IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 10, Issue 4, Ver. II (Jul - Aug .2015), PP 42-48
www.iosrjournals.org

Analysis of Speech Coding Algorithms for Hindi Language

Sukriti Sharma[1], Charu[2]
[1, 2](Department of Electronics and Communication, Manav Rachna College of Engineering, Faridabad, India)

Abstract: *Speech coding is an algorithm used to analyze the special, non- stationary and intelligent speech signal in order to extract its important parameters and to compress it for the maximum utilization of available bandwidth. To achieve this, various speech coding algorithms have been effectively used. Out all these algorithms, Linear Predictive Coding (LPC) is the most powerful one as it provides accurate estimation of speech parameters and is computationally effective and used to represent the speech signal at reduced bit rates while preserving the quality of the signal. Voice- excited LPC is the algorithm proposed in this paper. This algorithm has been implemented using Hindi and English male and female voices and trade- offs between bit rates, delay, power signal to noise ratio and complexity are analyzed. It results in low bit-rates and better signal to noise ratio.*

Keywords: *Bit-Rates, Discrete Cosine Transform, Hindi and English Speech Signal, Linear Predictive Coding, Power Signal to Noise Ratio.*

I. Introduction

Digital transmission is used to provide more flexibility, reliability, privacy, security and cost effectiveness. Due to these reasons, there is a continuous need of digital transmission today in many applications like satellite, radio and storage media like CD ROMS and silicon memory. But now, these applications are band limited. Thus, it is required to reduce the number of bits of transmitted signal.

Speech coding is still a major subject in the area of digital speech processing in which the speech signals are analyzed in order to obtain its important parameters and to compress it to make maximum utilization of available bandwidth. But note that compression of speech signal should be such that it does not harm the intelligibility and quality of transmitted speech signal. To accomplish speech coding practically, number of voice coders or vocoders are employed which can be classified into three: waveform coders, source coders and hybrid coders. Waveform coders operate at high bit rates which lead to very good quality speech. Source coders operate at very low bit rates and reconstructed speech is 'robotic' sounding. Hybrid coders use elements of both waveform and source coders and produces good reconstructed speech at average bit rates. [1]

The vocoder employed here is a source vocoder- modified version of LPC-10. This speech coder is analyzed using subjective and objective analysis. Subjective analysis includes listening of encoded Hindi and English speech signals and making the judgment of its quality which will depend on the opinion of the listener. Objective analysis includes computation of power signal to noise ratio between original and encoded Hindi and English speech signals which will be included within the performance analysis. [2]

II. Technical Approach

The complete cycle of speech production in humans can be summarized as air is pushed up from the lungs through vocal tract and is up through mouth to generate speech as shown in Fig. 1. The air flow from the lungs is called the excitation signal which causes the vocal cords to vibrate which play major role in shaping the sound produced. In technical terms, lungs acts as a source of the speech and vocal tract as a filter that produces different types of sounds that in turn forms a speech.

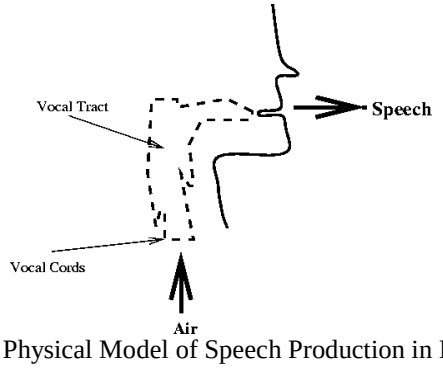

Fig. 1. Physical Model of Speech Production in Humans.

This human speech production model is the model which is used in LPC. The idea behind it is separating source from filter during production of sound and this model is used in both analysis and synthesis part of LPC and is derived from mathematical approximation of vocal tract produced as shown in Fig. 2. The air travelling through vocal tract is the source which can be either periodic for voiced sound produced and random for unvoiced sounds.

