other existing techniques. It covers three separate but interconnected parts. In the primary part, new algorithms for robust pitch tracking and for automatic detection of glottal closure instants are developed. This step is necessary as accurate glottal analysis prepare to perform pitch-synchronous speech frames. In the secondary part, a new non-parametric method based on Complex Cepstrum is proposed for glottal flow estimation. In addition, a way to achieve this decomposition is investigated. This comparative study of glottal flow estimation possibilities are also given. The random samples of voiced speech can be passed into several speech processing advancements. This requires the precise locations of the Glottal Closure Instants (GCIs) are available. Thus detection of GCIs are directly taken place from the speech waveform. A new procedure to determine GCIs, called the Speech Event Detection using the Residual Excitation And a Mean-based Signal (SEDREAMS) algorithm, is Identified. The usefulness of glottal information for voice pathology detection and expressive speech analysis is detected. In the another part, a new excitation modeling called Deterministic plus Stochastic Model of the residual signal is proposed. This model is to provide quality of the outspoken speech. The glottal analysis obtained from this method is useful for identification of speech signal characteristics.

Keywords: Speech Processing, Pitch Estimation ,Speech Analysis, Speech Synthesis, Speaker Recognition, Speech, Glottal flow, Source-tract Separation, Pitch Estimation, Glottal Closure Instant location detection.

I. INTRODUCTION

In speech processing, Glottal Closure Instants (GCIs) are examined to the objectives of significant excitation of the vocal speech. These exact time events correspond to the directions of high energy in the glottal signal during pronounced speech. Knowing the GCI location of particular period in speech processing. For speech path, closed-phase LP autoregressive analysis process have been implemented for further estimation of prediction coefficients, which results in a better analysis of the vocal tract responses. These methods exactly require the detection of GCIs. The wide Range of applications are exactly predicted and these possibilities are located. In additive speech, it is well known that path of a reference instant is mandatory to delete concatenation discontinuities.

GCI has also been used for voice transformation

- Voice quality enhancement
- Speaker identification
- glottal source estimation, speech coding and transmission

Many methods have been examined to predict the GCIs directly from speech waveforms. The earliest attempts depend on the mean of the auto covariance set . A study of the use of the Linear Prediction (LP) residual was identified in .Further, as GCIs correspond to instants of significant excitation, it is assumed that a large value in the LP residual is informative about the GCI location.. An approach based on a weighted nonlinear prediction was proposed in an algorithm based on a wavelet demodulation was performed. Some techniques also deploy the phase properties due to the impulse-like nature at the GCI by computing a group delay function . The DYPSA algorithm, ,enables to retain the most likely part of the speech samples . In GCIs are located by the center-of-gravity based signal and then refined by using minimum-phase group delay functions derived from the amplitude spectrum. More recently, an approach is used to detect discontinuities in frequency by limiting the analysis around a single frequency. Pitch tracking involves the estimation of contours with average fundamental frequency for voiced tracts. Such a system is of particular interest in several applications of speech processing, such as speech coding, analysis, synthesis or recognition. Pitch tracking method Defaming the Overtones contained in the spectrum of the residual signal. The idea of using a summation of harmonics for detecting the fundamental frequency is not new the use of a sub harmonic summation so as to verify for the process of unreal pitch. This approach was inspired by the use of spectral and cepstral comb filters the use of the Sub harmonic-to-Harmonic Ratio for estimating the pitch frequency and for voice quality analysis.

Glottal Closure Instants (GCI) location Detection is important and is directly evaluated from the speech waveform. The Examination of GCI using Speech Event Detection using Residual Excitation and the Mean Based Signal (SEDREAMS) algorithm. Speech coding enables parameter estimation by audio signal processing techniques to design speech based algorithms in order to represent the Sub band coding and Glottal Closure Instant (GCI) by the means of SEDREAMS algorithm The proposed structure significantly Decrease the error and exact locations of Glottal Closure Instants (GCIs) are found using SEDREAMS algorithm.

