Iraqi Journal of Information and Communication Technology(IJICT)
Vol. 5, Issue 1, April 2022

www.ijict.edu.iq

ISSN:2222-758X
e-ISSN: 2789-7362

A COMPARATIVE STUDY OF SPEECH CODING TECHNIQUES FOR ELECTRO LARYNX SPEECH PRODUCTION

Zinah J. Mohammed [1], Abdulkareem A. Kadhim [2]
[1,2] College of Information Engineering, Al-Nahrain University, Baghdad, Iraq
{Zinah.jaffar, abdulkareem.a}@coie-nahrain.edu.iq [1,2]

Received:4/6/2021, Accepted:10/6/2021
DOI:10.31987/ijict.5.1.185

***Abstract*- Speech coding is a method of earning a tight speech signals representation for efficient storage and efficient transmission over band-limited wired or wireless channels. This is usually achieved with acceptable representation and the least number of bits without depletion in the perceptual quality. A number of speech coding methods have already been developed and various speech coding algorithms for speech analysis and synthesis are used. This paper deals with the comparison of selected coding methods for speech signals produced by the Electro Larynx (EL) device. The latter is a device used by cancer patients with their vocal laryngeal cords being removed. The used methods are Residual-Excited Linear Prediction (RELP), Code Excited Linear Prediction (CELP), Algebraic Code Excited Linear Predictive (ACELP), Phase Vocoders based on Wavelet Transform (PVWT), Channel Vocoders based on Wavelet Transform (CVWT), and Phase vocoder based on Dual-Tree Rational-Dilation Complex Wavelet Transform (PVDT-RADWT). The aim here is to select the best coding approach based on the quality of the reproduced speech. The signal used in the test is speech signal recorded either directly by normal persons or else produced by EL device. The performance of each method is evaluated using both objective and subjective listening tests. The results indicate that PVWT and ACELP coders perform better than other methods having about 40 dB SNR and 3 PESQ score for EL speech and 75 dB with 3.5 PESQ score for normal speech, respectively.**

keywords: **RELP, CELP, ACELP, Speech coding based on wavelet technique, PESQ.**

I. INTRODUCTION

The Speech signal is one of the powerful forms of human contact. It is a particular kind of non-stationary signal which is difficult to form and analyze. The human speech signal is typically sampled at 8 kHz and its band is restricted to 200-3400 Hz and includes complete "speech information" . Coherence, intelligibility, and other parameters play a special role in the speech signals analyzes. Modern technology reveals this fact simply by using various techniques that are used to process an enormous amount of data and increase the information size carried from one point to another [1]. This process is called speech coding. The three major types of speech codes are [2]:

- Waveform codecs with high quality and bit rates.
- Speech voice coders (vocoders) with low synthetic sound and bit rate.
- Hybrid codecs invest from both waveform codecs and speech vocoders techniques and provide a better compromise in quality and bitrate.

Over the past few decades, many algorithms for speech coding have been introduced and simulated. Various speech coding algorithms employ different techniques for speech recognition and synthesis. The great concern is devoted to the speech coder's performance covering the analysis of speech content after decoding [3]. Several researchers have studied and analyzed different speech coders. Some experiments that compare a wide variety of types of vocoder were presented [2]. Comparisons of the preferential test with the natural stimulus revealed that sinusoidal vocoders can provide the superior output of vocoded expression. Multi-dimensional scaling is conducted on the listener responses to analyze similarities in

