International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Significance of Intelligent Pause Detection Protocol (IPDP) Over Other Protocols used for Speech Processing

R. Chinna Rao, D. Elizabath Rani, S. Srinivasa Rao

Abstract: Analysis of human voice based on its pitch can be used in detecting the pauses. The available algorithms for pause detection could succeed to some extent and lot of scope for better performance still exists. The proposed intelligent pause detection protocol (IPDP) is the convergence of (i) calculation of Mean/RMS peak value from human voice pitch (ii) estimating the pause using MLE algorithm and (iii) optimizing the bandwidth utilization of Vocoders using DTX algorithm. The work carried projects better pause removal than the existing standard methods.

Keywords* : *Pause, Vocoders, Human Voice, Protocol, Speech Processing.

I. INTRODUCTION

There are several methods to identify the events occurred in human voice. The most commonly used method is a 3-phase representation. In phase (i), pause (P) occurs where speech is not available. In phase (ii), No Voice (N) is identified due to non vibration of Vocal cords of human being [5]. In phase (iii), Voice (V) is identified due to vibrations of vocal chords of human being and the resultant wave form is periodic in nature. Two standard methods used for this purpose are Short Time Energy (STE) and Zeros Crossing Rate (ZCR). Each method has its own limitations in defining the thresholds on temporary basis. The accuracy results obtained with these two methods i.e. Zero Crossing Rate and Short Time Energy are very less approximately coming to 65% when compared to manual approach of detecting the pauses.

In this paper, the proposed protocol detects the silence/pause part from the Human Voice sample in three states, out of which in the first state, the RMS/Mean value from Human Voice pitch is calculated which is given as the input to the second state in which the pauses are estimated using MLE algorithm and finally proceed to the third state in which the bandwidth utilization is optimized using DTX algorithm.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
R. Chinna Rao *, Research Scholar at GITAM University, Visakhapatnam, Assistant Professor in Department of ECE at MRCET, Secunderabad, Telangana, India.
Dr. D. Elizabath Rani, Professor, Department of EIE, GITAM University, Visakhapatnam, Andhrapradesh, India.
Dr. S. Srinivasa Rao, Professor, Department of ECE, MRCET, Secunderabad, Telangana, India.

Using measure of correctness parameter, the performance of the algorithms is calculated keeping manually approach of speech as a reference. The experiments are done taking different human voices of different pitches (both male and female). The result shows better Pause detection than ZCR and STE algorithms. A speech signal may get attenuated with different types of noise in addition to Gaussian noise which is nothing but background noise. Therefore all types of noise are to be properly detected and consider while detecting the pauses in the human voice.

The paper is organized as follows: In section 2 we describe the literature survey, Section 3 presents the IPDP protocol along with a brief discussion about computational complexity and calculation of performance measurement, Section 4 projects the results and section 5 is about conclusions.

II. LITERATURE SURVEY

The Modeling of speech signal with the help of suitable Probability distributions brings the better performance for many speech signal processing algorithms. In probability theory, a random variable with a Gaussian distribution is said to be normally distributed and is called a normal deviate. The probability density of the normal distribution is

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

With μ and σ^2 as mean and variance of random variable x. This distribution is also called as normal distribution.

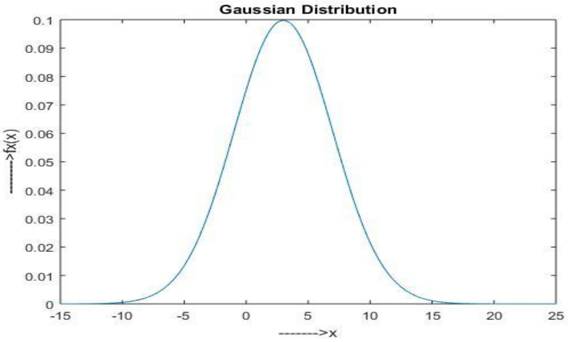

Fig.1. Gaussian Distribution Function

This probability distribution curve is a bell shaped one and symmetrical around the mean value. Increasing the mean(μ) value shifts this distribution curve to the right side and flattens when the standard deviation(σ) increases.