GCIs correspond to the positive zero-crossings of a filtered signal obtained by successive mixing of the speech waveform and followed by a mean removal operation. Comparative studies of the most popular approaches were performed. It was shown that the DYPSA algorithm and the technique proposed to clear outperform other state-of-the art models. In speech analysis, complex cepstrum is usually taken to de convolve the speech signal into a periodic pulse train and the vocal system impulse response. The Complex Cepstrum (CC) based technique is highly accurate than ZZT-based method. Both approaches are Examined as two similar ways to reach the same goal of separating the maxima and minima-phase contributions from a signal Z-transform. The impact of windowing, which plays a crucial role on the decomposition quality.

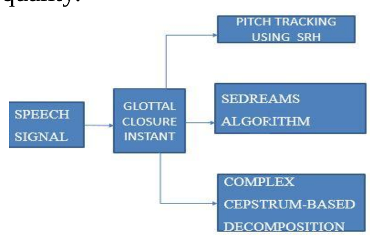


Fig. 1 Estimation of the GLOTTAL analysis

II. PITCH TRACKING WITH SRH METHOD

The pitch tracking method depends on the examination of the remaining signal. For this, an ARMA of the spectral path is examined by the speech signal s(t) and the residual signal r(t) is processed by inverse filtering. The whitening process has been processed and eliminated both the noise and the vocal tract vibrations. For each Hamming-windowed frame, intact various paths of the removing residual signal r(t), the amplitude spectrum R(f) is computed. R(f) has been additively flat top envelope for voiced approach of waveform, presents peaks at the overtone which tends to combine with the fundamental frequency F0. With this spectrum, each frequency in the range [F0,min, F0,max],the Summation of Residual Harmonics (SRH) is defined as:

$$SRH(f) = R(f) + \sum_{k=2}^{N_{harm}} [R(k \cdot f) - R((k - \frac{1}{2}) \cdot f)].$$

In path of Reference term R(k · f) in the product exponential, this equation takes the contribution of the Several harmonics taken into consideration. It is supposed to propose this expression reaches a maximum for f = F0. Morely, this is true for the harmonics present in the periodic Frequency range [F0,min, F0,max]. For this reason, the decimation by R((k− 1/2) · f) allows to significantly change the sudden depletion of the maxima of SRH at the even harmonics. The estimated tracking pitch value F0 for a given remaining frame

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, August 2017- Available at www.ijraset.com

is proportionally enlarged SRH(f) at particular time. Despite it may not work well for low-pitched waveforms for which the third harmonic may be examined in the initial range [F0,min, F0,max]. Although the successive path to incorporate a correction in Equation by decimating the term in R((k ± 1/3) · f), no evaluative correction was observed .For this reason, the proposed algorithm is in detailed in several steps. In the primary step, the proposed process is performed using the full range [F0,min, F0,max], from which the mean ,pitch frequency F0,mean of the examined speaker is calculated. In the secondary step, the final pitch tracking is evaluated by applying the same process but in the range [0.5 · F0,mean; 2 · F0,mean]. Atmostly assumed that a normal speaker will not pass these cross range. Prediction of restricting the range of F0 for a given speaker is similar to what has been proposed in (for the choice of the window length

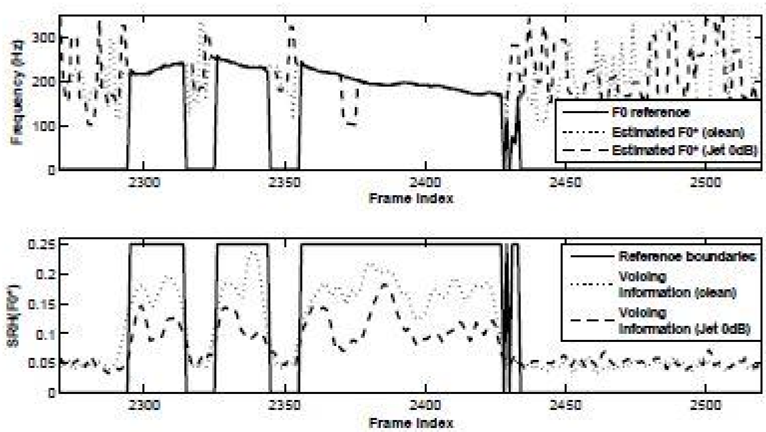


Fig.2Illustration of the proposed method in clean and noisy speech (using a jet noise with a SNR of 0 dB). Top plot : The pitch ground truth and the estimates F_0 . Bottom plot : The ideal voicing decisions and the values of SRH

The Summation of Residual Harmonics (SRH) is proposed both for pitch estimation and for the determination of voicing boundaries. SRH is shown to lead to a significant improvement, while its performance is comparable to other techniques in clean conditions.

