Journal of Green Engineering (JGE)

Volume-10, Issue-12, December 2020

Performance Analysis of IPDP Protocol Enabled Vocoders Using Wire Shark for Acoustic Application

[1]R Chinna Rao, [2]P.V.Y Jayasree, [3]S. Srinivasa Rao

[1]*Research scholar Department of Electronics and communication Engineering, GITAM university, Andhra Pradesh, India.*
E-mail: rayudu.chinnarao@gmail.com
[2]*Professor &HOD, Department of Electronics and communication Engineering, GITAM university, Andhra Pradesh, India.*
E-mail: jpappu@gitam.edu
[3]*Professor, Department of Electronics and communication Engineering, Malla Reddy College of Engineering & Technology, Secunderabad, India..*
E-mail: ssrao.atri@gmail.com

Abstract

In general, the operation of Vocoders is based on modeling of speech waveform's segment (or frame) on the order of 20 milliseconds. The parameters of speech model are determined, quantized, coded, and transmitted through the communication channel. To synthesize speech, the transmitted values are decoded, reconstructed, and used at the receiver. It is in this context, we provide a basic feasible solution by name Intelligent Pause Detection Protocol (IPDP) which would ensure performance enhancement of vocoders used in mobile communications. In this paper, we make an attempt to analyze IPDP protocol enabled Vocoders with the other commonly used Vocoders using Wire Shark simulator.

Keywords: Vocoders, Wire shark, Protocols, Communication channel,Acostic.

1 Introduction

In mobile communications, Vocoders are being extensively utilized in large numbers increasingly for providing better communication. In regard to the optimization of bandwidth, speech processing is one of the very essential applications where voice can be transmitted with accuracy if the programming of vocoders is done with efficient speech processing algorithms. In this connection, identifying pauses in human voice in mobile communication is very essential. The mobile manufacturers are facing this problem most commonly and they incorporate vocoders for better communications.

A robust protocol is developed in this case that helps in ensuring the reliable performance of vocoders through the reduction of bandwidth utilization during pauses occurrence in human voice which becomes a dire requirement.

In practice, various algorithms adaptability is limited to either energy utilization by human voice or zero crossing of human voice. Adaptability of various algorithms that are in practice is restricted to either zero crossing of human voice or energy utilization by human voice. Intelligent Pause Detection Protocol (IPDP) is, on the other hand, more flexible and estimates the pauses that are likely to happen in a human voice according to its pitch and thereby trains the Vocoder not to transmit during the pauses. This ensures the enhancement of performance of vocoders used in mobile communications[1-10].

Intelligent Pause Detection Protocol (IPDP) probabilistically estimates the number of pauses thatoccur in the given human voice and accordingly instructs the vocoders not to transmit during those pauses thus optimizes the available spectrum and enhances the performance of vocoders. This is done in three steps, firstly by calculating the maximum and minimum pitch values from the human voice, secondly estimating the pauses from the obtained values and thirdly instructing the vocoders not to transmit any type of data during the pauses that are estimated. The better performance has been achieved empirically with the use of Intelligent Pause Detection Protocol (IPDP) when compared to the currently used speech processing protocols.

The research work is specifically aimed on determining the best possibilities of developing a reliable technique to enhance the performance of vocoders used in mobile communications where speech processing is of main concern.The protocol developed provides a basic feasible solution which would ensure performance enhancement of Vocoders by optimizing the bandwidth utilization with excellent speech processing protocol used in a dense and hazard mobile communications environment. We call this technique as “**I**ntelligent **P**ause **D**etection **P**rotocol **(IPDP)”.** This paper is organized as follows: Section 2 highlights the features of Wire Shark simulator tool. Section 3 projects the simulation results of various vocoders. In section 4, IPDP protocol with simulation results are presented. Section 5 demonstrates the performance comparison of vocoders with IPDP protocol and Section 6 gives the conclusion remarks.