When pauses in the speech are not considered then the speech follows the Laplace Distribution in its time domain representation. Here in this paper the mean value and the variance are estimated using MLE Algorithm. M.Usman,etl [1] observed performance for speech processing algorithms for various distribution functions by omitting pauses. J. P. Campbell, Jr etl [2] taken hidden markov model into account for minimizing the MSE error. K. Kitayama etl [3] shows how to detect the Noise endpoint by the help of Filled Pauses. S. E. Bou-Ghazale etl[4] stated how to detect speech signal from noise environment automatically with endpoint.[5-11] shows how the signal processing algorithms detecting the speech signals from noisy environment.

III. PROPOSED METHOD

A. IPDP Protocol

The protocol is implemented in three phases as described below:

Phase 1: The mean and rms value are calculated from the human voice pitch. If μ and σ are the mean and standard deviation of human voice pitch respectively then analytically we can write,

$$\mu = \frac{1}{1600}\sum_{i=1}^{1600} x(i) \tag{2}$$

$$\sigma = \sqrt{\frac{1}{1600}\sum_{i=1}^{1600} (x(i)-\mu)^2} \tag{3}$$

Phase 2: The occurrences of pauses are estimated using Maximum Likelihood Estimation (MLE) algorithm as shown below:

$$L[\Omega, x] = P[\Omega, x] \tag{4}$$

Where L[Ω, x] is Likelihood Function, P[Ω, x] is Probability Density Function, Ω is an estimator which is a function of x x is the samples of voice from which mean and rms values of human voice are calculated using equations (2) and (3). The Maximum Likelihood Function is given by

$$L[\Omega,x] = P[x|\Omega] = p^{x}(1-p)^{1-x} \tag{5}$$

Phase 3: Using DTX algorithm, data is not transmitted during pauses thereby optimizing the bandwidth of the vocoders in the mobile applications.

The basic principle of DTX is to switch the transmission ON only for those packets when there is active speech to transmit. Effectively, it implies that during silence regions, there is no need to transmit packets. Some sequences of Silence Indicator (SID) frames are transmitted intermittently to ensure continuity of link communication.

The protocol is shown in the flow chart given in figure 2. In Zero Crossing Rate (ZCR) method, the crossing of each sample is measured in an observed slot of x samples in y ms and number of zero crossings is calculated. In Short Time Energy method, the energy of each sample sequence is calculated and then added for x samples in y ms observation time slot. Then the sums of these are calculated to classify whether it is a voiced part or not. Finally, the proposed method is compared with the conventional STE & ZCR detection methods and concluded to be used for real time analysis.

B. Performance Evaluation

The Performance Evaluation of voiced samples from human voice signal is calculated as given below:

$$\%\ \text{Improvement} = 100 - (|N_{manual} - N_{IPDP}|/N_{manual}) \times 100$$

Where, N_{manual} is the total samples from manually labelled speech; N_{IPDP} is the no. of voiced samples from a developed protocol.