III. GCI DETECTION WITH THE SEDREAMS ALGORITHM

The Next Step of algorithm involves the Detection of GCI using the Speech Event Detection using the Residual Excitation And a Mean-based Signal (SE-DREAMS) algorithm. This is unique and appropriate process for finding both GCIs and GOIs from the outspoken waveform. This depends on GCIs, the determination of GOI locations by the SEDREAMS algorithm is projected. Further steps involved in this process are:

The Identification of Time period where GCIs are tentatively to happen.

Further Filtering of the GCI locations within these Time periods.

This processing involves the mean-based signal. Relating the speech waveform as s(n), the mean-based signal y(n) is defined as:

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

where w(m) is a black man windowing function of length 2N +1. The relative path of the window shape is not that important to calculate its time period which is useful to find the response of this filtering operation, and which stabilizes the accuracy of this method. The solidity of speech frequency is mis interpreted by the length of filter is depicted for the female speaker. Accessibly the division ranging in the two opposite forces. A too short window causes the excitation in the mean-based signal, has given rise to false alarms. On the other path, a too large window filtering will, affect the miss rate. The Pitching of GCI is 1.5 and 2 times the mean pitch period T0,mean. we defined the SEDREAMS a window whose length is 1.75·T0,mean.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, August 2017- Available at www.ijraset.com

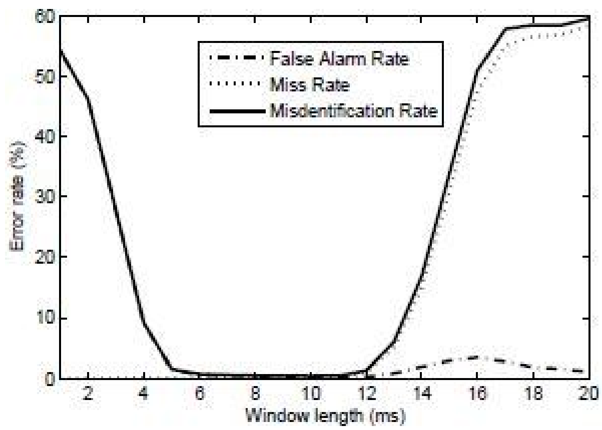


Fig. 3 Effect of the window length used by SEDREAMS on the misidentification rate for the speaker

More of a sudden it is examined that the mean-based signal lies with in the local pitch period. However the mean-based signal is not adequate to find itself for precise location of GCIs .The GCIs may happen at a non-constant relative position within the time period. Moreover, once minima and maxima of the mean-based signal are located, it is easy to derive short intervals of presence where GCIs are expected to occur more accurately

IV. GLOTTAL FLOW ESTIMATION WITH THE COMPLEX CEPSTRUM-BASED DECOMPOSITION

The Further step in algorithm involves the glottal flow estimation based on cepstrum and recognize the effective outcomes. This step, shows that complex cepstrum can be widely made to used for glottal flow estimation by dividing the causal and anti causal components of a windowed speech waveform as done by the Zeros of the Z-Transform (ZZT) decomposition. This one is similar to ZZT decomposition, windowing should be happened in such a way that windowed speech signals should portray the mixed-phase characteristics which finalize the processing of speech that the anti causal component are because of glottal flow open phase. The complex cepstrum-based approach is more advantageous and accurate when compared to ZZT decomposition.

The Decomposition of filtering components can be done by linear homo morphic filtering in the complex cepstrum domain, which involves the addition. In glottal analysis of speech signal waveform, complex cepstrum is adequately used to decompose the speech signal into a periodic pulse train and the impulse response of the specified vocal system. This method involves the pitch detection, voice coding, ,which helps in finding echo detection. Complex Cepstrum based decomposition. The complex cepstrum (CC) ^x(n) of a discrete speech signal x(n) is generally equated as

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}$$

$$\log[X(\omega)] = \log(|X(\omega)|) + j\angle X(\omega)$$

$$\hat{x}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log[X(\omega)]e^{j\omega n}d\omega$$

These Equations involves Discrete-Time Fourier Transform (DTFT), the complex logarithm and the inverse DTFT (IDTFT). This will make us to have a clear Idea that ^x(n) of an anti causal (causal) signal is zero for all n positive (negative). Remaining the negative part of the CC should be calculated by glottal contribution.one drawback involves the estimation of \X(!). This can be calculated by large number of points such that:

The phase evaluation can be easily calculated by means of unit circle in the grid.

