International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-2 Issue-4, April 2013

A New Technique for Artificial Bandwidth Extension of Speech Signal and its Performance Analysis

Tejal Chauhan, Shraddha Singh, Sameena Zafar

Abstract— ***In current scenario of wireless communication system, quality of voice output is degraded due to its limited bandwidth (300-3400 Hz) and power constraints which in turn offers speech sounding muffled and thin. Recent wireless systems involved in transmission of speech demands evolution of efficient and effective methods for maintaining quality of speech, especially at the receiving end. In order to obtain toll quality of speech and high intelligibility cum naturalness in wireless systems, NB speech coders should be upgraded to its counterpart WB coders (50-7000Hz). For the effective utilization of WB speech communication in wireless media, it is indeed necessary to upgrade both end devices and network to be WB compatible which is costly and time consuming affairs. In the meantime some techniques have been developed to artificially extend bandwidth of NB speech to WB at receiver which leads to improvement in the quality of recovered speech. Amongst all elements of the communication system (channel, transmitter and receiver), quality and intelligibility of voice at receiver side majorly depend upon channel condition. Many techniques are adopted to mitigate the effect of the channel. In order to maintain quality and naturalness of voice at receiver side in various unpredictable channel conditions, AMR (Adaptive Multi Rate) NB is considered to be one of the obvious potential candidates. AMR NB is operated on various modes of bitrate between 4.75 and 12.2 kbps. Depending upon the channel conditions, specific mode of operation is selected dynamically. For example, Low bit rate mode of operation is selected in bed channel conditions, that allows more error protection bits for channel coding and vice versa. Since inception, many speech coding techniques like CELP, ACELP and RPE-LTP are adopted in different applications in 2G and 3G. In this paper, implementation of ABWE algorithm is developed on CELP based GSM AMR 06.90 NB Coder using MATLAB simulation; further Subjective (MOS) and Objective (PESQ) analysis are carried out to judge the overall performance of developed coder. The evaluated results for both analyses clearly advocate that BWE coder offers significant improvement in recovered speech quality in comparison with legacy GSM AMR NB decoder.***

Index Terms— ***ABWE, AMR, CELP, GSM, Speech coding, Steganography, Subjective Analysis, Objective Analysis.***

Manuscript published on 30 April 2013.
***** Correspondence Author (s)
Tejal Chauhan*, Research Student, E&C, Patel College of Science and Technology, Bhopal, India.
Shraddha Singh, Assistant Professor, E&C, Patel College of Science and Technology, Bhopal, India.
Sameena Zafar, Assistant Professor, E&C, Patel College of Science and Technology, Bhopal, India.

I. INTRODUCTION

Currently, in telephone networks bandwidth is limited to 300-3400 Hz, results in degradation of speech quality.
One of the solutions to achieve high quality speech in communication system by introducing Wide Band (WB) coding scheme. In WB communication system, the transmission of signal having cut-off frequency of at least 7 kHz is used for better speech quality in terms of increased intelligibility and naturalness. Unfortunately the problem in the utilization of WB coders and communication is an expensive up-gradation of current Narrow Band (NB) coders and transmission to its counterpart WB system and in which compatibility as well as software cum hardware up-gradation is a major issue. Also it is not feasible to suddenly replace of entire current NB coding system. Current speech transmission system is a mixture of traditional NB terminals and new WB terminals. During the long transitional period between NB and WB system, speech quality can be enhanced without much modification of existing network infrastructure by Artificial Bandwidth Extension (ABWE) [1,2] by implemented only at the receiving side. Producing the WB speech at receiver side by estimating missing high frequency components of the speech from available NB speech is called ABWE [2]. WB speech, which is artificially regenerated by stand-alone ABWE, still gives limited performance [3]. Therefore this research introduces a novel approach of transmitting the High Band features like envelope, pitch and gain information from HB portion of WB speech as side information for regeneration of efficient WB speech at receiver. These parameters are embedded in the NB coded bitstream by steganography and transmitted to the receiver. At receiver, using legacy NB decoder the NB signal is recovered and using extracted side information, regeneration of the HB signal is performed. Finally summing this reproduced HB signal with the recovered NB signal generates WB signal having speech quality significantly better than the legacy NB decoded speech signal.