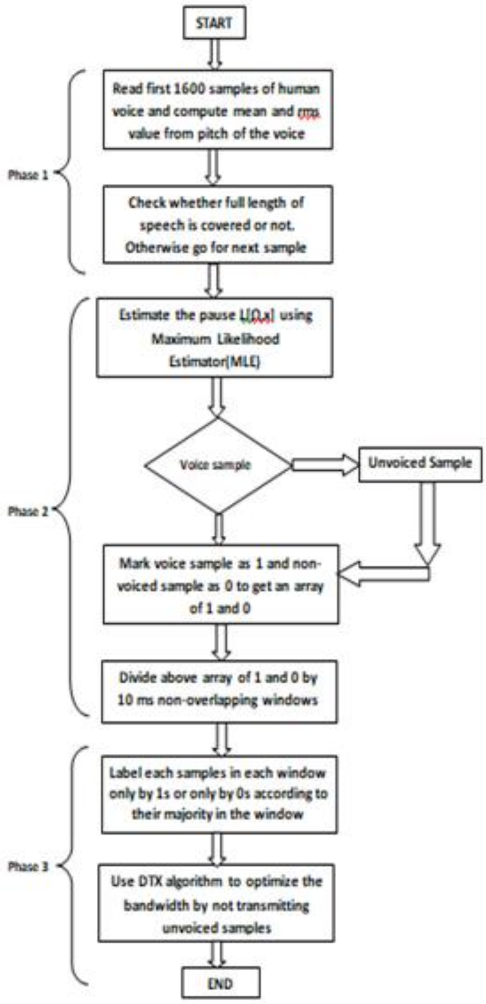

Fig. 2. Flowchart depicting IPDP Protocol

IV. RESULT AND DISCUSSION

Various male and female voices are tested using IPDP Protocol and compared with ZCR Protocol and STE Protocol which are shown below.

A. IPDP Protocol

The following results shows how IPDP protocol can be used for detecting the pauses based on human voice (both male and female voices).