The aliasing effect in the Speech waveform is completely overridden.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, August 2017- Available at www.ijraset.com

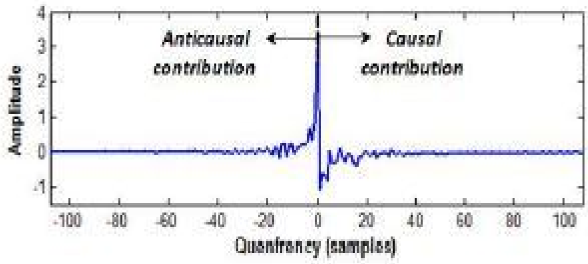


Fig.4.Corresponding complex cepstrum where the separation of maximum and minimum-phase components can be linearly achieved.

V. EVALUATION AND RESULTS

In general a Speech Signal is taken and Examined to Several steps of examination. The First step involves the pitch tracking using summation of residual Harmonics and filtered to have lower and upper speech bands by taking FFT for the time domain response of the filter. The Second step involves the resulting frequency domain by identifying the exact location of GCI using SEDREAMS 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 Third step involves the GCI for estimating the fundamental frequency which is known as pitch tracking. Here we analyze the residual signal which is obtained by glottal flow by separating the causal and anti causal signals .black window is used for finding the spectrum and for each frequency in the range [fo,min, fo,max], and the Summation of Residual Harmonics (SRH) was computed and results are shown be

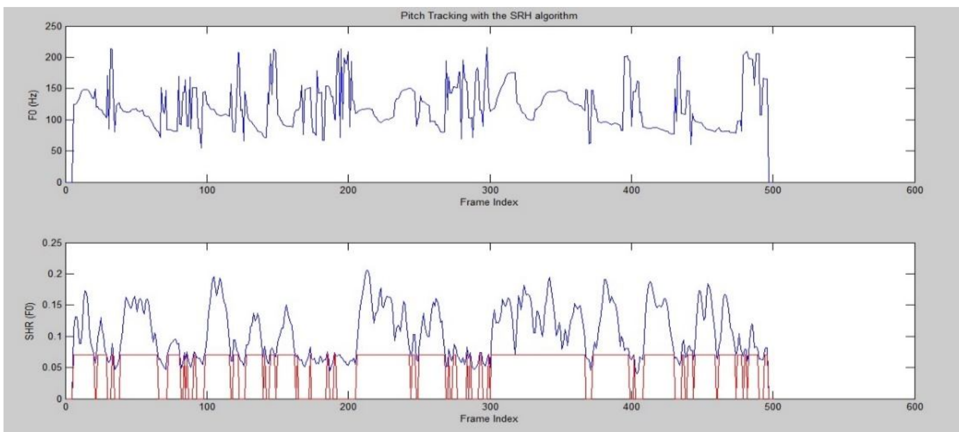


Fig. 5(a) Effect of the window length used by SEDREAMS

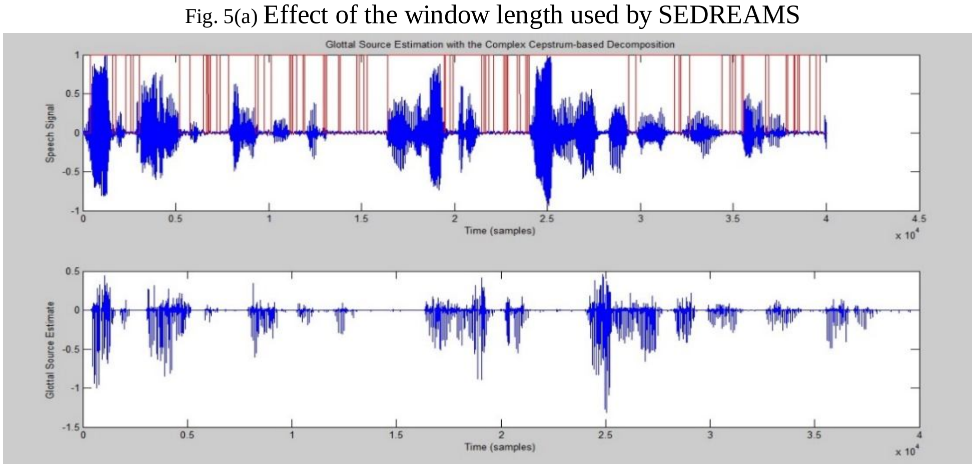


Fig. 5(b) Glottal Source Estimation with the Complex Cepstrum based Decomposition