Here, in this paper WB speech is splitted into NB and HB having cut-off frequency 3.4 kHz. NB is encoded using GSM AMR 06.09 NB coder by selecting three various bit rates according to channel condition. Side information is extracted using LPC method from HB which is watermark embedded and transmitted along with encoded NB bitstream to the receiver. At receiver, along with decoding NB bitstream, HB speech can also be reproduced from the watermarked extracted HB features. Finally WB signal is generated by summing NB and HB version of speech.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Paper is organized as follows: Overview of basic GSM AMR 06.90 NB coder is given in sec.2. Detailed process of BWE coder implementation is explained in sec.3, proposed method of steganography in NB coder in Sec.4. Sec.5 demonstrates overall performance analysis and results of proposed coder. Finally concluding remarks are given in sec.6.

II. ADAPTIVE MULTI RATE CODER

To maintain good speech quality under varying channel conditions, a technique called AMR, is utilized. AMR has several operating codec modes which are switched adaptively in accordance with the dynamics of the channel (good or bad channel). The process of dynamically switching due to varying channel conditions is known as AMR adaptation [4]. AMR has several codec modes, for example; AMR full-rate and AMR half-rate.

As shown in Fig. 1, AMR full-rate has 8 codec modes operated in kbps, i.e.: 12.2, 10.2, 7.95, 7.4, 6.7, 5.9, 5.15, and 4.75; whereas AMR half-rate has 6 codec modes operated in kb/s, i.e.: 7.95, 7.4, 6.7, 5.9, 5.15, and 4.75. The above codec modes are a precipitate of the ETSI standardization in 1999.

Fig. 2 is a block diagram of AMR codec operation. AMR working is a two-step process; uplink channel and downlink channel. In uplink channel, MS sends information to BTS which contains speech data, codec mode indicator uplink (MIu), and codec mode request downlink (MRd).

Main information sent from MS via BTS is Speech data with bit rate determined by its speech coder. Codec mode indicator uplink (MIU) & codec mode request downlink (MRd) are sent along with the speech data. This process is repeated continuously for the next speech data sent. MIu comes from codec mode command uplink (MCu) to determine codec mode which will be used for uplink transmission from MS to BTS. MRd comes from computation of error condition at downlink channel from BTS to MS (downlink quality measurement). After MRd arrives at BTS, the process of downlink mode control is carried out to produce codec mode [4].

Uplink channel performance process is very dependent on downlink channel performance process, and vice versa. In downlink channel, BTS send information to MS which contains codec mode command uplink (MCu), speech data andcodec mode indicator downlink (MId).

At first, MCu comes from the computation of error condition in uplink transmission (uplink quality measurement), which is then processed through uplink mode control to produce codec mode command uplink (MCu). Arriving at MS, MCu is then used to give codec mode order which will be used for uplink transmission from MS to BTS. MId comes from codec mode command downlink (MCd) which is used to determine codec mode to be used for downlink transmission from BTS to MS [4].

The computation of the error condition quality in uplink and downlink channel (uplink quality measurement and downlink quality measurement on Fig. 2) is determined by the ratio of signal carrier (C) to interference (I), which expressed in dB. For example, as can be seen in Fig. 3, the quality of channel condition has a value from 2 to 22 dB. One from the three codec modes will be selected (12.2, 7.95 or 5.9 kbps) in accordance with its respective channel condition. The quality

Fig. 1 AMR full-rate and half-rate codec [4]

of channel condition is computed on the parts of channel measurement (Fig. 2), as well as on the mobile station and BTS.

The working process of the uplink and downlink channel in AMR is going on repeatedly, means that if there is a sudden change in the channel condition, change in the codec mode will also take place in accordance with the channel condition at that moment. Codec mode is determined by the quality of the channel condition. If the channel condition is good, then there will be no significant error correction, so that the bit rate speech coding sent is higher than the bit rate channel coding (error correction). On the contrary, if the channel condition is bad, then a significant error correction will be needed. In this case, the information sent will have much correction, so that it will produce good voice quality [4].

A. *CELP Coder*

The CELP block diagram is shown in Fig. 4. Voice signal input is the conversion result of human voice – an analog signal to digital signal which is carried out by ADC (Analog to Digital Converter).

The buffer process and LP (Linear Prediction) analysis is used to estimate the vocal impulse response system at each frame, which then produces pitch delay and LP coefficient. The pitch delay will be used in the pitch synthesis filter and the LP coefficient (ai) at LP synthesis filter. Before the process of pitch synthesis filter, pitch filter coefficient (b) will be produced first from the computation using the pitch delay (P).The process of LP synthesis filter before the processing of LP coefficients in the block is carried out by converting LP coefficient (ai) to become the reflection coefficient of LP [5]. To obtain gain parameter and codebook index, perceptual weighting filter process will be done and then error minimization process is carried out. In the error minimization block, gain and codebook index which will be used in the next block will be determined [5].

After all the parameters are obtained, voice compression process can be carried out. Each block will work as shown in Fig. 4. Therefore, the compressed voice signal can be formed and adapted to the AMR codec mode.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

B. CELP Based AMR Coder

For comparing the performance of each AMR codec mode of operation, simulation using MATLAB program is carried

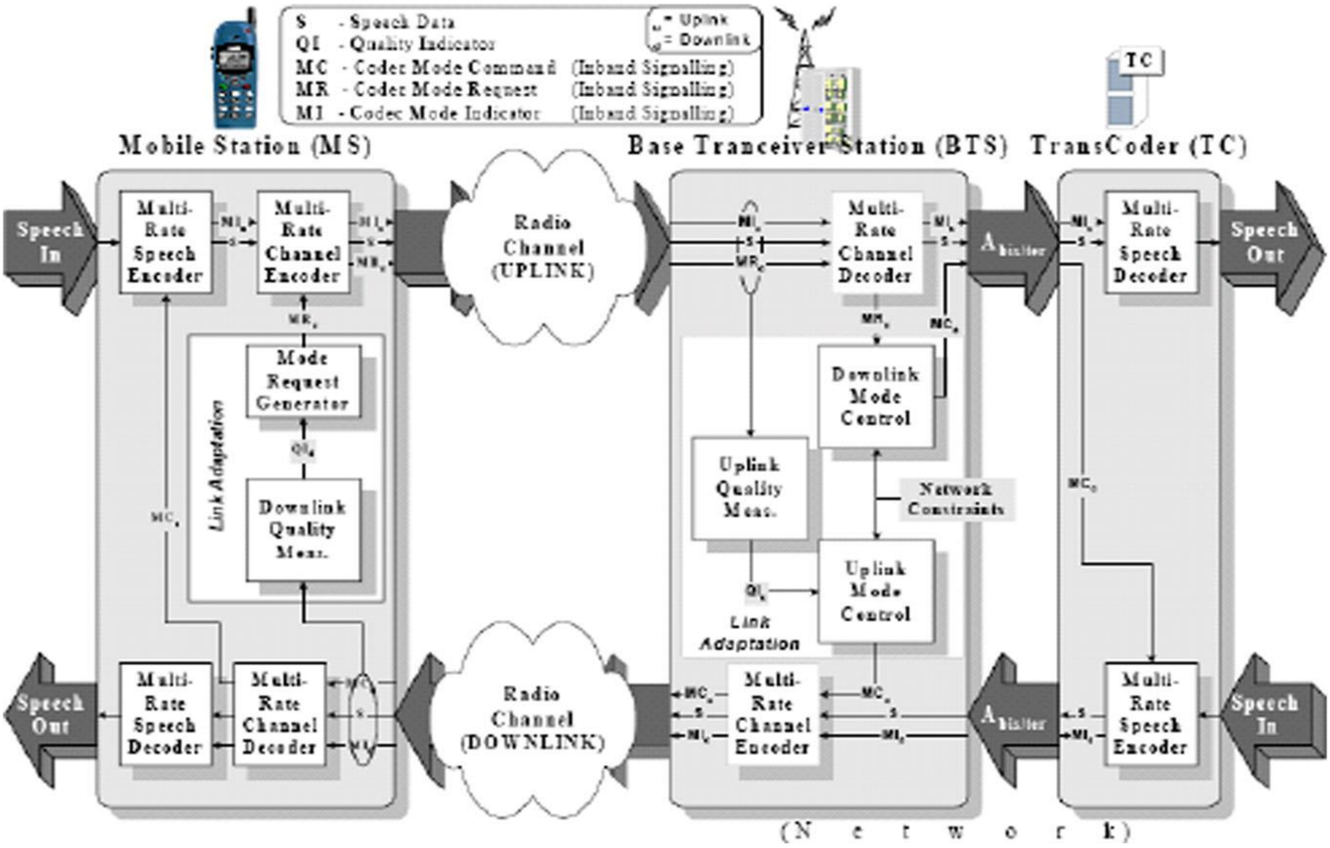

Fig. 2 Block Diagram of AMR Codec Operation [4]

out using CELP speech coding technique and is in sync to the bit rate used in AMR as per ETSI standards. The simulations considered Standard GSM Cellular System architectural configuration, having a 20 ms mother frame, consisting of four sub-frames within the mother frame, with each sub-frame having a length of 5 ms.

As can be seen in Fig. 5, Speech and C/I ratio are provided as an input to the MATLAB e-test bench of CELP based AMR Codec. Depending upon the provided C/I ratio, the particular mode of full rate AMR Codec is selected. MATLAB program has provision to continuously monitor the change in C/I ratio so that appropriate full rate mode is selected for subsequent frames as per fig.3 [8]. After selection of Codec mode, Linear Prediction analysis is carried out to produce the parameters like Frame length (N), Block length (L), Order of filter (M), LP parameter (c), Codebook index (Cb) and Pitch index (Pidx) are computed and provided to CELP analysis part of AMR Codec. Information parameters relating to the AMR codec include: linear prediction coefficient (a), pitch lag (P), codebook index (K), gain (Θ_0) and pitch filter coefficient (B), been investigated, studied and recorded through the simulation e test bench created using MATLAB. Same set of parameters are provided to CELP Synthesis block for the purpose of AMR decoding which ultimately results into recovered speech signal So (i).
Here CELP based full rate GSM AMR coder is implemented. As each codec mode has different bit allocation, the total bits is in accordance with the bit allocation for AMR codec as per ETSI standards 1999.

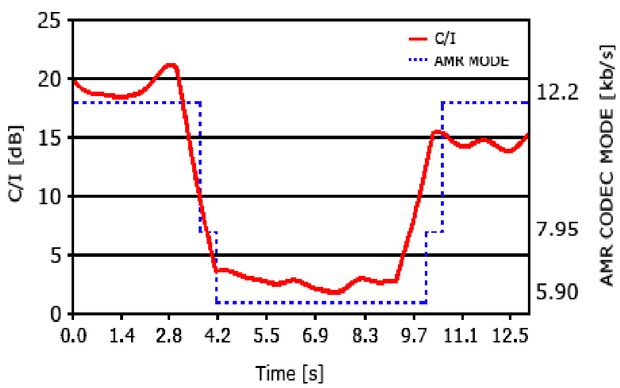

Fig. 3 AMR Codec mode selection [4]

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

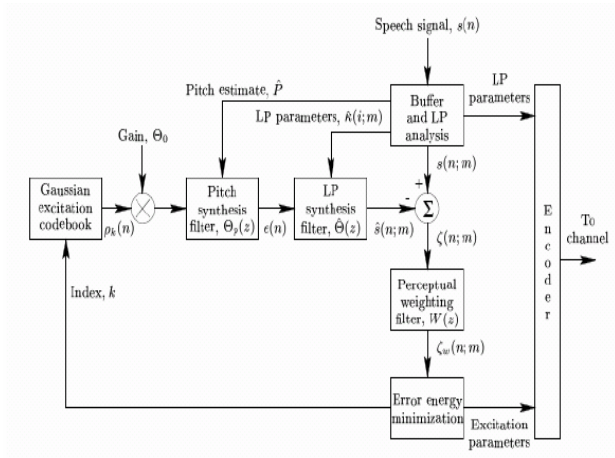

Fig. 4 CELP speech coder

III. ABWE WITH SIDE INFORMATION

Input given to the system is WB speech signal with sampling frequency of 16 kHz. This signal is separated into NB and HB signal using the LPF and HPF respectively (having cut- off frequency equal to 3.4 KHz) followed by down sampling both of them with the factor of two. Produced NB signal is framed with 160 samples / frame and given as an input to the GSM 06.90 AMR encoder to encode the NB signal which generates NB bit stream. At the same time, HB signal is utilized to capture HB features like LP coefficients [11], (called side information) with the use of all pole LPC filter of order 8 for each frame. The LPC coefficients are non-uniformly quantized as per their subjective importance. Table 2 demonstrates bit allocation for HB parameters for various AMR bit rates.

Quantization of this parameter generates number of bits with respect to selected bit rate mode to be transmitted as HB side information. These bits of side information are then embedded to watermark embedding algorithm to hide into GSM AMR NB bitstream selected.

At receiver, NB bitstream of selected bit rate is applied to Watermark extraction algorithm which separates out the HB side information from the input bitstream. Then NB bitstream is (frame wise) supplied to GSM AMR decoder (legacy NB decoder) to finally reproduce NB speech after interpolating it by the factor of two and by passing it to LPF with cut-off frequency of 3.4 KHz. Non uniform de-quantization is performed to recover HB features from the side information bits received from watermark extraction block. These reproduced features are then supplied to various blocks like Analysis filter, Excitation Regeneration and Synthesis Filter to effectively reconstruct HB speech after HPF with cut-off frequency of 3.4 KHz. Finally both NB and HB recovered signals are summed up to reconstruct WB speech at receiver.

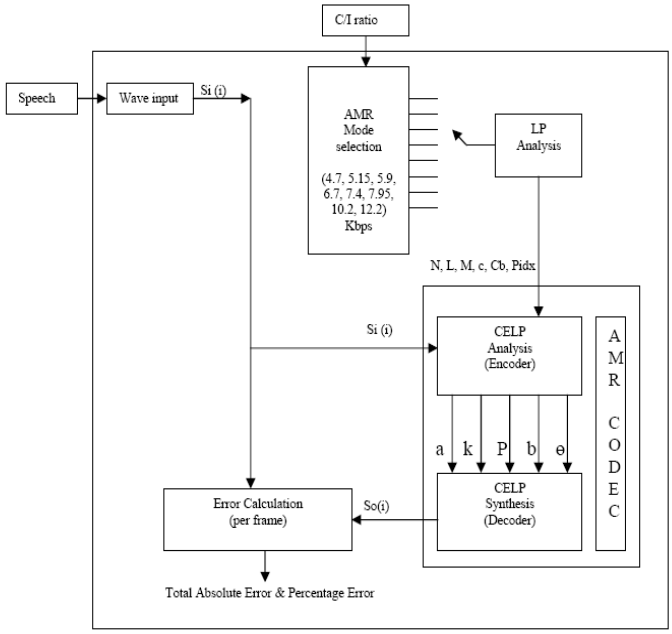

Fig. 5 MATLAB implementation of CELP based AMR Codec

Table: 1 AMR bit rate selection

AMR Kbps	a	P	K	Θ_0	B	Total Bits
12.2	24	8,8,8,8	17,17,17,17	15,15,15,15	15,15, 15,15	244
10.2	24	8,8,8,8	15,15,15,15	11,11,11,11	11,11, 11,11	204
7.95	23	8,8,8,8	12,12,12,12	7,7,7,7	7,7,7,7	159
7.4	24	8,8,8,8	11,11,11,11	5,5,5,5	5,5,5,5	148
6.7	22	8,8,8,8	10,10,10,10	5,5,5,5	5,5,5,5	134
5.9	22	8,8,8,8	9,9,9,9	4,4,4,4	3,3,3,3	118
5.15	19	8,8,8,8	8,8,8,8	3,3,3,3	2,2,2,2	103
4.75	19	8,8,8,8	7,7,7,7	2,2,2,2	2,2,2,2	95

IV. DATA HIDING IN PROPOSED CODER

In this research, GSM AMR NB bitstream of the selected bit rate mode is considered as carrier signal over which side information (frame wise) is embedded. This process is carried out in Watermark Embedding algorithm block of fig. 6. To embed side information bits into NB bitstream, few quantized parameters of GSM AMR coder like codebook index (K), gain (**Θ_0**) and pitch filter coefficient (B) are considered. GSM AMR encoded bitstream is actually arranged according to its subjective importance as mentioned in GSM 05.03 channel coding standards [6].

For embedding and masking of these side information bits in AMR bit rate of 5.9 kbps, last 2 bits of sub frame of codebook index(K), last 2 bits of gain (**Θ_0**) and last bit of pitch filter coefficient (B) have been chosen and toggled using LSB insertion method. In the case of AMR bit rate of 7.95 kbps, last 3 bits of sub frame of code book index, last 2 bits of gain and last 2 bit of pitch filter coefficients and in AMR bit rate of 12.2 kbps,

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

last 4 bits of sub frame of codebook index, last 3 bits of sub frame of gain and last 2 bit of pitch filter coefficient are chosen as per GSM 05.03 channel coding standard using LSB insertion approach. Reason behind not selection of information parameters of AMR coder like: linear prediction coefficient (a), pitch lags (P) for embedding side information is that change in bits of these parameters may severely degrade the quality of recovered speech which has also been witnessed by varying them on trial and error based method. At receiver, Watermark Extraction algorithm receives selected NB bitstream and as per the predefined locations, side information are extracted and used for further process of HB speech reproduction simultaneously for NB legacy AMR NB decoding.

Table 2: Bit allocation of HB Parameters

Parameter	AMR kbps	Resolution	Total bits/frame
LPC Coefficient	5.9	3,3,3,3,2,2,2,2	20 bits
	7.95	4,4,4,4,3,3,3,3	28 bits
	12.2	6,6,5,5,4,4,3,3	36 bits

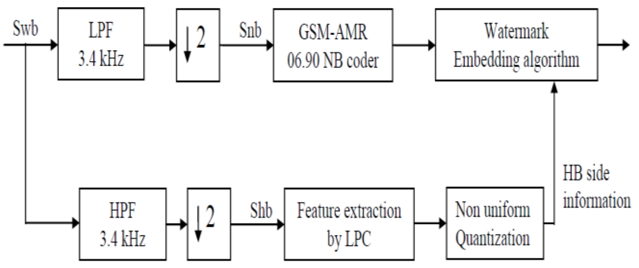

Fig. 6 BWE with side information at Transmitter

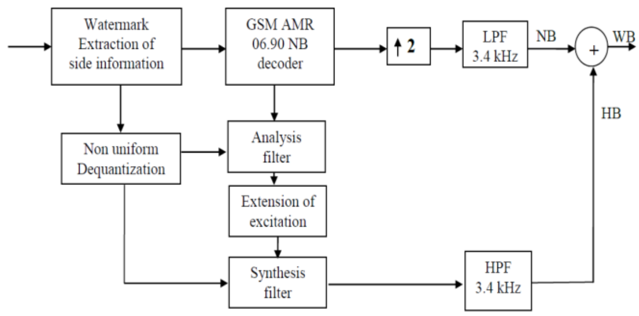

Fig. 7 BWE with side information at Receiver

V. PERFORMANCE EVALUATION AND RESULTS

In this paper, performance of proposed method of BWE with side information is evaluated using both objective and subjective analysis.

A. Subjective Analysis

Here, in Subjective Analysis, Mean Opinion Score rating is carried out for three different clean wave files chosen from wideband corpus [12] as given by cc-17.wav having total no. of samples equal to 64768, cc-21.wav having 52992 samples, cc-24.wav having 106240 samples and cc-27.wav having 125440 samples.

MOS analysis is conducted in quiet environment and with high quality headphones. For this analysis, thirty un-trained listeners are chosen to participate in MOS rating. Out of which fifteen listeners are men and fifteen are women listeners. Each listener is offered with total of eighteen (3 wave files of legacy NB decoded speech output and other 3 wave files of bandwidth extended WB decoded speech for all three sample wave files) wave files. Ratings given by all thirty listeners (for each case) are then averaged to produce final MOS ratings. As observed from Table 3, obtained results for MOS scores clearly advocate the performance of proposed ABE WB coder (for three wave files) over legacy NB decoder. It is evident from the results that MOS scores [7] for all four decoded WB wave files are significantly improved in the case of BW extended WB speech in comparison with the legacy NB decoded speech.

B. Objective Analysis

In the Objective Analysis, Perceptual Evaluation of Speech Quality (PESQ) has been evaluated as per [8]. As can be observed from Table 2, PESQ scores for BW extended speech are quite higher than the PESQ scores of corresponding NB legacy decoded speech. Also it can be witnessed from Table 2 that obtained PESQ scores (Objective Analysis) for all wave files are quite comparable with respect to MOS ratings (Subjective Analysis) computed by listening tests.