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

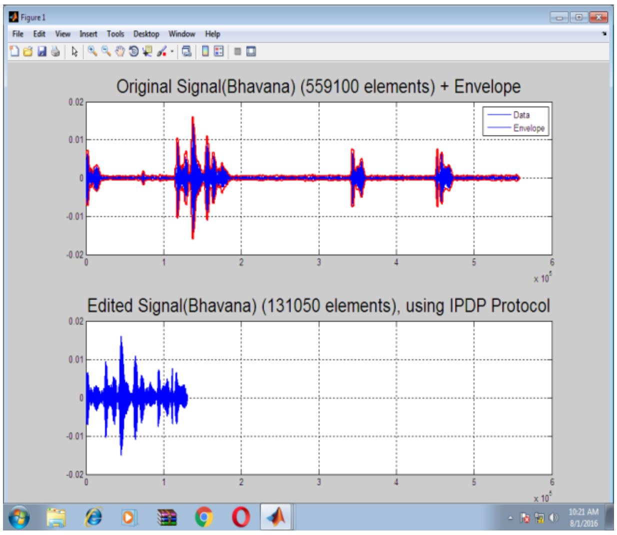

Fig. 3. Pause detection of Female Voice 1 using IPDP Protocol

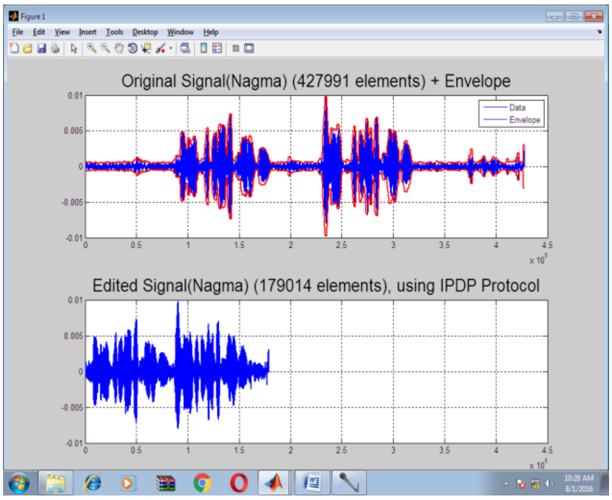

Fig.4. Pause detection of Female Voice 2 using IPDP Protocol

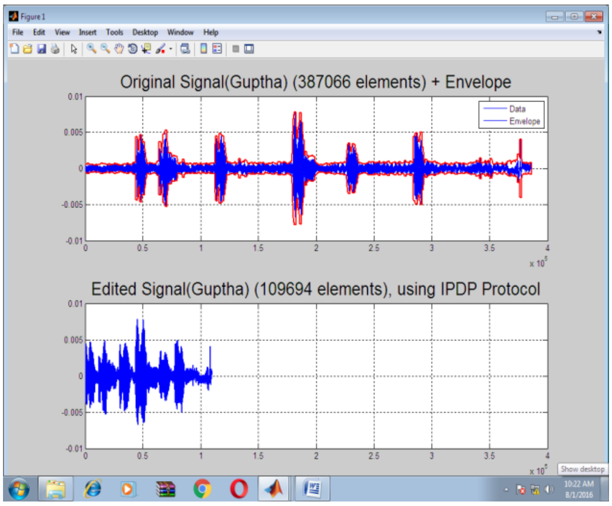

Fig. 5.Pause detection of Male Voice 1 using IPDP Protocol

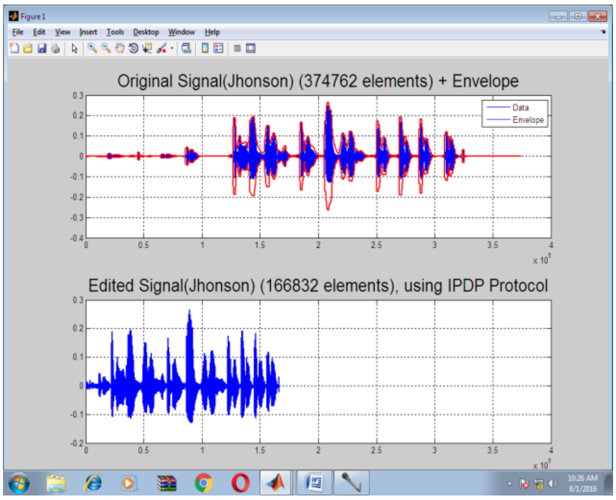

Fig. 6. Pause detection of Male Voice 2 using IPDP Protocol

B. ZCR Protocol

The following results shows how ZCR protocol detects the pauses in human voice (both male and female voices)