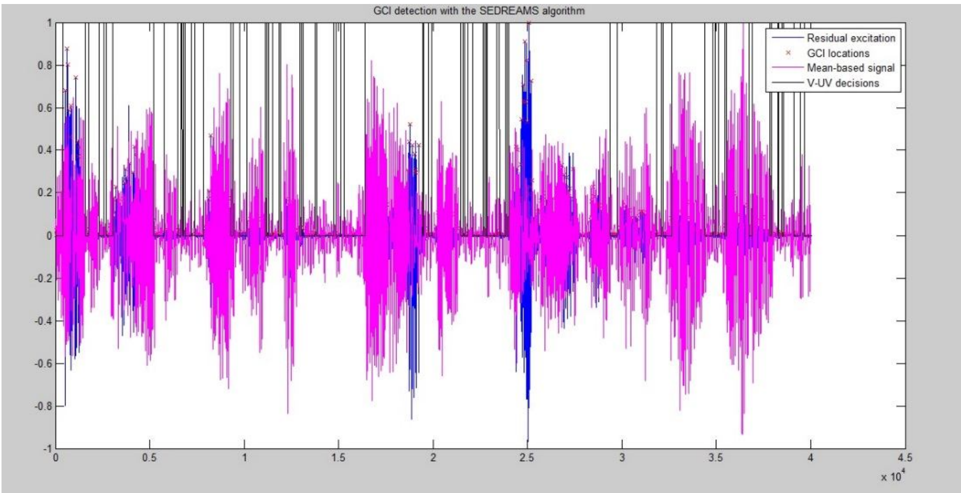


Fig. 5(c) GCI Detection with the SEDREAMS Algorithm.

WAVEFORMS	POSITIVE POLARITY WITH PROBABILTY	POSITIVE POLARITY WITH PROBABILITY	AVERAGE PROBABILITY
MALE VOCALTRACK	76.07%	75.63%	75.85%
FEMALE VOCALTRACK	74.32%	73.25%	73.78%

Table 5.(c). Performance of speaker independent vocal track system

VI. CONCLUSION

The performance of this system degrades completely with the noisy speech and hence cannot be used in real time applications like Automatic Speech Recognition (ASR)