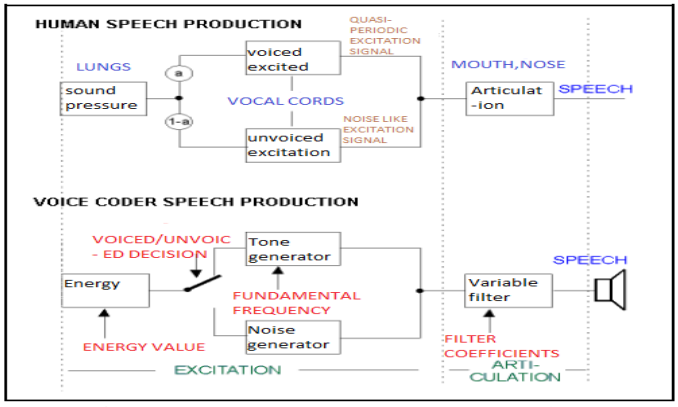

Fig. 2. Human vs. Voice Coder Speech Production.

II.I LPC Model Implementation

The speech signals are analyzed and synthesized using LPC technique which is the method used to estimate the basic parameters like pitch, formants and spectra of input speech signal. The block diagram of LPC vocoder is shown by Fig. 3.

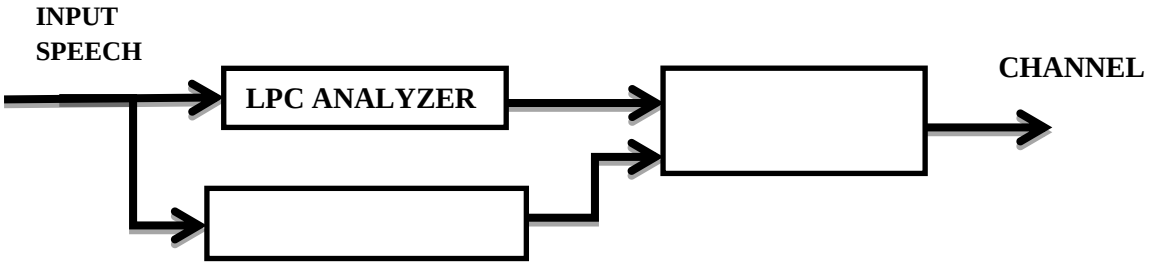

Fig. 3. Block Diagram of LPC Vocoder

II.I.I Sampling

The speech signal is sampled at an appropriate frequency to capture all the necessary frequency components needed for speech processing and recognition. 10 kHz is typically the sampling frequency as most of the speech energy is included in frequencies below 4 kHz (but some women and children violate from this fact).

II.I.II Segmentation

Properties of speech signal change with time. Thus, to process effectively, it is necessary to work frame by frame for which speech is segmented into blocks. The length of the blocks in LPC analysis is between 10ms and 30 ms as within this small interval, the speech signal remains roughly constant.

II.I.III Pre- emphasis

The spectral envelope of speech signal has high frequency roll off due to radiations of sound from lips and these high frequency components have low amplitude that increases the dynamic range of speech spectrum. The speech signal is processed using time- varying digital filter, defined by equation (1).

$$H(z) = 1 - \alpha z^{-1} \quad (1)$$

The filter described in (1) is a pre-emphasis filter which is used to boost the high frequencies in order to flatten the spectrum. Denoting x[n] as input to filter and y[n] as output, the difference equation (2) is applied.

$$Y[n] = x[n] - \alpha x[n] \quad (2)$$

Value of α is near 0.9. To maintain same spectral shape for synthetic speech, it is filtered by de-emphasis filter, defined by equation (3), whose system function is the inverse of pre-emphasis filter.

$$G(z) = 1/(1 - \alpha z^{-1}) \quad (3)$$

II.I.IV Voice detector

The purpose of voicing detector is to determine which frame is voiced or unvoiced. Voice detector is one of the most critical components of LPC coder as misclassification of voicing will result in disastrous consequences on the quality of synthetic speech. A simple voicing detector can be implemented by employing 'Zero Crossing Rate (ZCR)' technique in which if rate is lower than a certain threshold then the frame is considered out to be voiced else unvoiced. ZCR of frame ending at time instant, m is given by equation (4).

$$ZCR(m) = \frac{1}{2\sum_{n=m-N+1}^{m} |sgn(y[n]) - sgn(y[n-1])|} \quad (4)$$

where, sgn (.) is the sign function returning ±1 depending on the operand.

II.I.V Pitch estimation

Pitch or fundamental frequency is one of the most important parameters of speech analysis. Here, autocorrelation function is employed to estimate correct pitch period for voiced or unvoiced frames. If frame is unvoiced then white noise is used with pitch period, T=0 and if frame is voiced, impulse train with finite pitch period, T becomes the excitation of LPC filter as represented by Fig. 4.

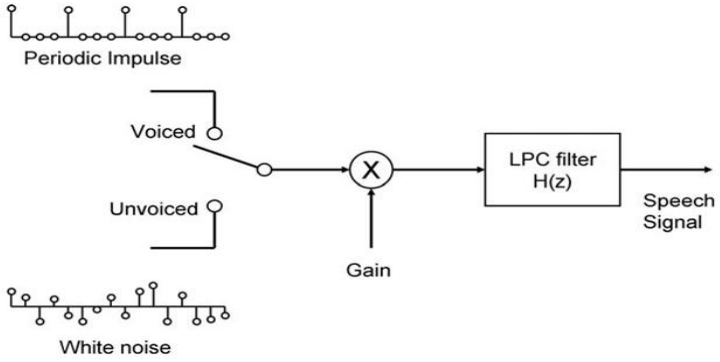

Fig.4. Mathematical Model of Speech Production

II.I.VI Coefficient determination

The prediction coefficients which can be estimated by minimizing the mean square error between the reconstructed and the original speech signal using equations (1) and (2). For efficient estimation, Levinson-Durbin Recursion algorithm is employed.

II.I.VII Gain calculation

For unvoiced case, prediction error is given by equation (5).

$$p = 1/N \sum_{n=0}^{N-1} e^2[n] \quad (5)$$

Where, N as the length of frame

For voiced case, prediction error is given by equation (6).

$$p = 1/[N/T]\, T \sum_{n=0}^{[N/T]T-1} e^2[n] \quad (6)$$

And N is assumed to be N > T.

For unvoiced case, gain (G) is given by equation (7).

$$G = \sqrt{p} \quad (7)$$

For voiced case, the impulse train power having amplitude of G and pitch period, T and interval of [N/T] T must be equal to p.

II.I.VIII Quantization

Usually, direct Quantization of the predictor coefficients is not employed. To ensure stability of the coefficients (the poles and zeros must lie within the unit circle in the z-plane) a relatively high accuracy (8-10 bits per coefficients) is needed. This comes from the effect that small changes in the predictor coefficients lead to relatively large changes in the pole positions. There are two possible alternatives. One of them is the partial reflection coefficients (PARCOR). These are intermediate values during the calculation of the well-known Levinson-Durbin recursion. Quantizing the intermediate values, Line Spectral Frequencies (LSFs) is less problematic than quantifying the predictor coefficients directly as LSFs are less sensitive to quantization noise that ensures more stability. Thus, a necessary and sufficient condition for the PARCOR values is $|k_i| < 1$.

II.II Voice- Excited LPC Vocoder

To improve the quality of sound, voice-excited LPC vocoder is employed. Fig. 5. represents the block diagram of voice-excited LPC vocoder. [4] Its main difference to plain LPC is use of excitation detector instead of pitch detector in plain LPC.

The main purpose behind voice-excited LPC is to avoid the detection of pitch and use of impulse train for synthesizing the speech. Instead, it is better to estimate the excitation signal. As a result, input signal is filtered with the estimated system function of LPC analyser. The filtered signal thus obtained is called residual signal which when transmitted to the receiver will result in good quality. Also, high compression rates can be achieved by computing discrete cosine transform (DCT) of residual signal in which the most of the energy is contained in first few coefficients.