2 Literature survey

Different kinds of lossless and lossy data compression methods such as the hybrid based speech compression, parametric based speech compression, and wave form-based speech compression have been discussed in the literature survey. The technique of parametric based code is higher in the complexity of implementation but it has an ability to achieve better compression ratio. It's an interesting fact that the investigation on different vocoders influence on the telecommunication in the future work. Basic principal of a vocoders like system for speech compression and transmission, they were basing on the fact that the analysing and synthesizing subsystems of the different vocoders could be used as experimental devices in basic research hIt has discussed regarding the popular VoIP codes' performance including G.711, G723.1 and G.729A implemented on VoIP network based on RTP.

By analyzing the parameters of quality of service (QoS) like delay variation, Mean Opinion Score (MOS), and jitter, the simulation results have been presented that show more delay in G.723.1 and G.711 than the G.729A. The objective of analyzation and comparing the VoIP performance for G.729 over an MPLS (Multiprotocol Label Switching) and IP Networks is extended for this study by a research team. VoIP or Voice over Internet Protocol is an evolving technology in producing a high-quality service through the incorporation of many methods for voice communication[11-15].

The performances of H.323 and SIP protocols based on the hybrid codes like G.729 and G.711 were investigated that helps to determine the QoS parameters' performances easily such as MOS, packet loss, end-to-end delay, RTP delay, and jitter. The integration of signaling and codec protocol such as G.729 and H.323 was providing most appropriate results in terms of minimum delays like end-to-end delay or RTP delay with an acceptable value of jitter for obtaining the efficient call quality. An approach for Removing silence, based on different methods like windowing and Framing based on various parameters are spectrogram, frequency magnitude spectrogram, and thresholding. The separation of voiced or unvoiced part of speech with an approach in an efficient and simple manner is presented in the paper. But, a plan will study in future work for improving the results in developing a new method.

3 proposed methodology: Performance Analysis of IPDP Protocol

Intelligent Pause Detection Protocol (IPDP) is developed in such a way that for any given human voice (either male or female voice) it calculates the maximum, minimum, mean and RMS value of the pitch of the voice. Using Maximum Likelihood Estimation (MLE), the protocol estimates the pauses that could possible. Once the pause of a particular voice is detected, using DTX algorithm, the protocol instructs the Vocoder to manage the pauses efficiently and band bandwidth can be efficiently utilized so that the spectrum efficiency can be enhanced. The algorithm for IPDP is as follows:

Step 1: Receives the Human Voice.

Step 2: From the received human voice, mean, maximum, minimum, and rms values are estimated.

Step 3: Using Maximum Likelihood Estimation (MLE) mathematical model, silence/pauses are estimated.

Step 4: Using DTX algorithm, vocoders are trained not to transmit packets.

Hence, with the help of IPDP protocol, the performance of Vocoders is enhanced by efficient utilization of bandwidth allocated for speech transmission by mobile networks

3.1 Various Human Voices are Given as Inputs to IPDP

Which in turn calculated the maximum, minimum; mean and rms values of the pitch of the voice are calculated for a window size of 20 milliseconds.

The following figures 1 shows Proposed Flowchart depicting IPDP Protocol. Figure 2 shows Comparison of Existing Methods and Invented Method.

% Improvement = 100 – $(|N_{manual} - N_{IPDP}|/N_{manual})$ x 100

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

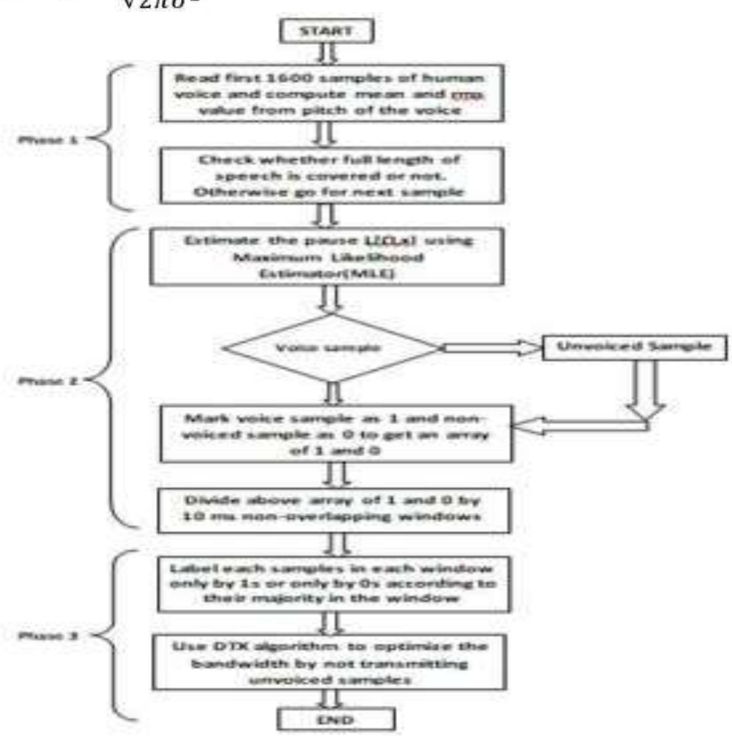

Figure 1 Proposed Flowchart depicting IPDP Protocol

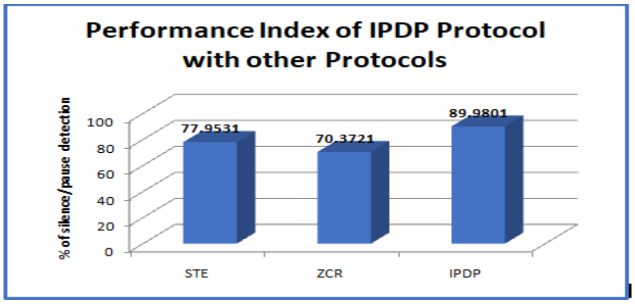

Figure 2 Comparison of Existing Methods and Invented Method

3.2 Mathematical Modeling for New Proposed Method

This method is execution by following formulas using:
Phase 1: By using the human voice pitch, the values of RMS and MEAN are determined in case σ and μ are representing the standard deviation and mean of human voice.

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

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

Phase 2*:* the Based on the Maximum Likelihood Estimation (M.L.E), the pauses circumstances are determined:
$L\ [\Omega, x] = P\ [\Omega, x]\ L\ [\Omega,x] = P[x|\Omega] = p^x(1\text{-}p)^{1-x}$

3.3 Maximum Likelihood Estimation (MLE)

For a given model, the parameters' values are estimated using a technique of maximum likelihood estimation. Here, these values are termed as the maximum likelihood estimates (MLE).

For estimating the more robust parameters, MLE can be utilized extensively. The definition of MLE can described as a method which uses to determine the given model's population parameters such as rms value for Normal, standard deviation, mean, and rate for Poisson (lambda), etc. by using the sample data in a way that the given model's observed data retrieval probability is improved.

The Probability Density Function (f(x|Ω)) has to be estimated for a given observed data of the model among all densities of probabilty that are likely to produce the data most.

A likelihood function L (Ω;x) is assumed in MLE. Here, x represents the set of observations and Ω indicates the distribution parameter vector. . The value of Ω is determined that improves the likelihood for given observations like values of x.

By using the function's derivative with respect to Ω and equate it to 0, the above function's maxima or minima is determined because the zero slop refers to the maxima or minima.

The first derivative of function of LL (Ω;x) with respect to Ω and equate it to 0 is considered to estimate the log likelihood function's maxima.

In this way, consider the second derivative of function of LL (Ω;x) with respect to Ω and it is negative.

The spectral efficiency is achieved with the concept of Discontinuous Transmission (DTX) in which the bandwidth is expensive.

The switching of transmission ON only is the basic principle of DTX for the packets if there is an active speech for transmission of data. The packets transmission is not required during silence regions. For ensuring the link communication continuity, the transmission of some sequence frames of Silence Indicator (SID) is required.