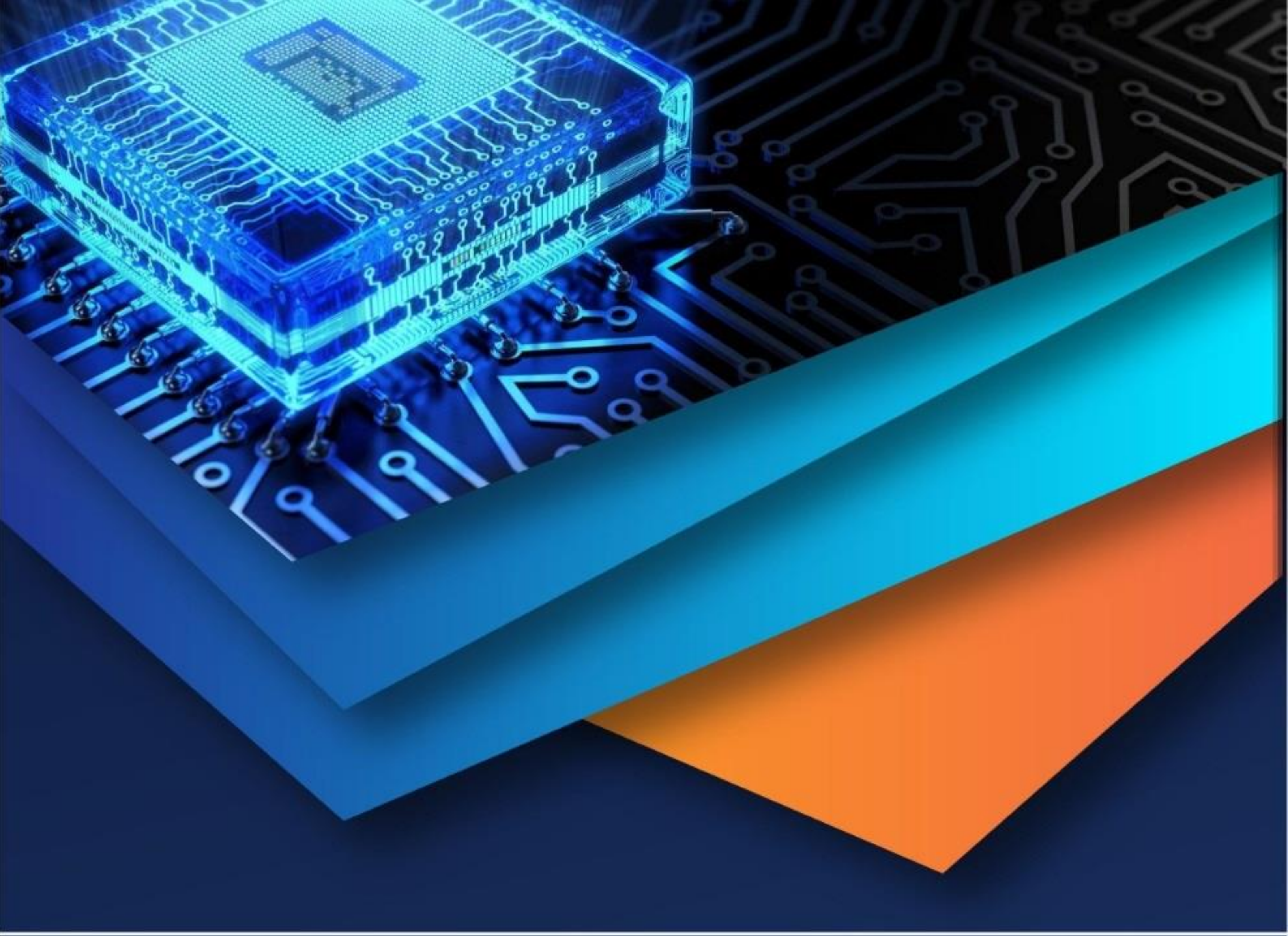
A. This system will improve the efficiency and the error rate is reduced when compared to delta modulation encoding systems. The proposed structure is used with SEDREAMS algorithm to analyze the glottal closure instant (GCIs) locations.

B. We have not included the mapping functions corresponding to source characteristics (shape of the glottal pulse), duration patterns and other speaker-specific features for applying the complex cepstrum for mixed-phase decomposition, allowing the estimation of the glottal source. The importance of a suited windowing has been highlighted. Interestingly some

C. The significant differences between the voice qualities were observed in the excitation.

REFERENCES

[11 Schafer R. W. and Rabiner L. R., "A Digital Signal Processing Approach to Interpolation," Proc. IEEE, Vol. 61, pp. 692-702, June 2003.
[2] Roberts R. A. and Mullis C. T. Digital Signal Processing. Addison-Wesley, Reading. Mass, 2006..
[3] J. Makhoul, "Linear prediction: A tutorial review," Proc. IEEE, vol. 63, no. 4, pp. 561–580, Apr. 1975.
[4] 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.
[5] 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.
[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] Ashraf M. Aziz, "Subband Coding of Speech Signals Using Decimation and Interpolation", AerospaceSciences & AviationTechnology, ASAT- 13, May 26 – 28, 2009.
[9] Blahut R. E. Fast Algorithms for Digital Signal Processing, Addison-Wesley, Reading, Mass, 1985.[accessed Aug 20, 2017].
[10] Andreas I.Koutrouvelis, George P.Kafentzis, NikolayD.Gaubitch, and Richard Hesdens ,"A Fast Method for High ResolutionVoiced/Unvoiced detection and Glottal Closure/Opening Instant Estimation of Speech", IEEE/ACM Transactions onAudio, Speech, and Language Processing,Vol 24,No.2,February 2016



10.22214/IJRASET

INDEX COPERNICUS 45.98

ISRA JIF IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY