Intelligible Semantic Level Speech Compression Algorithm by Preserving Emotional Content

Firos A
Department of Computer Science & Engineering,
Rajiv Gandhi University
Doimukh, Arunachal Pradesh, India
firos.a@rgu.ac.in

Prof.Utpal Bhattacharjee
Department of Computer Science & Engineering,
Rajiv Gandhi University
Doimukh, Arunachal Pradesh, India
utpal.bhattacharjee@rgu.ac.in

***Abstract*—**Speech encoding refers to compression for transmission or storage, possibly to an unintelligible state, with decompression used prior to playback. This paper attempts to formulate the semantic level compression technique on speech signals by preserving its prosodic features. LPC analysis will be done to identify the feature of the input speech. GMM will be used to preserve the emotional content while encoding. ANN will be utilized to identify the best features for encoding. Using such semantic based coding will highly reduce the computational overhead in speech coders.

Keywords— Speech coding; G.723.1, iLBC; fuzzy clustering; Windowing; ANN

I. INTRODUCTION

It is called the Three-dimensional sound compression of Google and its founders insist on strict anonymity of its compression method. On this particular compression method, it has assembled a method for encoding multiple directional audio signals using an integrated codec by a wireless communication device.

The clinking use and buzzing popularity are proof that the Three-dimensional sound compression already a success with its the ability to capture, compress, and transmit three-dimensional (3-D) audio. And it's no wonder — looking at what the menu has in store for the VoIP enthusiasts. The modern VoIP can offer recording a plurality of directional audio signals and a taste-bud sating dessert of transmit three-dimensional (3-D) audio with ultimate compression.

Still, the mystery that's impossible to crack is: who among the VoIP codec will give the best compression rate. The speech codec founders often opt to remain secret and inform the users with a best value rather than its average score. The codecs considered for the comparative study: G. and iLBC, does it slightly differently and usually teams up with best or highly talented VoIP solutions networks that create a special meal for its users.

This paper is organized as follows. Section II describes various existing speech coding techniques. Section III describes the proposed speech coders and its Features. Section IV describes The Comparative Study. Finally, Section V concludes the paper.

II. EXISTING SPEECH SYSNTHESIS TECHNIQUES

All around world, similar secret VoIP codecs are springing up which take the formality out of the voice chatting experience and inject a new sociable element. They offer an enticing combination of superlative clarity that we were not able to make till now, combined with best sample rate and variable bit rates — and carefully chosen — network standards.[1]

Even though these are exclusive, low-delay CELP, Lossy codecs like G.728 are available they are not widely uses since it does not offer VBR ,Stereo and Multichannel access for its users.[1] Whenever some codec adds a the favor ,high compression rate beyond certain level it goes through the weakness [2]. so needs a vets on it and adds them on the selection list for the suitable speech communication standards only after that.

The coding strategy for this study was limited to standards within the following 14 sectors defined according to the ITU-T Perceptual evaluation of speech quality (PESQ) tool Sources (P.862 (02/01):

Narrow-band speech coding

- G.723.1, G.726, G.728, G.729, iLBC and others for VoIP or videoconferencing
- Full Rate, Half Rate, EFR, AMR for GSM networks
- SMV for CDMA networks

Wide-band speech coding

- G.722, G.722.1, Speex and others for VoIP and videoconferencing
- AMR-WB for WCDMA networks
- VMR-WB for CDMA2000 networks

The G.722 uses a lossy sub-band ADPCM algorithm with a sampling rate of 16kHz . It works on bit rate of 64 kbit/s (comprises 48, 56 or 64 kbit/s audio and 16, 8 or 0 kbit/s auxiliary data) and will have 14 bits/sample with a latency of 4ms.it support constant bit rate (CBR) and does not support variable bit rate(VBR)[3]. The G.722.1 uses a Modulated Lapped Transform, (based on Siren Codec), Lossy algorithm. It works on bit rate of 24,32 kbits/sec and will have 16

978-1-5386-1887-5/17/$31.00 ©2017 IEEE

bits/sample with a latency of 40ms.it support constant bit rate (CBR) and does not support variable bit rate(VBR). The G.722.1C uses a Modulated Lapped Transform, (based on Siren Codec), Lossy algorithm. It works on bit rate of 24,32 kbits/sec and will have 16 bits/sample with a latency of 40ms.it support constant bit rate (CBR) and does not support variable bit rate(VBR).

The G.722.2 (AMR-WB) uses a multi-rate wideband ACELP, Lossy algorithm with a sample rate of 16KHz. It works on bit rate of 6.60, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05, 23.85 kbit/s and will have 14 bits/sample with a latency of 25ms.it support constant bit rate (CBR) and variable bit rate(VBR). The G.723 uses a ADPCM, Lossy algorithm with a sample rate of 8KHz. It works on bit rate of 24, 40 kbit/s and will have 13 bits/sample..it support constant bit rate (CBR) and does not support variable bit rate(VBR). The G.723.1 uses a MP-MLQ, ACELP, Lossy algorithm with a sample rate of 8KHz[4]. It works on bit rate of 5.3, 6.3 kbit/s and will have 13 bits/sample with a latency of 37.5 ms .it support constant bit rate (CBR) and does not support variable bit rate(VBR).

iLBC, another codec does it slightly differently. The iLBC uses a block independent linear predictive coding Lossy algorithm with a sample rate of 8KHz. It works on bit rate of 15.2 kbit/s for 20 ms frames, 13.33 kbit/s for 30 ms frames. It support constant bit rate (CBR) and does not support variable bit rate (VBR)[5].

Shift the scene to G.729 and the tables of modern VoIP coding techniques. The recent turnaround of the speech coding like G.729(used in videoconferencing also built largely on this simple insight, as well as on the related fact that broad coverage strengthens coding standards for a sophisticated VoIP. There is an important lesson here for the proposed by the recent developments of speech coding standards.

III. PROPOSED ALGORITHM

. The Codec Description

- The input speech will be given for LPC analysis for getting the LPC$e(n) = x(n) - \sum_{k=1}^{p} \alpha_k \; x(n-k)$. The result will be the smaller varying error coefficients according to the syllable, $\alpha_1, \alpha_2, \alpha_3$,----------$\alpha_p$. These parameters are to be found with the help of LPC . We get the this signal in Its z - transform X(z) =$\frac{1}{1-\sum_{k=1}^{p} \alpha_k \, z^{-k}} \; E(z)$,. The result will be have p equations and p unknowns ($\alpha_1, \alpha_2, \alpha_3$,----------$\alpha_p$) at every 20ms . so we need to find $\alpha_k' s$ on every 20ms. Since this is not computational efficient , auto correlation method wil be usesd.so , we get $S_n(m)$=S(m+n)w(m) ; where (m) is the window ; 0<=m<=N-1. So we have E_n=$\sum_{m=0}^{N+p-1} S_n^2(m)$. With this we will get $\emptyset_n(i,k)$= $R_n(i-k)$, where R_n (k)= ∑_(m=0)^(N-1-k)▒〖S_n (m) 〗 S_n (m+k) ; R_n (k) is going to be even function. With this for i—1,2,----p we will get a matrix

$$\begin{bmatrix} R_n(0) & R_n(1) & R_n(2) & \dots\dots\dots & R_n(p-1) \\ R_n(1) & R_n(0) & R_n(1) & \dots\dots\dots & R_n(p-2) \\ R_n(2) & R_n(1) & R_n(0) & \dots\dots\dots & R_n(p-3) \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \\ R_n(p-1) & R_n(p-2) & R_n(p-3) & & R_n(0) \end{bmatrix}$$

$$\mathbf{X}\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ . \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ . \\ R_n(p) \end{bmatrix}$$

interestingly , since the diagional elements are the some and its a Toeplits matrix , its computationally easy for LPC for computing its α_s

- At the time LPC is going ahead with the analysis of probabilistic features Gaussian Mixture Model (GMM)-based emotional voice analysis will be done in the same frame to find the prosodic features. Simultaneously the features of the speech signal are extracted by the MFCC block. The total number of samples chosen in a frame is 256 and overlapping samples with the adjacent frame will be 128. We acquire MFCC cepstral coefficients at the output of MFCC block. In GMM, K-mean algorithm is used to obtain a cluster number specific to each observation vector and sets the centroid of the observation vector. After clustering the model, it returns one centroid for each of the cluster K and refers to the cluster number closest to it. K-mean algorithm is described as the squared distances between each observation vector and its centroids. In the training section parameters of GMM model are produced iteratively by expectation-maximization (EM) algorithm. Euclidean distance is found out between observation vector and its cluster centriods to match the spoken word with the present database[3]. The proposed method is depicted in the figure 1.
- the matrices α ,e and w will be taken into feed forward neural network, A feed forward neural network algorithm includes the following steps:
    1. Initialize weights and biases to small random numbers.
    2. Present a training data to neural network and calculate the output by propagating the input forward.
    3. changing in numbers of hidden layers and transfer function for every hidden layer and for output layer and also changing in number of neurons in every hidden layer until reach to maximum recognition and language identification rate or to minimum error.

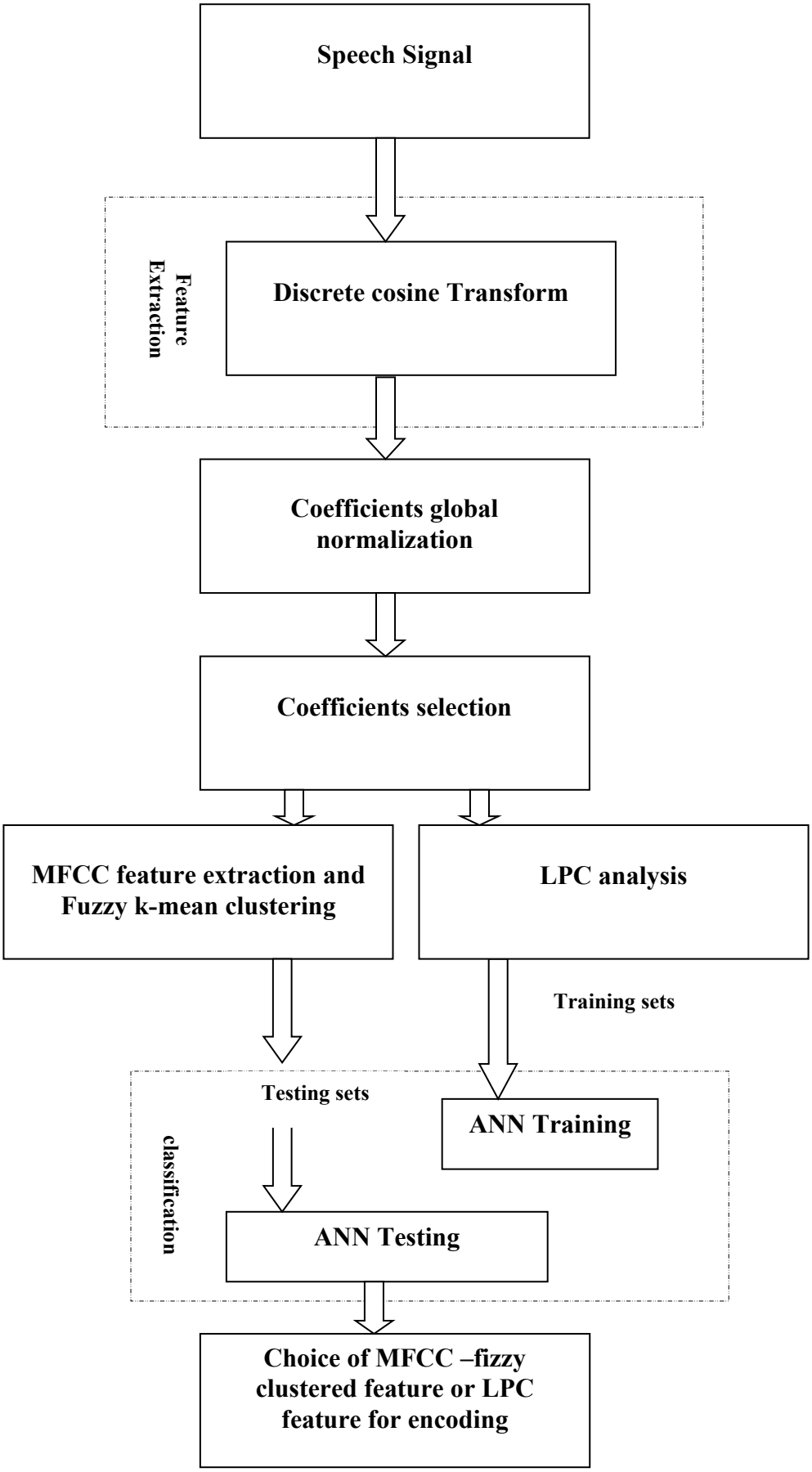

Fig 1. The Proposed encoding method

- The feed forward neural network will give Choice of MFCC –fizzy clustered feature or LPC feature for encoding

IV. RESULTS AND DISCUSSION

The study starts with comparative analysis of proposed method with the algorithm: iLBC

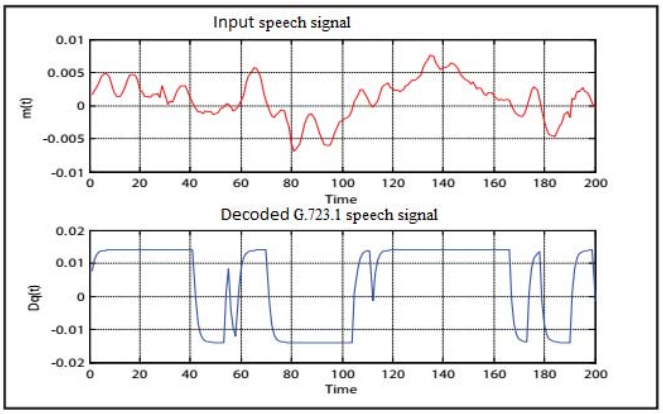

Fig.2 Encoding and Decoding for iLBC

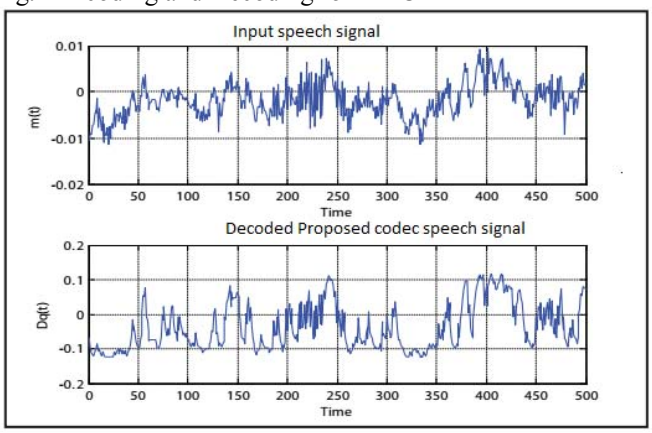

Fig.3 Encoding and Decoding for proposed systems

MATLAB simulation of the input voice for iLBC and proposed coders have been plotted graphically (Fig.1-Fig.2). The proposed method reproduces the signal more closely to the original signal as compared to other coders . It is noted that as the bit-rate goes down, the computation requirements increases highly for different bits used. This is the motivation for the proposal of semantic based speech code. LP estimation for iLBC is depicted in fig-3.This introduces a delay as well as an increase in the cost of implementation. However, for equal number of bits used bandwidth in G.723.1and iLBC is reduced highly than waveform coders, making them most suitable in bandwidth scarcity situations.

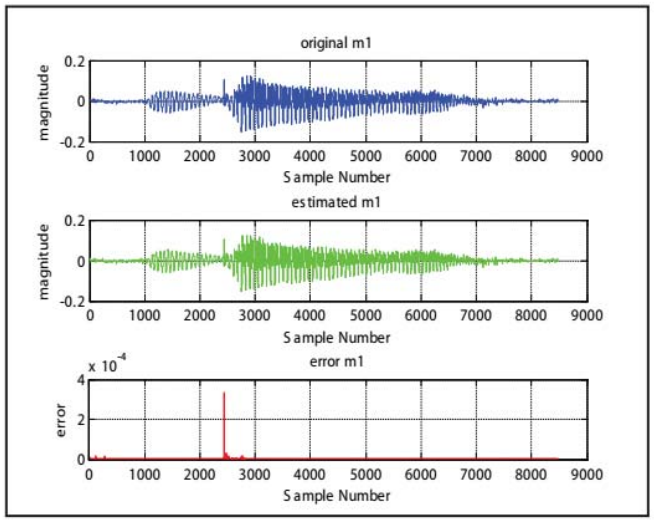

Fig 4. LP estimation for proposed method

V. CONCLUSION

In this paper, a novel semantic based speech compression methods which achieve the best possible speech quality low bit rate, with constraints on complexity and delay.

This paper proposes a mechanism to encoding speech by preserving its emotional content. The emotional preservation was achieved with the help of GMM using GMM where in the semantics of the speech may be identified with its emotional content. Since the accuracy is a concern in GMM, LPC also will be incorporated and a better choice of either GMM or LPC feature for decoding will be done with the help of ANN.

Speech coding algorithms are improving day by day to address the issues of speech communication standards. Even though this issue is addressed and solved, the VoIP industry demands lower bit energy efficient speech codes.

References

[1] Ying-Hui Lai, Fei Chen , Yu Tsao ,"Adaptive Dynamic Range Compression for Improving Envelope-Based Speech Perception: Implications for Cochlear Implants " Springer, Emerging Technology and Architecture for Big-data Analytics,pp 191-214, April 2017

[2] Stanislaw Gorlow ; Joshua D. Reiss ."Model-Based Inversion of Dynamic Range Compression" IEEE, IEEE Transactions on Audio, Speech, and Language Processing , Page(s): 1434 - 1444 ,Volume: 21 Issue: 7, July 2013

[3] Virendra Chauhan, Shobhana Dwivedi, Pooja Karale, Prof. S.M. Potdar "SPEECH TO TEXT CONVERTER USING GAUSSIAN MIXTURE MODEL(GMM) ", International Journal of Engineering Research and Applications (IJERA), ISSN: 2248-9622, Vol. 2, Issue 3, May-Jun 2012, pp.1169-1173.

[4] Dhinesh Babu L.D, P. Venkata Krishna, "Honey bee behavior inspired load balancing of tasks in cloud computing environments", Applied Soft Computing 13 (2013), pp.2292–2303.

[5] Matthias Schmidt,Niels Fallenbeck,Matthew Smith,Bernd Freisleben,"Efficient Distribution of Virtual Machines for Cloud Computing",Proceedings of the 2010 18th Euromicro Conference on Parallel, Distributed and Network-based Processing,IEEE Computer Society Washington, DC,(2010), pp.567-574