The current DTX schemes has the Voice Activity Detector (VAD) which is the most essential part and designed with the motive of balancing the clipping speech segment risk in regard to the risk that noise classify as speech. A step change causes the background noise at the receiver if the VAD uses for switching on and off the transmitter. The annoying feature can perceive by the receiver due to the sudden change in noise level. If the transmitter is switched off, the producing of noise at the decoder is a way to restrict the sudden change. This noise is same as the background noise at the transmitter. To reconstruct the background noise accurately, the average background data should have to transmit to the receiver periodically by the transmitter if silence is detected. This noise is termed as the comfort noise. The DTX algorithm implementation is showed in the following flow chart.

4 Results and Discussions

4.1 Wire Shark Simulator

Wireshark simulator tool is a free open source tool used for wireless network analysis and trouble shooting, software protocol development, performance analysis of various vocoders used in mobile networks and can be used to validate the newly developed speech processing protocols.

Wireshark is an encapsulated tool which understands various speech processing protocols and networking protocols. It has the capability of resolving and projecting various parameters along with their meaning. It is well supported with pcap so that the packets used for transmission of all types of information can be captured. The salient features of Wireshark simulation tool are mentionedbelow:

- ❖ Data can be retrieved from live captured data packets or from already captured data packets. From a live network connection “from the wire”, information can retrieve or read using already-captured packets file.
- ❖ From different types of network supported protocols such as Ethernet Protocol, Loopback Protocol, Point-to-Point Protocol (PPP), and IEEE 802.11 Protocol, etc., real-time data can be retrieved or read.
- ❖ Retrieved data can be viewed using GUI or utility command line or TShark.
- ❖ Retrieved data can be modified or edited using “edit cap”feature.
- ❖ Display filter feature can be used to fine tune the retrieved data.
- ❖ Newly developed protocols can be dissected by creating Plug-ins.
- ❖ The media flow can be played if compatible encoding is used. VoIP calls can be traced out in the traffic that is captured.
- ❖ The raw form of USB data can also be identified.
- ❖ The connections used in wireless networks can be customized as long as they are enabled with Ethernet protocol.
- ❖ The simulation tool allows customizing timers, filters and network settings so as to capture the required traffic.

4.2 Performance Analysis of Various Vocoders

4.2.1 G.711 ALawVocoder

G.711 ALawVocoder which is commonly used for speech processing in mobile networks is simulated for a payload of 160 bytes for which 2500 packets are required out of which 2132 are actually used without any droppings or error prone. Figure 3 shows the screen shot of G.711 ALawVocoder using Wiresharksimulator.

Figure 3 Performance analysis of G.711 ALawVocoder

4.2.2 G.711 µLawVocoder

G.711 µLaw Vocoder which is commonly used for speech processing in mobile networks is simulated for a payload of 160 bytes for which 2222 packets are required out of which 2118 are actually used without any droppings or error prone. Figure 4 shows the screen shot of G.711 µLaw Vocoder using Wiresharksimulator.

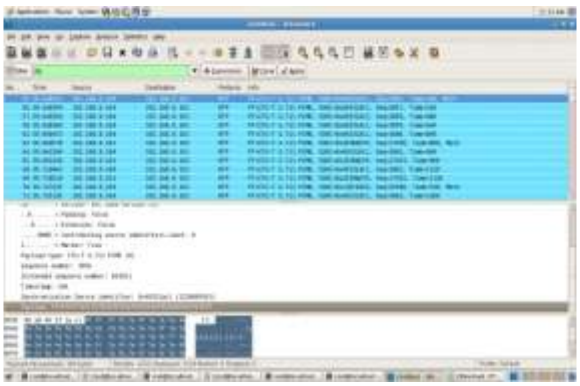

Figure 4 Performance analysis of G.711 µLaw Vocoder

4.2.3 G.722 Vocoder

G.722 Vocoder which is commonly used for speech processing in mobile networks is simulated for a payload of 160 bytes for which 2458 packets are required out of which 2173 are actually used without any droppings or error prone. Figure 5 shows the screen shot of G.722 Vocoder using Wireshark simulator.

Figure 5 Performance analysis of G.722 Vocoder

4.2.4 GSMVocoder

GSM Vocoder which is commonly used for speech processing in mobile networks is simulated for a payload of 160 bytes for which 2359 packets are required out of which 2149 are actually used without any droppings or error prone. Figure 6 shows the screen shot of GSM Vocoder using Wireshark simulator.

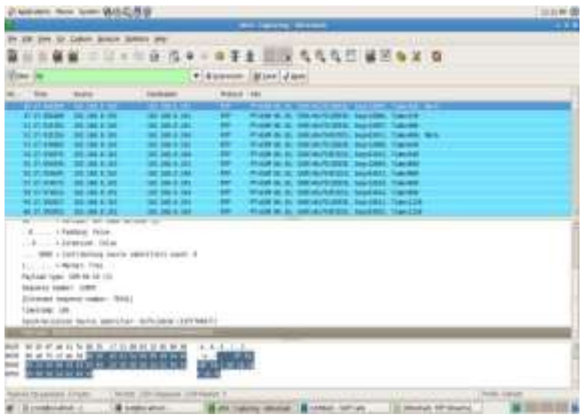

Figure 6 Performance analysis of GSM Vocoder

4.2.5 ILBCVocoder

ILBC Vocoder which is commonly used for speech processing in mobile networks is simulated for a payload of 160 bytes for which 1774 packets are required out of which 1440 are actually used without any droppings or error prone. Figure 7 shows the screen shot of GSM Vocoder using Wireshark simulator.

Figure 7 Performance analysis of ILBC Vocoder

4.2.6 SPEEXVocoder

SpeexVocoder which is commonly used for speech processing in mobile networks is simulated for a payload of 160 bytes for which 2356 packets are required out of which 2134 are actually used without any droppings or error prone. Figure 8 shows the screen shot of GSM Vocoder using Wireshark simulator**.**

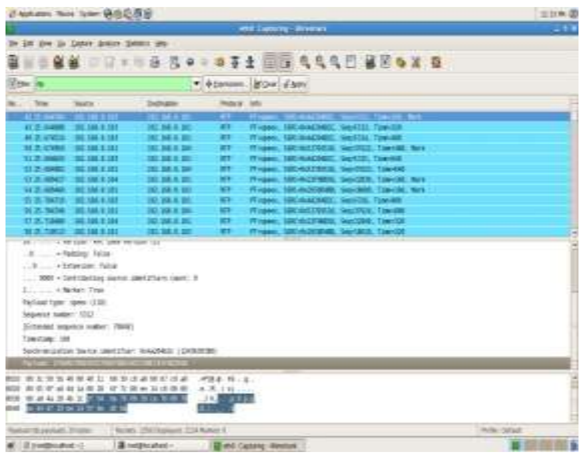

Figure 8 Performance analysis of SpeexVocoder

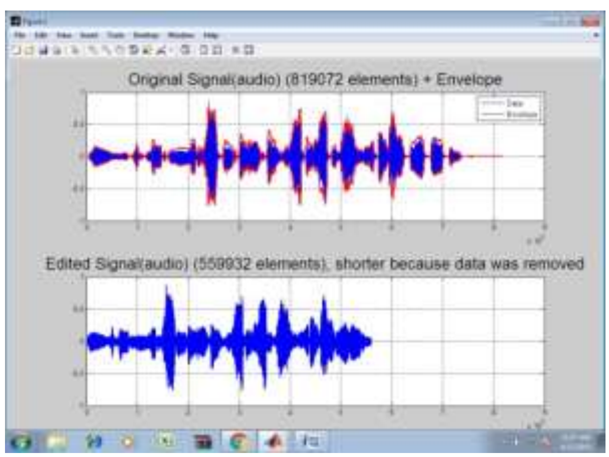

Figure 9 Original Test Signal 1 with Pauses and Edited Signal Avoiding Pauses

In figure 9, Original Test signal 1 requires 819072 elements + envelope, whereas using IPDP protocol, it requires 559932 elements only, thereby optimizing the bandwidth and enhancing the performance of vocoders used for speech processing in mobile networks.

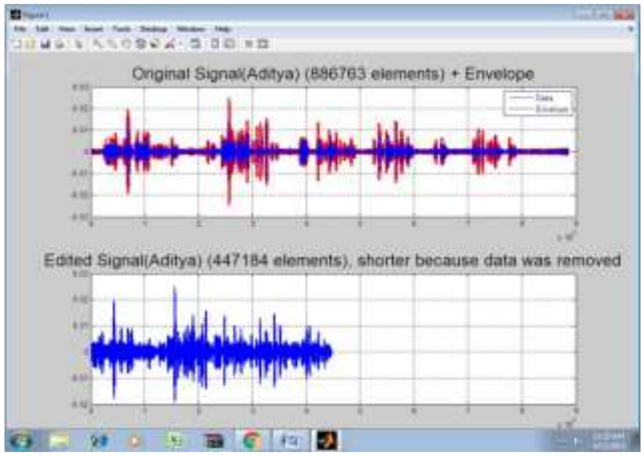

Figure 10 Original Test Signal 2 with Pauses and Edited Signal Avoiding Pauses

In Figure 10, Original Test Signal 2 requires 886763 elements + Envelope, whereas using IPDP Protocol, it requires 447184 elements only, thereby optimizing the bandwidth and enhancing the performance of vocoders used for speech processing in mobile networks.

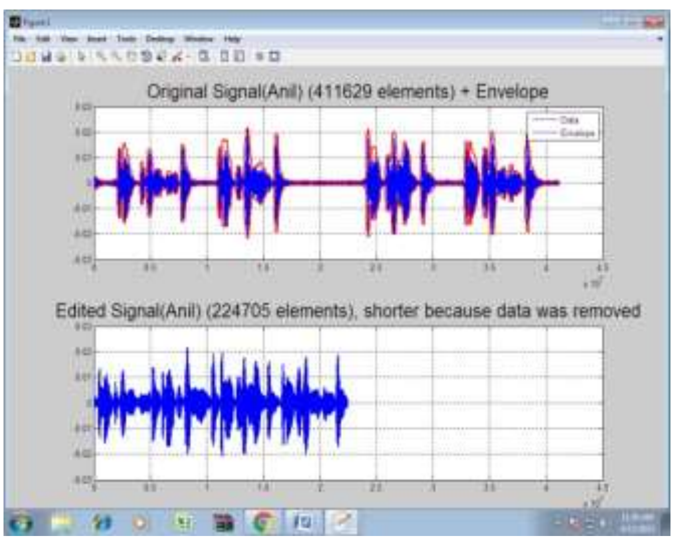

Figure 11 Original Test Signal 3 with Pauses and Edited Signal Avoiding Pauses

In Figure 11, Original Test Signal 3 requires 411629 elements + Envelope, whereas using IPDP Protocol, it requires 224705 elements only, thereby optimizing the bandwidth and enhancing the performance of vocoders used for speech processing in mobile networks.

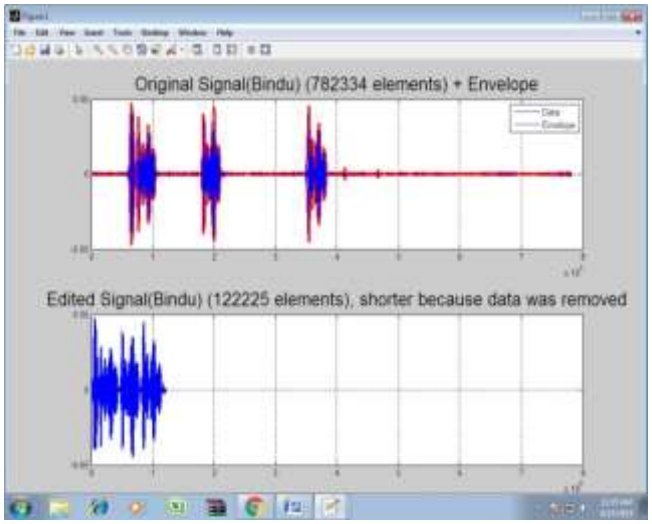

Figure 12 Original Test Signal 4 with Pauses and Edited Signal Avoiding Pauses

In Figure 12, Original Test Signal 4 requires 782334 elements + Envelope, whereas using IPDP Protocol, it requires 122225 elements only, thereby optimizing the bandwidth and enhancing the performance of vocoders used for speech processing in mobile networks.

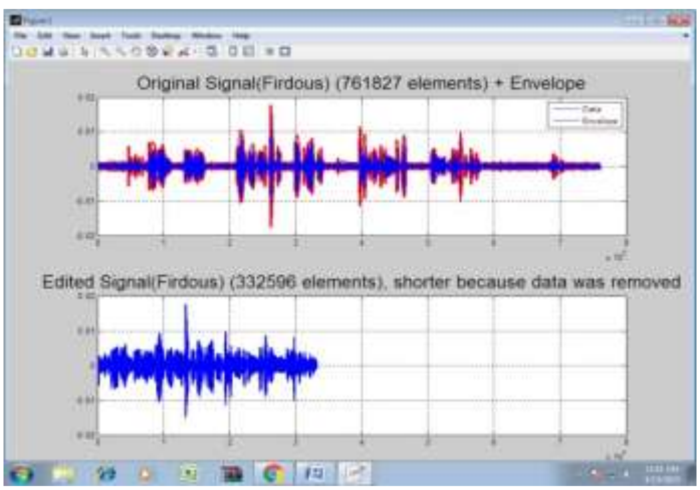

Figure 13 Original Test Signal 5 with Pauses and Edited Signal avoiding pauses

In Figure 13, Original Test Signal 5 requires 761827 elements + Envelope, whereas using IPDP Protocol, it requires 332596 elements only, thereby optimizing the bandwidth and enhancing the performance of vocoders used for speech processing in mobile networks.

4.3 Performance Analysis of Various Vocoders with IPDPProtocol

The following table 1 presents the percentage of data that can be saved without transmitting unnecessary information when the speech has pauses of various vocoders and IPDP enabled Vocoder (proposed). Figure 14 shows Performance Analysis of IPDP Protocol with other Various Vocoders.

Table 1 Performance Analysis of Various Vocoders with IPDP Protocol enabled Vocoder

Type of Vocoder	Original packets of data to be transmitted	Actual packets of data transmitted	Percentage of packets of data Saved
G.711ALaw	2500	2132	14.72 %
G.711µLaw	2222	2118	4.69 %
G.722	2458	2173	11.59 %
GSM	2359	2149	8.91 %
ILBC	1774	1440	18.83 %
SPEEX	2356	2134	9.42 %
IPDP	3200	2187	31.7 %

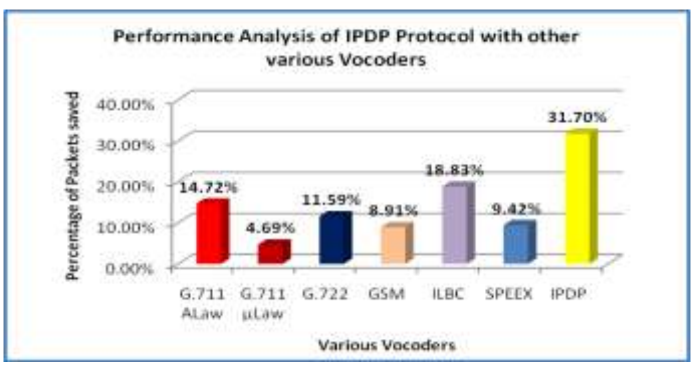

Figure 14 Performance Analysis of IPDP Protocol with other Various Vocoders

5 Conclusions

The testing of data of different human voices is done and computed the parameters by utilizing Intelligent Pause Detection Protocol (IPDP). With the help of MLE model the silence/pause for each human voice is estimated and accordingly vocoders are trained in such a way that no transmissions are done during pauses. Hence the performance of vocoders used for speech processing in mobile networks can be enhanced with IPDP protocol.