Table 3: Subjective and Objective Analysis of the proposed method of BWE with side information

AMR codec Mode	**Speech sample**	**PESQ**		**MOS**	
		Legacy NB decoded speech	**BW extended speech**	**Legacy NB decoded speech**	**BW extended speech**
12.2 kbps	cc-17	2.9181	3.0774	3.43	3.72
	cc-21	2.9574	3.1635	3.46	3.79
	cc-24	3.0064	3.1816	3.55	3.82
7.95 kbps	cc-17	2.9635	3.1643	3.47	3.79
	cc-21	3.100	3.2631	3.64	3.88
	cc-24	3.111	3.3329	3.66	3.93
5.9 kbps	cc-17	3.0715	3.272	3.62	3.89
	cc-21	3.1737	3.4353	3.71	3.98
	cc-24	3.2891	3.4455	3.81	3.97

VI. CONCLUDING REMARKS

Inherent limitation incurred by the NB wired and wireless networks is it offers limited quality of recovered speech. Remedy to it is to somehow improve the intelligibility and naturalness in recovered speech so that overall speech quality increases. Here we proposed a novel approach to extend the bandwidth of recovered speech signal at receiver without modification and up-gradation of current NB transmission channels under varying channel condition using CELP based AMR NB coder. Paper presented ABE algorithm with side information using Linear Prediction method in which HB parameters are scalar quantized, embedded and transmitted into AMR NB bitstream for a given bitrate mode. These parameters, when decoded at receiver, will be utilized to artificially reproduce WB speech. In order to evaluate the performance of developed coder, Subjective (MOS score) and Objective (PESQ score) analysis are conducted on three selected WB speech corpuses. As per the observation from Table 3, both MOS and PESQ scores offer satisfactory values.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

It has also been demonstrated that in comparison with NB legacy decoder for any selected bitrate mode of AMR, ABE WB decoded speech offer significant improvements both in MOS and PESQ scores for all WB input speech corpuses. Also both PESQ (Objective) and MOS (Subjective) analysis results are quite comparable, as could be advocated from obtained results demonstrated in Table 3.

REFERENCES

[1] P. Jax and P. Vary, ―Bandwidth extension of speech signals: A catalyst for the introduction of wideband speech coding?,‖ IEEE Commun.Mag., vol. 44, no. 5, pp. 106–111, 2006.

[2] P. Jax and P. Vary, ―On artificial bandwidth extension of telephone speech, Signal Process., vol. 83, no. 8, pp. 1707–1719, 2003.

[3] Peter jax and Peter vary, ―An upper bound on the quality of artificial bandwidth extension of narrowband speech signals,‖ in Proc. of ICASSP, vol. 1, Orlando, FL, USA, 2002, pp. 237–240.

[4] Eko Pryadi, Kuniwati Gandi, Herman Kanalebe. "Speech Compression Using CELP Speech Coding Technique In GSM AMR", IEEE Conference, 2008A.

[5] Xiao Jianming, Li Xun, Wan Lei, "Software Simulation in GSM Environment of Federal Standard 1016 CELP Vocoder", International Conference on Communication Technology, Oct. 22-24, 1998, Beijing, China

[6] ETSI, Channel coding (GSM 05.03 version 8.9.0 (2005-01), release 1999), pp. 12-19 & 98.

[7] D. Malkovic, ―Speech Coding Methods in Mobile Radio Communication Systems‖, 17th International Conference on Applied Electromagnetics and Communications, Oct-2003, Croatia.

[8] ITU-T 2000, ―Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs‖, ITU-T Rec. P. 862, 2001.

[9] K.Jarvinen "Standardization Of the Adaptive Multi Rate Codec", IEEE Conference 2000.

[10] Bernd Geiser and Peter Vary, ―Backward Compatible Telephony in Mobile Networks: CELP Watermarking & Bandwidth Extension‖, International Conference on Audio and Speech Signal Processing, 2007.

[11] Hannu Pulakka and Paavo Alku, "Bandwidth Extension of Telephone Speech Using a Neural Network and a Filter Bank Implementation for Highband Mel Spectrum" , IEEE traransactions on audio, speech, and language processing, vol. 19, no. 7, september 2011

[12] http://www.repository.voxforge1.org/downloads/SpeechCorpus/Trunk / Audio/Original/16kHz_16bit/

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)