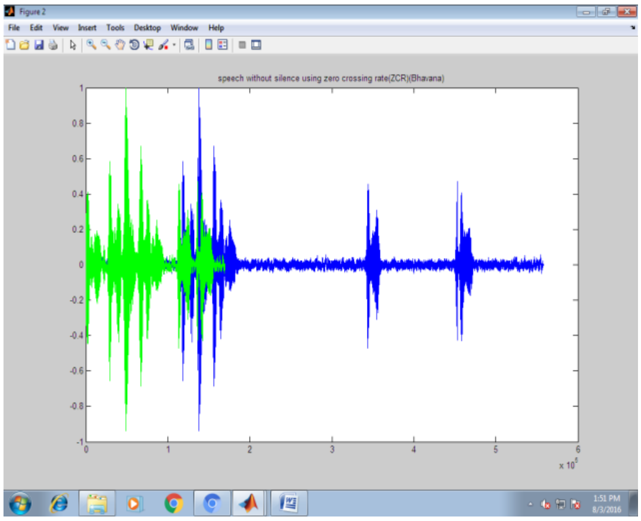

Fig. 7. Pause detection of Female Voice 1 using ZCR Protocol

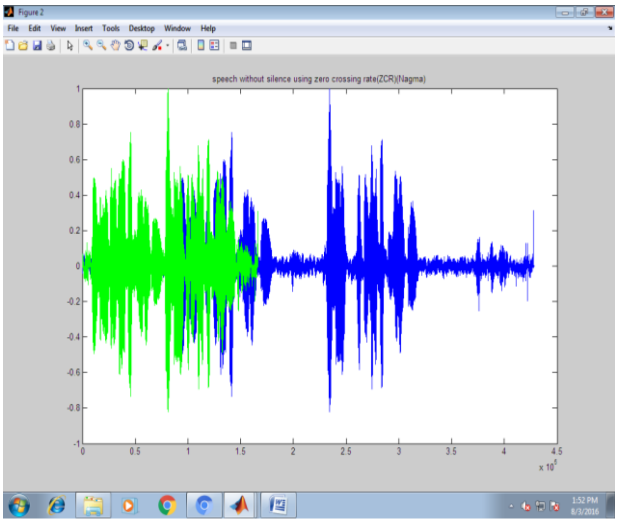

Fig. 8. Pause detection of Female Voice 2 using ZCR Protocol

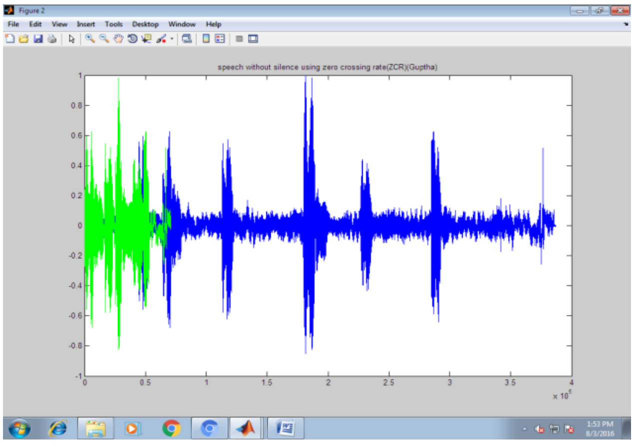

Fig. 9. Pause detection of Male Voice 1 using ZCR Protocol

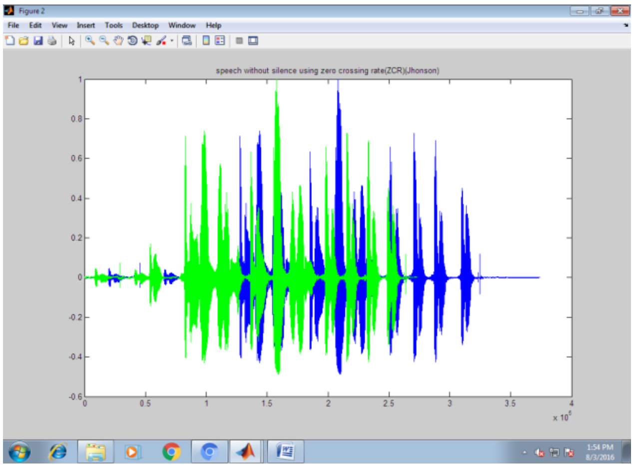

Fig. 10. Pause detection of Male Voice 2 using ZCR Protocol

C. STE Protocol

The following results shows how STE protocol detects the pauses in human voice (both male and female voices).