References

[1] J. P. Campbell, Jr., “Speaker Recognition: A Tutorial”, Proceedings of The IEEE, Vol.85, no.9, pp.1437-1462, 1997.

[2] Koji Kitayama, MasatakaGoto, Katunobu Itou and Tetsunori Kobayashi, “Speech Starter: Noise-Robust Endpoint Detection by Using Filled Pauses”, 8th European Conference on Speech Communication and Technology, Eurospeech, pp. 1237-1240, 2003.

[3] S. E. Bou-Ghazale and K. Assaleh, “A robust endpoint detection of speech for noisy environments with application to automatic speech recognition”, IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP, Vol. 4, 2002, pp. 3808–3811, 2002.

[4] A. Martin, D. Charlet, and L. Mauuary, “Robust speech / non-speech detection using LDA applied to MFCC”, IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001.

[5] K. Ishizaka and J.L Flanagan, “Synthesis of voiced Sounds from a Two-Mass Model of the Vocal Chords”, Bell System Technical J., Vol. 50, no. 6, pp. 1233-1268, 1972.

[6] Atal, B. Rabiner, L., “A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition”, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 24 , no. 3 , pp.201 – 212, 1976.

[7] D. G. Childers, M. Hand, J. M. Larar, “ Silent and Voiced/Unvoied/ Mixed Excitation(Four-Way), Classification of Speech”, IEEE Transaction on ASSP , , Speech, and Signal Processing, Vol. 37, no.11, pp. 1771-1774, 1989.

[8] Mark Greenwood and Andrew KInghorn, “SUVing: Automatic Silence/Unvoiced/Voiced Classification of Speech", Presented at the university of Sheffield.

[9] Richard. O. Duda, Peter E. Hart, David G. Strok, “Pattern Classification”, A Wiley-interscience publication, John Wiley & Sons, Inc, Second Edition, 2001.

[10]Sarma, V.; Venugopal, D., “Studies on pattern recognition approach to voiced-unvoiced-silence classification”, Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '78. , Vol. 3, pp. 1 – 4, 1978.

[11]L. R. Rabiner and R.W. Schafer, “Digital Processing of Speech Signals”, First Edition, Pearson Education, Prentice-Hall,1978

[12]Available online: http://cslu.ece.ogi.edu/nsel/data/SpEAR_technical.html.

[13]J. L. Flanagan, “Speech Analysis, Synthesis, and Perception”, 2nd ed., Springer-Verlag, New York, 1972.

[14]L. R. Rabiner and B. H. Juang, “Fundamentals of speech recognition,” 1st Indian Reprint, Pearson Education,1993

[15]R Chinna Rao, Elizabath Rani, S Srinivasa Rao, “Basic Frame work of Vocoders for Speech Processing”, Soft Computing and Signal Processing, 2019.

Biographies

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 asAssistantProfessor, Dept. of ECE in Malla reddy College of Engineering and Technology, Secunderabad, India. PhD Scholar of GITAM, Visakhapatnam,A.P.

P.V.Y.Jayasree received her B.E degree in Electronics &Communication Engineering from College of Engineering, GITAM affiliated to Andhra University in 1989.M.E in Electronics Communication Engineering from Andhra University, Visakhapatnam in 1999.Ph.D inElectromagnetic interference and Compatibility from JNTUK, Kakinada in 2010.Presently working as a professor & HOD, Department of Electronics and communication Engineering ,GITAM University, A.P India, her Interests are Electromagnetic interference and Compatibility, Antennas and Microwaves.

S.Srinivasa Rao, received the B.Tech degree from Madras Institute of Technology, Anna University, and the M.Tech and Ph.D from JNTUHyderabad,Telangana, India. Presently working as Professor and Head of the Deportment 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.