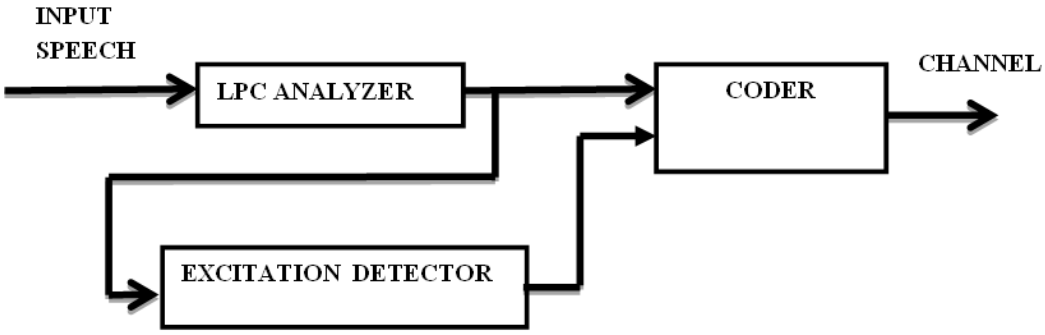

Fig. 5. Voice-Excited LPC Vocoder

III. Comparison Between Hindi and English Speech Signals

All the Indian languages have natural languages that share several features and sounds with the other languages of the world as one cannot expect a language or a group of languages entirely composed of speech sounds that cannot be found anywhere else. Presence or absence of voicing in a speech sound gives rise to distinction of voiced- unvoiced sounds. This basic distinction that is found in English speech signal is employed in Hindi speech signal to a great extent.

Languages differ by the 'amount' of voicing that is present in it. English voiced plosives are considered to be 'partially' voiced as compared to 'fully' voiced plosives. On the other hand, in an Indian language such as Urdu or Hindi, release aspiration does not play a key role in distinguishing unvoiced and voiced plosives. The reason is that these languages maintain a contrast between unvoiced aspirated and unaspirated plosives, whereas English does not have such a contrast. Hindi speech signal utilizes the feature of aspiration to separate their unvoiced aspirates from their unvoiced unaspirates, whereas English speech signal uses the same feature of aspiration to separate its voiced from voiceless plosives. The quality of Voiced sounds in Hindi speech signal is of 'modal' variety. Modal voice is generated by regular vibrations of the vocal folds at any frequency within the speaker's normal range.

Pitch is the fundamental frequency and an important parameter of speech coding. All natural languages use relative variations in pitch to bring out intonational differences like differences between interrogative and declarative sentences or emotional and attitudinal differences on the part of the speaker.

As compared to consonants, vowels of Hindi speech signal do not have those significant different features. Hindi speech signal is supposed to have syllable-timed rhythm whereas English speech signal has a stress-timed. As stress does not have any phonemic value in Indian languages, it does not control the quality as well as the quantity of vowels in a word. Thus, Hindi speech signal does not exhibit drastic changes in the quantity and quality of a vowel which usually depends upon the syllabic stress.

IV. Mean Square Error

The difference between the original and reconstructed speech signal is computed which is called error signal, denoted by 'err' and mean square error (MSE) is computed by taking the average of squares of sample values of err. The value of MSE should be as low as possible and is given by equation (3):

$$MSE = \{\sum err^2\} \tag{8}$$

TABLE I shows the comparison of both Hindi and English Speech Signals in terms of MSE for both Plain LPC and Voice- Excited LPC and it reflects that MSE of English speech signal is more than Hindi speech signal in both LPC algorithms.

Table. I Comparison of MSE for Plain LPC and Voice- Excited LPC using Hindi And English Speech Signals.

Vocoder Type	Hindi Speech Signal	English Speech Signal
Plain LPC	1.0529	1.3623
Voice- Excited LPC	0.00414	0.0051

V. Performance Analysis

Fig. 6. represents the waveforms of Hindi speech signal "मेरा नाम सुकृति शर्मा है, मैं एमटेक ईसीई की छात्रा हूँ|" and Fig. 7. represents the waveforms of English speech signal "My name is Sukriti Sharma, I am from M.Tech ECE" with number of samples in x-axis versus amplitude in y-axis resulted by implementing both of the LPC techniques in MATLAB R20013a.