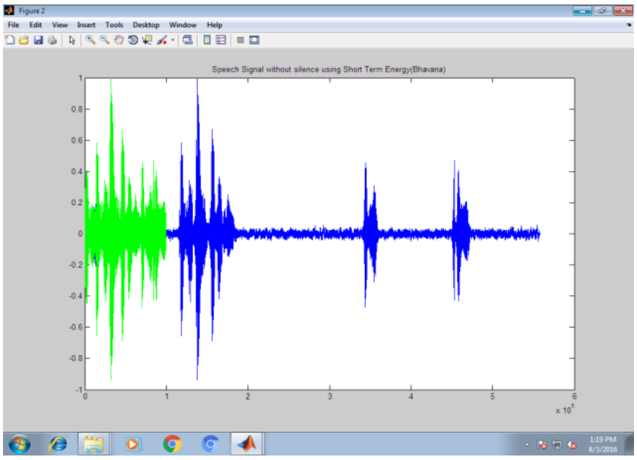

Fig.11. Pause detection of Female Voice 1 using STE Protocol

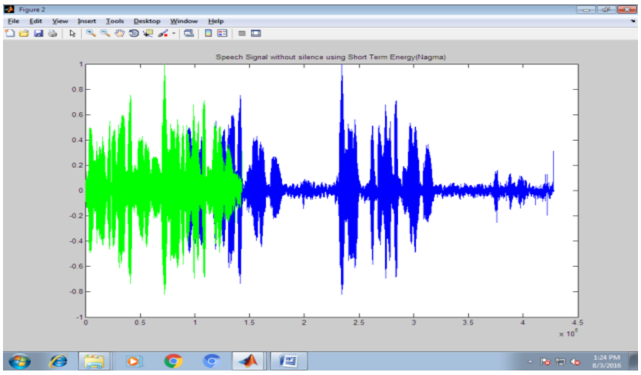

Fig 12. Pause detection of Female Voice 2 using STE Protocol

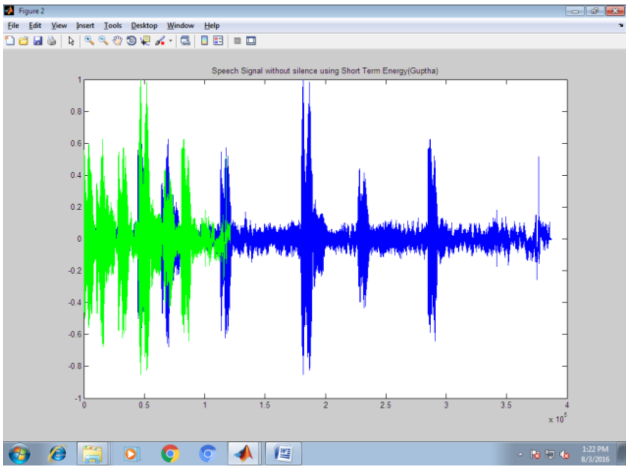

Fig.13. Pause detection of Male Voice 1 using STE Protocol

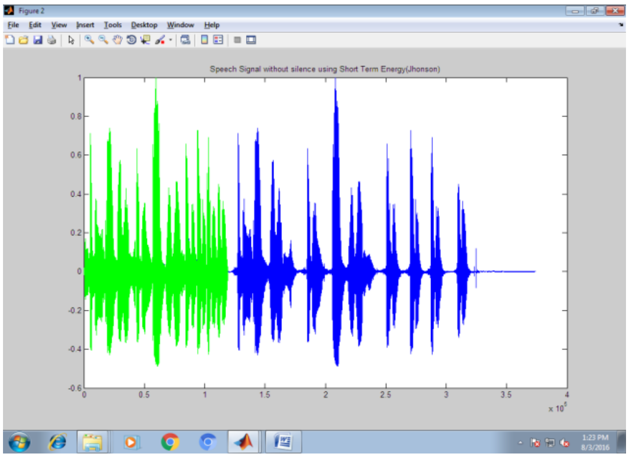

Fig. 14. Pause detection of Male Voice 2 using STE Protocol

D. Performance Evaluation of IPDP, ZCR and STE Protocols

Performance Index of IPDP using % Improvement criteria is given below

Table I: Performance Evaluation of IPDP using Percentage Improvement criteria

Phrases	STE	ZCR	IPDP
Percentage of pause Detection	77.9531%	70.3721%	89.9801%

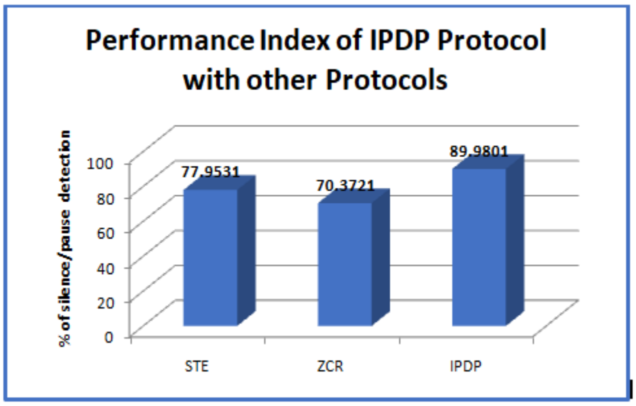

Fig. 15. Performance comparison bar graph of IPDP Protocol with other protocols

V. CONCLUSION

A new pause end point detection protocol by name Intelligent Pause Detection Protocol (IPDP) is presented. Using Intelligent Pause Detection Protocol (IPDP), various human voices are tested and their parameters are calculated. Also the performance index is calculated and compared with other existing protocols such as ZCR and STE protocols. The results prove that IPDP protocol has shown better performance in detecting silence/pause in the human voice than STE and ZCR algorithms.

REFERENCES

1. Usman, M., Zubair, M., Shiblee, M., Rodrigues, P., Jaffar, S.“Probabilistic modeling of speech in spectral domain using maximum likelihood estimation”, Symmetry, 10(12) Volume 10, Issue 12 , pp.1-15, **2018**.
2. J. P. Campbell, Jr., “Speaker Recognition: A Tutorial”, Proceedings of The IEEE, Vol.85, No.9, pp.1437-1462, Sept.1997.
3. Koji Kitayama, Masataka Goto, Katunobu Itou and Tetsunori Kobayashi,“Speech Starter: Noise-Robust Endpoint Detection by Using Filled Pauses”, Eurospeech 2003, Geneva, pp. 1237-1240.
4. S. E. Bou-Ghazale and K. Assaleh, “A robust endpoint detection of speech for noisy environments with application to automatic speech recognition”, in Proc. ICASSP2002, vol. 4, 2002, pp. 3808–3811.
5. Martin, D. Charlet, and L. Mauuary, “Robust speech / non-speech detection using LDA applied to MFCC”, in Proc. ICASSP2001}, vol. 1, 2001, pp. 237–240.
6. K. Ishizaka and J.L Flanagan, “Synthesis of voiced Sounds from a Two-Mass Model of the Vocal Chords,” Bell System Technical J., 50(6): 1233-1268, July-Aug., 1972.
7. Atal, B.; Rabiner, L., “A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition” Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on , Volume: 24 , Issue: 3 , Jun 1976, Pages: 201 - 212.
8. D. G. Childers, M. Hand, J. M. Larar, “ Silent and Voiced/Unvoied/ Mixed Excitation(Four-Way), Classification of Speech”, IEEE Transaction on ASSP, Vol-37, No-11, pp. 1771-74, Nov 1989.
9. Mark Greenwood and Andrew KInghorn, “SUVing: Automatic Silence/Unvoiced/Voiced Classification of Speech", Presented at the university of Sheffield.
10. L. Flanagan, Speech Analysis, Synthesis, and Perception, 2nd ed., Springer-Verlag, New York, 1972.
11. R Chinna Rao, Dr Elizabath Rani, Dr S Srinivasa Rao, “Basic Frame work of Vocoders for Speech Processing”, ICSCSP 2K18, June 22-23, 2018, MRCET, Secunderabad, Telangana, India.

AUTHORS PROFILE

Mr. R. Chinna Rao, received his B.Tech degree in Electronics & communication from JNT University. M.Tech from Malla Reddy College of Engineering & Technology. He is currently working as Assistant professor, Dept. of ECE in Malla reddy College of Engineering and Technology, Secunderabad, India

Dr. Elizabath Rani , received the B.Tech degree from Madurai Kamaraj University,M.Tech from Bharathiar University and Ph.D from Andhra University Visakhapatnam. Presently working as Professor and Head of the Department at Gandhi Institute of Technology and Management, Visakhapatnam. She has 31 years of experience in the field of teaching. She is a member of professional bodies like MISTE, IETE and SEMCE(I).

Dr. S.Srinivasa Rao, received the B.Tech degree from Madras Institute of Technology, Anna University, and the M.Tech and Ph.D from JNTU Hyderabad, Telangana, India. Presently working as Professor and Head of the Department at Malla Reddy College of Engineering and Technology, Secunderabad. He has 24 years of experience in the field of teaching. He is a member of professional bodies like IEEE, ISTE and IETE.