ISSN:2222-758X
e-ISSN: 2789-7362

terms of quality between the vocoders. The results showed that the vocoders with a sinusoidal synthesis approach are perceptually distinguishable from the source-filter vocoders. A comparative study using different types of speech coding schemes is presented [4]. The performance of the tested coders was validated with a standard English language dataset. CELP coder reproduced the signal more closely to the original signal as compared to other coders. This is because CELP uses long-term and short-term linear prediction models [5]. A comparison study of CELP and Methodology of Parametric Coding methods like Linear Prediction of Mixed Enthusiasm (MELP) is already presented [6]. The analysis of the CELP speech coding technique showed that the technique provided an improvement over Linear Predictive Coder (LPC). It is an efficient coding technique having a bit rate in the range of 9.6 kbps to 16 kbps. The analysis of the MELP coding technique shows that this coder removes the voicing error in the two-state excitation model of LPC. The coder operates at an extremely low bit rate (2.4 kbps) and is mainly used by military and Federal Standards. In [7] a reconstruction of natural-sounding speech from whispers using a speech prosthesis based upon a modified CELP codec is introduced. This uses formant and pitch analysis allied with synthesis and reconstruction methods for missing pitch fundamentals and resonances. Listening tests demonstrate that the resulting speech quality exceeds that of the EL, the current most popular prosthesis for post-laryngectomies patients. The audible evaluation of the resulting samples shows that there is a little hissing around the consonants for the LPC-10e coder [8]. The companding approach assures a very clear distinction between vocals and consonants with just a little bit of hiss addition in steepest transitions for μ-Law. Adaptive Differential Pulse Code Modulation (ADPCM) codec variants are suggested and give similar results of subject appraisal to companding with an enhancement of noise at the end of some words where vocals reside. Emphasized consonants appear in the speech processed by MELP in all parts with a lower level of persistent noise and with no buzziness effect. Due to the recent emergence of machine-learning-based generative models for speech suggests a significant reduction in bit rate for speech codecs, the performance of generative models deteriorates significantly with the distortions present in real-world input signals. A robust speech codec using neural-network-based signal synthesis that encodes speech at 3 kb/s is presented in [9]. This system is suitable for, low-rate video calls, and fits in consumer devices as experimented with by running on a wide range of mobile phones. Experiments show that its quality is similar or better than state-of-the-art conventional codecs operating at double the rate. This paper presents a comparative study and analysis of six speech codecs namely; RELP, CELP, ACELP, Phase vocoder based on DT-RADWT (PVDT-RADWT), PVWT and CVWT. These coders have promising performance for input speech signals and are used here to improve the quality of the output of the EL device. A sample of speech is considered where the quality of the reconstructed version is measured in both subjective and objective senses. The paper is organized as follows: Section II gives an overview of speech coding techniques used in the tests, whereas Section III presents the conducted tests and the results. Finally, the conclusion of the paper is given in Section IV.

II. Speech Coding Methods

According to the automated speech processing model shown in Fig. 1 , speech basic parameters such as pitch for voiced speech and excitation energy and filter coefficients of the vocal tract model are extracted by the analyzer via the use of simple LPC analysis [4]. Fig. 1 represents a common block diagram used by vocoders. As with nearly all speech processing methods, a speech sample can be resembled by linearly combining the prior speech samples within a brief time segment.

ISSN:2222-758X
e-ISSN: 2789-7362

The vocoder formulates the vocal tract as a time-invariant all-pole digital filter. The decision transfer of a voiced/unvoiced and pitch detection utilizes the excitation signals generation. The analysis is carried out on the signal's short temporal (frames). Every 15 to 30ms, the filter parameters are modified. By measuring the spectral envelope with the aid of LPC analysis, the excitation signal is represented. The coding methods used in the work are described next.

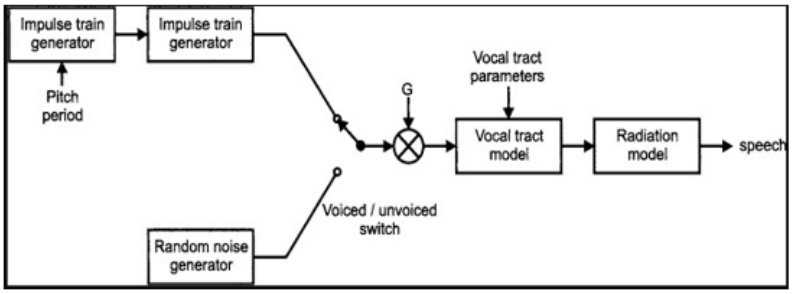

Figure 1: General speech production model [5]

A. Residual-Excited Linear Prediction Coding

An example of prior methods of signal excitation for LPC vocoders is the Residual-Excited Linear Prediction (RELP) coding [2], which is based on linear prediction error (residual) signals for excitation without the need for speech voiced/unvoiced or pitch detection. According to the accuracy of the synthesized speech needed, the RELP vocoder bit rate is ranging between 6 to 9.6 kbps [3]. Its compression rate is reasonable as the RELP vocoder needs a series of residual signals to excite the vocal tract model and synthesize the voice signals [2]. The speech signals are windowed through the Hamming window. The longitude size of the Hamming window is 196 samples. An analysis block of 20 ms is used (136 samples). A block of 30 samples from the next and previous blocks are added to the analysis block to perform overlapping. This overlapping assures ease transitions to prevent sudden shifts within analysis blocks [3]. Fig. 2 shows the RELP vocoder block diagram. The residual error is calculated by subtracting the original speech signal from the synthesized at the encoder. The residual error is quantized, coded, and transmitted to the decoder. At the decoder, the signal is synthesized by adding the residual error to the signal generated from the model. The residual signal is low-pass filtered at 1000 Hz in the analyzer to reduce the bit rate. In the synthesizer, it is rectified and spectrum flattened (using highpass filter), the lowpass and highpass signals are summed and the resulting residual error signal is used to excite the LPC model.

B. Code Excited Linear Prediction Coder

In Code Excited Linear Prediction (CELP) coder the signal is divided into frames of usually 20 ms size to perform the Analysis by Synthesis (AbS) technique. The parameters of the synthesis LPC filter are calculated for each frame followed by defining the excitation signal of this filter. Mainly, the error between the input and the restored speech signals is minimized by the excitation signal when filtered by the LPC synthesis filter [2]. In CELP, as an AbS technique, the receiver is a part of the transmitter. In the absence of channel defects, the receiver would receive a speech signal equal

ISSN:2222-758X
e-ISSN: 2789-7362

to that generated by the transmitter. Splitting the input speech into frames is the first step of CELP encoding [5]. CELP operates with normal low pass filters with vector quantization excitation being used as in most modern coding schemes having a bit rate range of 4 to 8 kbps. This approach is generally used for toll quality at 16 kbps. To emphasize significant frequencies, the error signal is perceptually weighted and reduced by utilizing the excitation signal. The excitation signal is adjusted within the frame over four blocks. The CELP coder has a frame length of 20 ms and a block duration of 5 ms to identify the excitation. The block diagram for the CELP vocoder is presented in Fig. 3 [6]. The encoding and decoding take place at the encoder besides the parameters that minimize the error signal energy. LP analysis is used to get the vocal system impulse response in each frame. The error signal is perceptually weighted to emphasize important frequencies and it is minimized by optimizing the excitation signal. The excitation signal is updated over four blocks within the frame. The encoder needs information about LP coefficients which are gain, pitch filter, pitch delay, and codebook index.

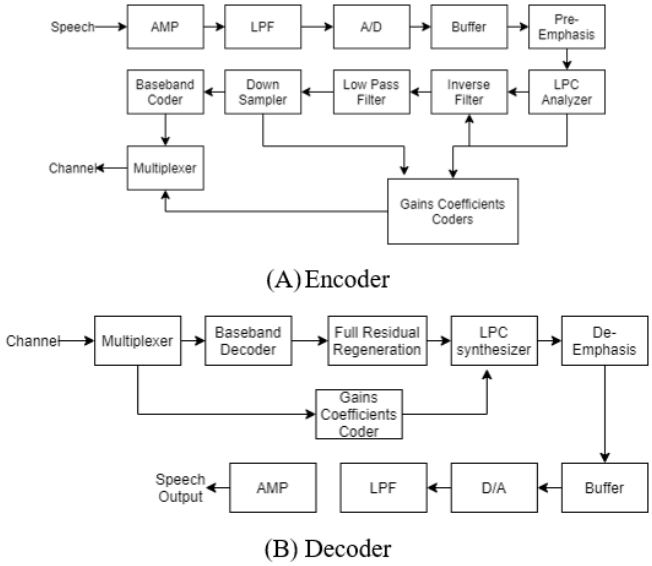

Figure 2: RELP vocoder block diagram

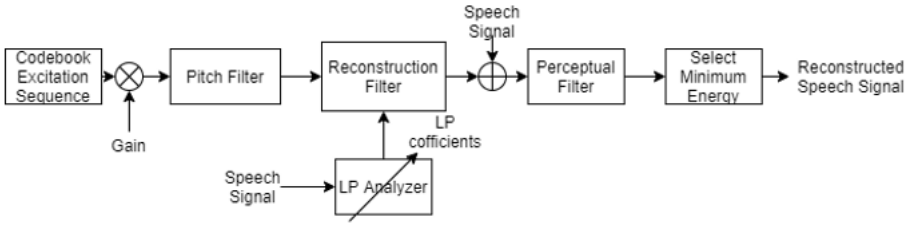

Figure 3: CELP vocoder block diagram

ISSN:2222-758X
e-ISSN: 2789-7362

C. Algebraic Code Excited Linear Predictive Coder

Algebraic Code Excited Linear Predictive (ACELP) coder is a CELP coder variant that has an LP model followed by the AbS algorithm. ACELP such as G.723.1works at rates of 2.4 to 8 kbps [10]. This coder works on two-time scales of 240 sample frames separated into 60-length sub-frames [11]. A collection of parameters is created for each speech frame and sent to the decoder. This means that the frame time reflects a lower device delay limit and the encoder must wait for at least a frame before the encoding process can even begin [12]. The coding process starts by removing the DC component using HPF. Then the Linear Prediction analysis is done for each sub-frame this creates groups of LP coefficients. The output from the HPF is also fed to the formant filter to extract the filter coefficients of the speech formants, which is used to estimate the pitch of the sound. Both of the outputs from the formant filter and the pitch estimation will be used to get the harmonics associated with the speech using the harmonic weighting filter. The decoder on the other works on a frame-by-frame basis in four main steps. The speech parameters received from the coder will be fed to both the LP analysis to reproduce the filters from the coefficients that have been sent after decoding them. The excitation codebook is searched to generate the excitation signal which is then passed through the synthesis filter whose output is input to the formant post filter. The final gain scaling to maintain the energy level of the original signal will be adapted in the final step of the speech reconstruction [13]. Fig. 4 shows encoder and decoder operations in the ACELP.

D. Speech Coding Using Wavelets Transform

The time and frequency representations are essential methods in many signal processing applications ranging from diagnostic data, material and system damage or failure detection, to image and audio processing [14]. The previously mentioned coders are not effective in demonstrating time and frequency resolutions at the same time. For instance, filter banks in spite of providing excellent temporal resolution, they are failed in providing adequate frequency resolution. Wavelet transform resolved the restrictions of former methods by providing time and frequency resolution with reduced temporal resolution [15]. The wavelet transforms of a signal decomposes the original signal into wavelet coefficients at varying scales. The signal in the wavelet domain is represented by coefficients, and all data operations can be performed using only the associated wavelet coefficients [14]. A family of wavelet packets including Haar, Daubechies, Symlets, Coiflets, Meyer ...etc is already available to be used for speech signal analysis. The main challenge is now to select the best wavelet-packet-decomposition (WPD) method which provides the best speech reconstruction [15]. The Symlet is more symmetrical and orthogonal wavelets proposed as modifications to the original Daubechies family. Using Symlet performs better when used in speech signal analysis and has better SNR of the reconstructed signal. Practically it has been verified that any increase in the scale value of wavelet-based speech coder decreases the quality of the speech signal. Three different vocoders based on wavelet transform are considered in this paper. These are described next;

1) Phase Vocoder Based on Wavelet Transform

   PVWT used wavelet transform instead of Short-Time Fourier Transform (STFT) for time-frequency analysis, processing and resynthesize system. Originally, Phase Vocoder (PV) was intended to be a coding method for reducing the bandwidth of speech signals [15]. The coder method starts by windowing a signal into short segments, then the

ISSN:2222-758X
e-ISSN: 2789-7362

wavelet transform using Symlet function with 8 scaling levels is applied to each segment separately and the resulting spectral snapshot can be manipulated in a variety of ways. After the desired spectral changes, the resynthesis is handled by the inverse wavelet transform to return each segment to the time domain. The modified segments are then summed [16]. Fig. 5 shows the PVWT block diagram.

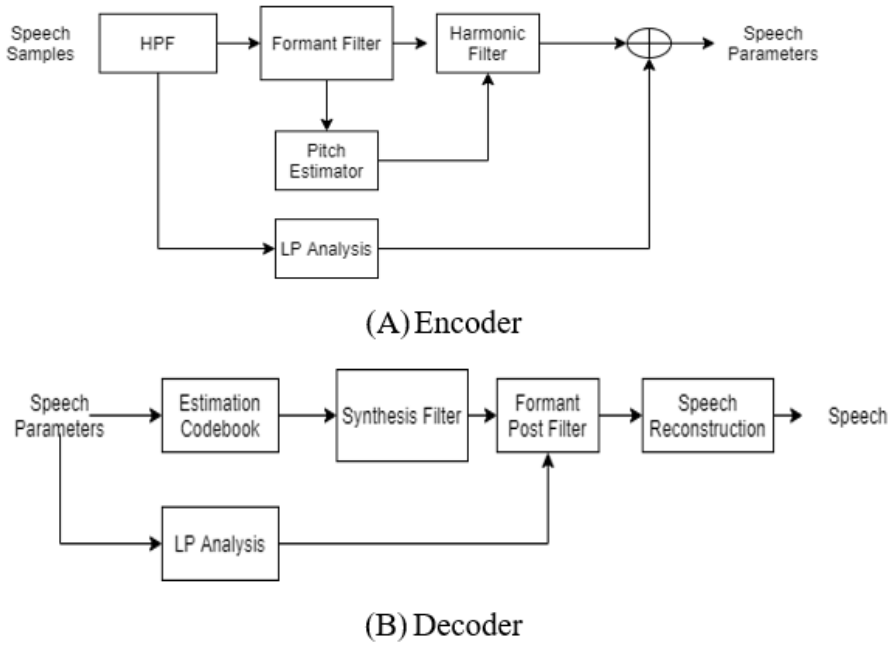

Figure 4: ACELP vocoder block diagram

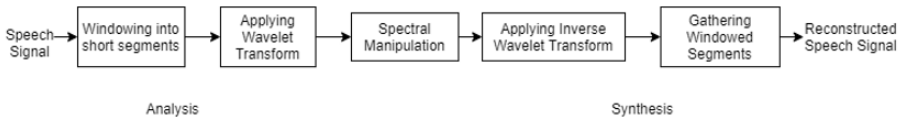

Figure 5: PVWT block diagram [15]

2) Channel Vocoder Based on Wavelet Transform

CVWT is used for bandwidth compression. The inputs are a modulating signal as a voice and a carrier signal, which can be melody or noise. Voice and melody are typically subband encoded through a filter bank, then rectified and followed by a low-pass filter for each subband component to estimate the energy in that particular bandwidth. After mixing the corresponding components of both signals in each subband, a synthesized sound is obtained with almost the same frequency components as the carrier. Given that the synthesized sound has similar envelope as to the modulating part in time domain, the resulting sound is music-like human speech [17]. Fig. 6 shows the block diagram of CVWT.

ISSN:2222-758X
e-ISSN: 2789-7362

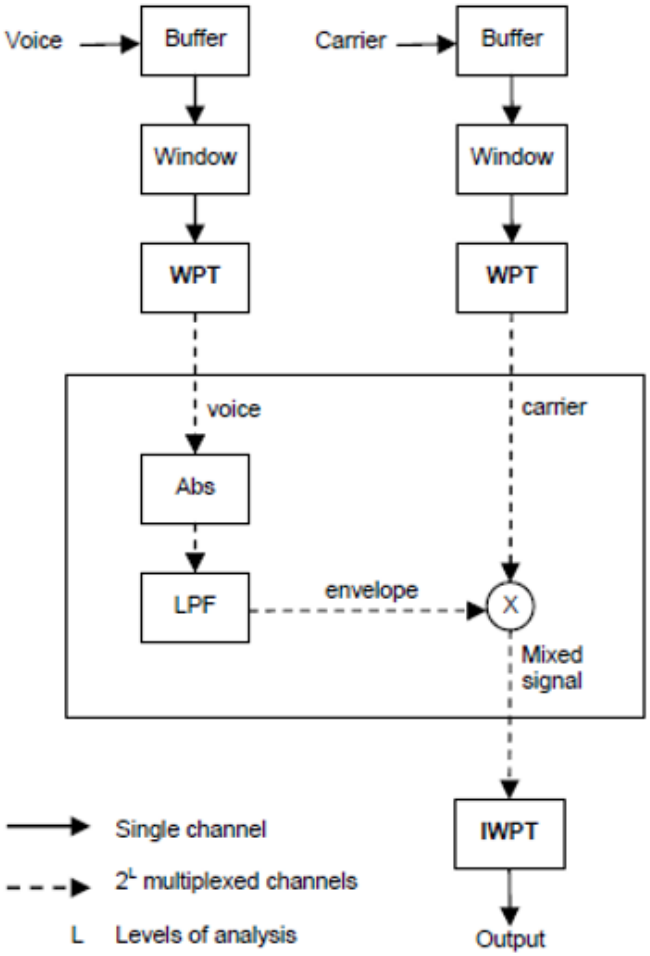

Figure 6: Block diagram of CVWT [17]

3) Dual-Tree Rational-Dilation Complex Wavelet Transform

The Dual-Tree Rational-Dilation complex Wavelet Transform (DT-RADWT) uses time-frequency atom quadrature pairs allowing the analytical signal to work similarly to the STFT using dyadic dual-tree complex wavelet transform [18]. A constant-Q transform property in DT-RADWT that is absent in the STFT makes it more suitable for scale-dependent models. In addition, the frequency resolution can be as high as required [15]. This feature allows DT-RADWT to be more applicable and suitable for oscillatory signal processing such as speech, audio, and diverse biomedical signals [19]. The RADWT's strong frequency resolution and the constant-Q property were inherited by the DT-RADWT. DT-RADWT is not limited to perform analysis or synthesis only, it can be useful for doing for both because of its tight frame property. It can also be used to process signals that are usually required both an analysis and a synthesis scheme. One basic approach is integrating it with a phase vocoder. DT-RADWT is realized by using two wavelet trees, one is the real tree and the other is an imaginary tree, operating in parallel on the same input as shown in

ISSN:2222-758X
e-ISSN: 2789-7362

Fig. 7. In DT-RADWT the second wavelet FBs (FBs of imaginary tree) are designed so that their impulse responses are approximately the discrete Hilbert Transform (HT) of those of the first wavelet FBs (FBs of real tree). Then, to process quadrature signals, this capability of taking the HT property of the imaginary tree in DT-RADWT, can be used to obtain 90 degrees phase shift effect [20]. The Q-factor of the wavelet transform depends on the parameters p, q, and s. Instead of being based on integer dilations, the dilation factor of the transform is q/p where the numbers q and p are coprime and satisfy $q > p$ [19], [20].

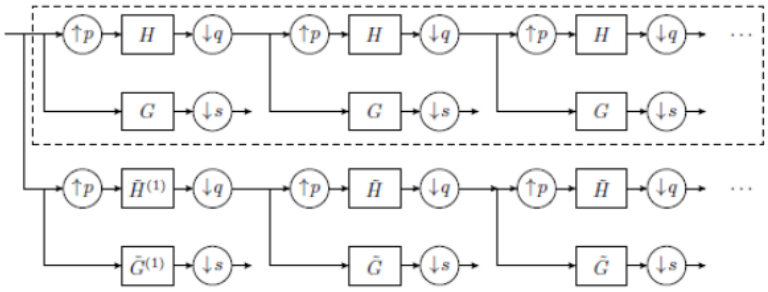

Figure 7: Main DT-RADWT components [19]

III. Simulation Tests and Results

Simulation tests are performed to measure the performance of speech coders mentioned in Section II. The speech codecs are implemented using MATLAB version R2020a. The speech '.wave' file is recorded using a sample of speech signal representing the sentence "Hello, I am OK are you ready" . This sentence is recorded using Intel smart sound technology microphone array at a 44.1 kHz sampling rate. The same sentence is also produced by using an EL device and processed by the considered encoders. The EL is a device that provides periodic mechanical vibration signals to replace the vocal fold vibration for the laryngectomees to reconstruct intelligible speech. The EL has been the most widely used method of speech rehabilitation for laryngectomees due to the advantages of easy learning, easy operating, and continuous output [21]. EL is a commercial small portable with a battery-driven device. The Servox standard model comes with two buttons, both programmable for two settings - one that controls loudness and another that controls intonation. The electrolarynx is applied transcervical and comes with an oral adaptor tube to be used transorally [22]. The device's vibrating coupler disk is placed against the neck. The coupler disk creates a signal that is carried into and filtered by the vocal tract the same as usual speech production [23]. Unfortunately, the EL speech quality and intelligibility are severely damaged by the strong noise presented in the EL speech, especially the radiated noise. In addition, the strong noise in the EL speech also creates difficulties for EL speech processing, such as voice activity detection and EL speech recognition. Thus, noise reduction is imperative to improve the quality and the intelligibility of EL speech [20]. In either case (normal or EL device-based speech), the constructed (decoded) speech signal is compared to the original one to determine the Mean Square Error (MSE). The latter is used to determine the corresponding signal-to-noise-ratio (SNR) in dB as shown in Table I. This represents the objective test. From Table I, it is clear that ACELP and PVWT have the highest SNR values, while RELP

ISSN:2222-758X
e-ISSN: 2789-7362

and CELP got the least SNR values, CVWT acts better than PVDT-RADWT. Fig. 8 presents the input signals and the corresponding decoded (reconstructed) signals for each encoder considered in the tests.

TABLE I
MSE and SNR of Reconstructed Speech for Different Coders

Speech coder	Bit rate (bps)	speech produced by EL Device		Normal speech signal	
		MSE	SNR (dB)	MSE	SNR (dB)
RELP	14705	0.49	8.20	0.45	9.34
CELP	9598	0.48	6.44	0.48	6.65
ACELP	5300	0.21×10^{-3}	40.39	1.42×10^{-6}	75.34
PVWT	7200	0.116×10^{-3}	39.87	0.12×10^{-3}	72.13
CVWT	3200	1.12×10^{-3}	24.34	1.13×10^{-3}	27.33
PVDT-RADWT	4500	1.1×10^{-3}	10.9	9.2×10^{-3}	13.37

Intelligibility and perceptual quality are the main subjective measures considered for speech coding. A subjective test is also considered in this paper. The test is performed using the Perceptual Evaluation of Speech Quality (PESQ) measure. PESQ is a quality assessment algorithm that works on the speech signal sample-by-sample and applies end-to-end testing. The MATLAB PESQ version 2.0 is considered in the test which provides a score in the range of -0.5 to 4.5. The result of the subjective test is shown in Fig. 9. The blue bars represent the EL speech while the red ones stand for normal speech. ACELP and PVWT introduce the highest PESQ score. The score is about 3 for EL device speech and 3.5 for normal speech for ACELP coding, while the corresponding scores for normal speech are about 3 and 3.47, respectively. CVWT performs better than PVDT-RADWT with EL speech, while RELP and CELP have the least PESQ score.

ISSN:2222-758X
e-ISSN: 2789-7362

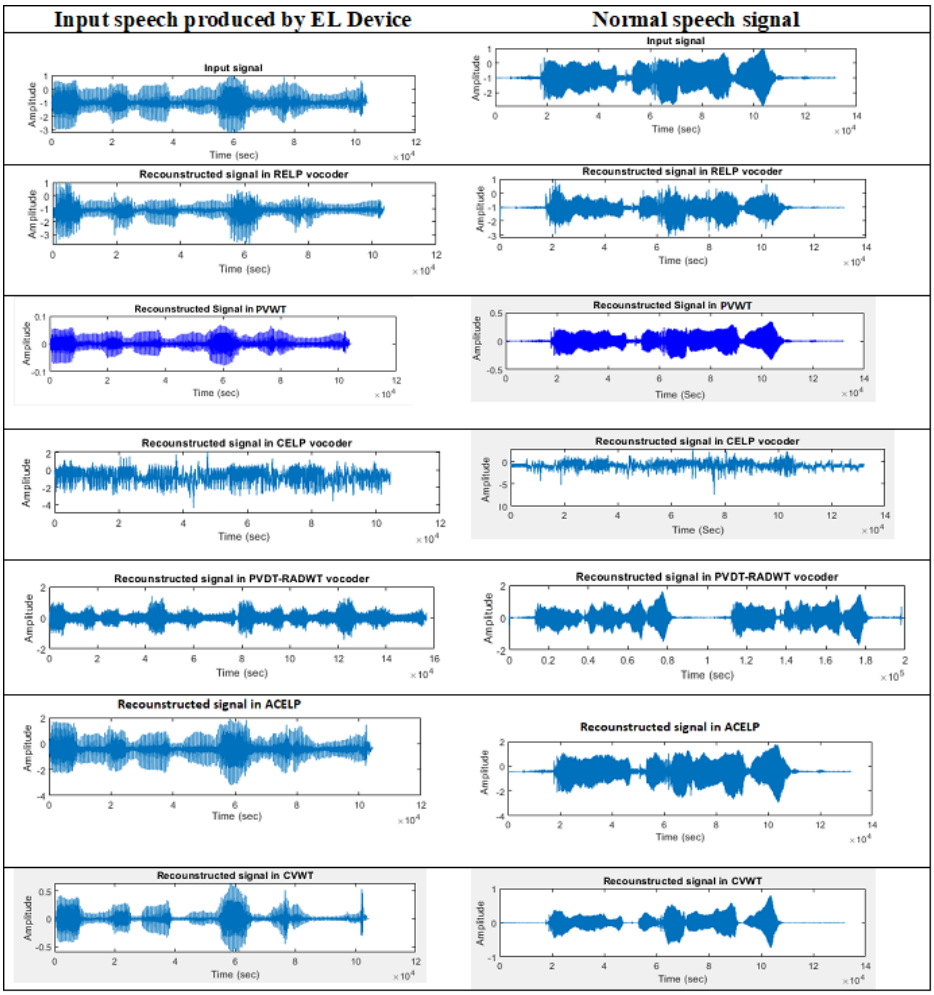

Figure 8: The input and reconstructed speech signal waveforms for all tested coders

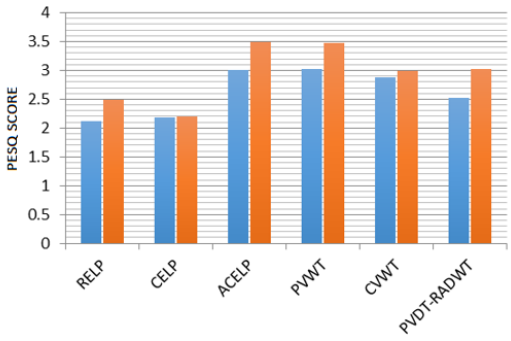

Figure 9: Results of subjective test for all encoders

ISSN:2222-758X
e-ISSN: 2789-7362

IV. CONCLUSION

A comparative study is presented in this paper for the performance of six different speech coding methods when operating on a sample of speech signal produced normally and by Electro-Larynx speech device aiming to adopt the most promising coder to improve the produced speech by such device. The coders considered are RELP, CELP, ACELP, PVWT, CVWT, and PVDT-RADWT. An optimum wavelet-based speech coder requires the selection of a wavelet function that has compact support for both time and frequency in order to minimize the reconstructed error and hence maximize the SNR. Both subjective and objective tests are considered, where the objective test is performed by calculating MSE and the corresponding SNR, while the subjective test considered PESQ algorithm that processes the reconstructed and original speech to produce MOS. The results showed that PVWT and ACELP coders perform better for both test samples: the normal speech and that of Electro-Larynx device than other methods. The SNR for Electro-Larynx speech is about 40 dB and MOS above 3.0. Thus, it is worthwhile to consider these coders to improve the produced speech by the Electro-Larynx device.

ISSN:2222-758X
e-ISSN: 2789-7362

REFERENCES

[1] S. Nagaswamy, "Comparison of CELP Speech Coder with a Wavelet Method" , M.Sc. thesis, University of Kentucky, North America, 2005.
[2] P.Hill, "Audio and Speech Processing with MATLAB" , CRC Press Taylor & Francis Group, University of Bristol, England, 2019.
[3] A. Taguchi, "Residual Excited Linear Predictive Vocoder System with TMS320c-6711 DSK and Vowel Characterization" , M.Sc. Thesis, University of Saskatchewan, Canada, 2003.
[4] Q. Hu, K. Richmond, J. Yamagishi and J. Latorre, "An Experimental Comparison of Multiple Vocoder Types" , 8th International Speech Communication Association (ISCA) Speech Synthesis Workshop, Barcelona, Spain, 2013.
[5] R. Ram, H. Kumar Palo, M. Mohanty and B.N. Sahu, "Speech Coding Techniques: A Comparative Study" , International Journal of Electronics & Communication Technology, Vol. 6, Issue. 3, pp. 29-33, 2015.
[6] R. Jage and S. Upadhya, "CELP and MELP Speech Coding Techniques" , IEEE Wireless Communications, Signal Processing and Networking (WiSPNET) Conference, India, 2016.
[7] H. Sharifzadeh, I. McloughlinIan and F. Ahmadi, "Reconstruction of Normal Sounding Speech for Laryngectomy Patients Through a Modified CELP Codec" , IEEE Transactions on Bio-medical Engineering, Vol. 57, No. 10, pp. 2448-58, 2010.
[8] I. Draganov and S. Pleshkova, "Comparative Analysis of Speech Coders" , Trends in Computer Science and Information Technology journal, Vol. 5, Issue. 1, pp. 1-4, 2020.
[9] W. Kleijn, A. Storus, M. Chinen, T. Denton, F. Lim, A. Luebs, J. Skoglund and H. Yeh, "Generative Speech Coding with Predictive Variance Regularization" , 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, Canada, 2021.
[10] A. Fuchs, M. Hagmuller and G. Kubin, "The New Bionic Electro-Larynx Speech System" , IEEE Journal of Selected Topics in Signal Processing, Vol. 10, Issue. 5, 2016.
[11] K. Xiao , S. Wang, M. Wan , and L. Wu, "Radiated Noise Suppression for Electrolarynx Speech Based on Multiband Time-Domain Amplitude Modulation" , IEEE/ Association for Computing Machinery (ACM) Transactions on Audio, Speech, And Language Processing, Vol. 26, No. 9, 2018.
[12] H. Elsayed, A. Abotaleb, E. Mahmoud and A. Zekry, "CS-ACELP Speech Coding Simulink Modeling, Verification, and Optimized DSP Implementation on DSK 6713" , IEEE International Conference on Promising Electronic Technologies (ICPET) , Gaza, Palestine, 2019.
[13] P. Kabal, "ITU-T G.723.1 Speech Coder: A Matlab Implementation" , Technical Report, Telecommunication and Signal Lab. , McGill University, 2004.
[14] R. Virulkar, A. Khandait, G. Bacher and A. Maidamwar, "Simulation of Conjugate Structure Algebraic Code Excited Linear Prediction Speech Coder" , International Journal of Advanced Computer Technology, Vol. 3, Issue. 3, 2014.
[15] M. Nasief, N. Messiha and H. Mansour, "Performance Evaluation of ACELP CODECs against the Compression Ratio and the Change in the Spoken Language and Accent" , Journal of Electrical Engineering, Vol. 13, No. 12, 2013.
[16] N. Holighaus, G. Koliander, Z. PriuÅ¡a and L. Abreu, "Characterization of Analytic Wavelet Transforms and a New Phaseless Reconstruction Algorithm" , IEEE Transactions on Signal Processing, No. 99, PP. 1-2, DOI:10.1109/TSP.2019.2920611, 2019.
[17] M. Mehrzad, M. Abolhassani, A. Jafari, J. Alirezaie, and M. Sangargir, "Cochlear Implant Speech Processing using Wavelet Transform" , International Scholarly Research Network Journal, Hindawi Publishing, Vol. 2012, pp. 1-6, https://doi.org/10.5402/2012/628706, 2012.
[18] M. Dolson, "The Phase Vocoder: A Tutorial" , Computer Music Journal, Vol. 10, No. 4 , pp. 14-27, 2003.
[19] C. Castaneda, "A Channel Vocoder using Wavelet Packets on a Reconfigurable Device" , Audio Engineering Society Convention (AES) 124th Convention, New York, USA, 2008.
[20] I. Bayram and I. Selesnick, "A Dual-Tree Rational-Dilation Complex Wavelet Transform" , IEEE Transactions on Signal Processing, Vol. 59, Issue. 12, pp. 6251-6256, 2011.
[21] M. Khare, N. Binh Nguyen and R. Srivastava, "Dual Tree Complex Wavelet Transform based Human Object Classification Using Support Vector Machine" , Journal of Science and Technology, Vol. 51, No. 4B, pp. 134-152, 2013.
[22] M. Eshghi, K. Tanaka, K. Kobayashi, H. Kameoka, and T. Toda, "An Investigation of Features for Fundamental Frequency Pattern Prediction in Electrolaryngeal Speech Enhancement" , 10th International Symposium on Computer Architecture (ISCA) Speech Synthesis Workshop, Vienna, Austria, September, 2019.
[23] G. Serbes, H. Gulcur and N. Aydin, "Directional Dual-Tree Rational-Dilation Complex Wavelet Transform" , 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1465-1468, Chicago, USA, 2014.