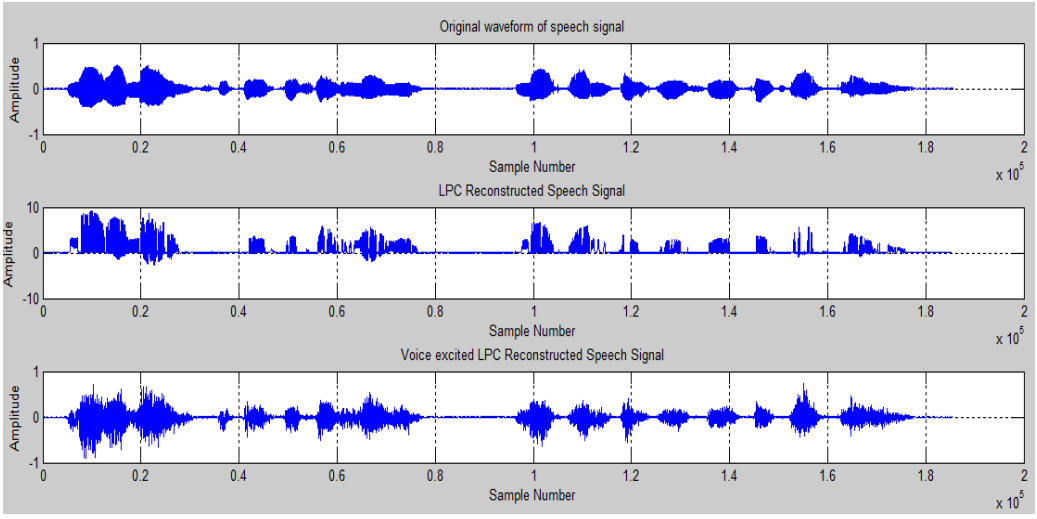

Fig. 6. Waveforms of Hindi speech signal (a) Original speech signal, (b) Plain LPC reconstructed Speech signal and (c) Voice-excited LPC reconstructed speech signal.

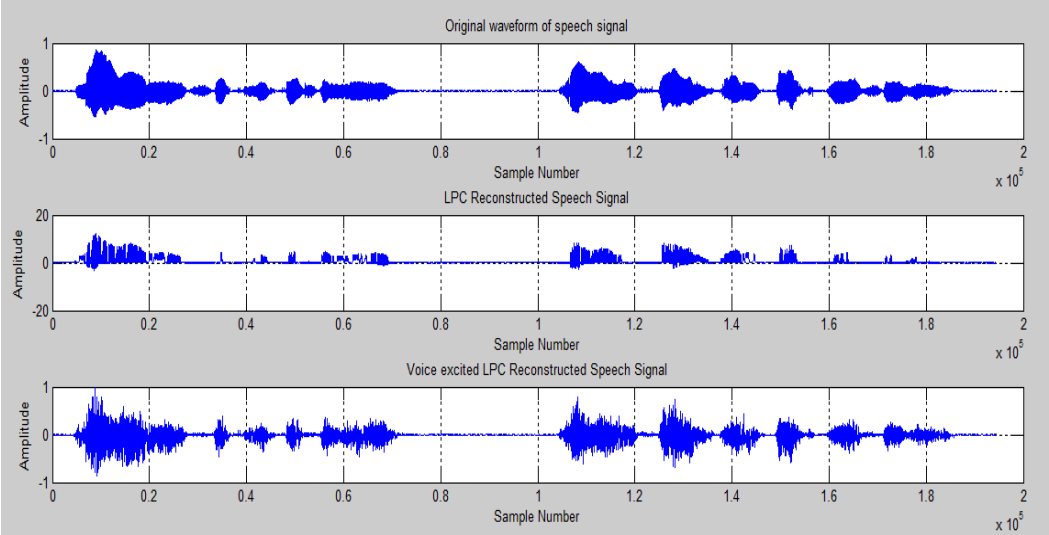

Fig. 7. Waveforms of English speech signal (a) Original speech signal, (b) Plain LPC reconstructed Speech signal and (c) Voice-excited LPC reconstructed speech signal.

Performance analysis is done with subjective and objective analysis where the original Hindi and English speech signals are compared with the plain LPC and voice-excited LPC reconstructed speech signals. In both the cases, Subjective analysis shows that the reconstructed Hindi and English speech signals have lower quality than original speech signal. The plain LPC reconstructed speech signal has low pitch and sound seems to be whispered. But, the reconstructed speech signal of voice-excited LPC appears to be more spoken; less

whispered and appears closer to original speech signal. On other hand, the objective analysis includes following mentioned parameters.

V.I Bit Rates

Bit rates in both the cases are lower than the original speech signal as shown by TABLE II and TABEL III. Here, following parameters are employed:

- Sampling rate Fs = 16000 Hz (or samples/sec.).
- Window length (frame): 20 ms which results in 320 samples per frame by the given sampling rate Fs.
- Overlapping: 10 ms, hence: the actual window length is 30ms or consists of 480 samples.
- There are 50 frames per second.
- Number of predictor coefficients of the LPC model = 10.

Table. II Bit Rates for Plain LPC

Parameters	**Number of bits per frame**
Predictor coefficients	10 bits k1 and k2 (5 each),10 bits k3 and k4 (5 each),16 bits k5, k6, k7, k8 (4 each), 3 bits k9,2 bits k10
Gain	5
Pitch period	6
Voiced/unvoiced switch	1
Synchronization	1
Total	54
Overall bit rate	(54bits/frame)*(50frames/second)= 2700 bits/second

Table. III Bit Rates for Voice- Excited LPC

Parameters	**Number of bits per frame**
Predictor coefficients	10 bits k1 and k2 (5 each),10 bits k3 and k4 (5 each),16 bits k5, k6, k7, k8 (4 each), 3 bits k9,2 bits k10
Gain	5
DCT coefficients	40*4
Synchronization	1
Total	207
Overall bit rate	(207bits/frame)*(50frames/second) = 10350 bits/second

Thus, it is clear that voice-excited LPC needs more than twice the bandwidth needed in plain LPC. This bandwidth increase results in better sound but still not perfect. [5]

V.II Computational complexity

In voice-excited LPC, autocorrelation employed in Plain LPC is omitted and instead DCT and its inverse are employed. But the total number of operations per frame are more in voice-excited than that of Plain LPC. Thus, the improved quality needs higher number of FLOPS (Floating-point Operations per Second). [6]

V.III Power Signal to Noise Ratio

It is given by equation (4).

$$PSNR = 10\log_{10}\{[\max(A)]/MSE\} \quad (9)$$

In equation (4), A is the number of samples of original speech signal. It is found that PSNR of plain LPC using both Hindi and English speech signals is negative that means it is noisier and noise is much stronger than the original signal but for voice-excited LPC, PSNR for both the signals is positive that means it is better but still does not sounds exactly like original speech signal. This is represented by TABLE IV.

Table. IV Comparison of PSNR for Plain LPC and Voice- Excited LPC Using Hindi and English Speech Signals.

Vocoder Type	**Hindi Speech Signal**	**English Speech Signal**
Plain LPC	-5.8592	-2.6750
Voice- Excited LPC	18.1933	21.5750

VI. Conclusion

Speech coding algorithms have been analyzed using two LPC algorithms: Plain LPC and Voice-excited LPC on Hindi and English languages. It has been found that the results obtained from Voice-excited LPC using English speech signal are more intelligible as compared to Hindi speech signal whereas from Plain LPC, the results are poor and barely intelligible for both English and Hindi speech signals. But through Voice-excited LPC, the improved quality of compressed reconstructed speech signal requires more number of bits per frame that leads to increased bandwidth requirement. Also, SNR for both the algorithms using Hindi and English Speech Signals were computed and compared and it has been found that sound of reconstructed speech signal due to Plain LPC has negative SNR for each language that results in noisy and whispered sound. On the other hand, Voice-excited LPC has far better sound and positive SNR for both Hindi and English speech signals. Since, the voice-excited LPC gives pretty good results with all the required limitations, and we can try to improve it. A major improvement can be the compression of the errors. If we send them in a lossless manner to the synthesizer, the reconstruction would be perfect.

References

[1]. L.R.Rabiner and R. W. Schafer, "Theory and Application of Digital Speech Processing Preliminary Edition".
[2]. The newest breeds trade off speed, energy consumption, and cost to vie for an ever bigger piece of the action. BY JENNIFER EYRE Berkeley Design Technology Inc.
[3]. B. S. Atal, M. R. Schroeder, and V. Stover, "Voice- Excited Predictive Coding System for Low Bit-Rate Transmission of Speech", Proc. ICC, pp.30-37 to 30-40, 1975.
[4]. M. H Johnson and A. Alwan, "Speech Coding: Fundamentals and Applications", to appear as a chapter in the encyclopedia of telecommunications, Wiley, December 2002.
[5]. Sukriti Sharma, Charu, "Lossless Linear Predictive Coding For Speech Signals", International Journal of Science, Technology & Management, Volume No 04, Special Issue No. 01, March 2015, ISSN (online): 2394-1537.
[6]. Orsak, G.C rt al, "Collaborative SP education using the internet and MATLAB" IEEE Signal Processing Magazine, Nov, 2009 vol 12, no6, pp 23-32.
[7]. H. Huang, H. Shu, and R. Yu, "Lossless Audio Compression In The New IEEE Standard For Advanced Audio Coding", IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP) 2014, pp 6984-6988.