Qualitative Spectral Parameter Coding for Hindi and English Speech Signals

Sukriti Sharma
Department of electronics and communication
Manav Rachna College of Engineering
Faridabad, India.
sukritisharmafbd@gmail.com

Charu
Department of electronics and communication
Manav Rachna College of Engineering
Faridabad, India.
charup123@gmail.com

***Abstract*— Speech signals are unique and special signals in a communication system due to their non-stationary and intelligent characteristics, thus they must be analyzed in order to extract their important parameters which maintain their quality. This analysis is also useful to sustain quality with a good compression ratio. To accomplish this, various kinds of 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. In this paper Voice- excited LPC and Qualitative Multi- Band Excitation (QMBE) algorithms are implemented on spoken Hindi language. The extracted parameters are compared with those achieved from the compression of spoken English language.**

Keywords- Bit-Rates; Discrete cosine transform; Linear Predictive Coding; Mean Square Error; Qualitative Multi- Band Excitation; Power Signal to Noise Ratio; Spoken Hindi and English Language.

I. INTRODUCTION

Now days, communication is carried out using digital transmission techniques. Digital communication provides more flexibility, reliability, privacy, security and cost effectiveness. It is widely used in satellite, radio and storage media like CD ROMS and silicon memory. The transmission is band limited and hence it is required to use minimum number of bits for coding the signals. At the same time the quality of the signal should not be compromised as the application demands. Similarly compression also reduces memory requirements for storage. In voice communication, various coding techniques have been developed for compression. Coding of voice signals is still a challenging task so that the compression does not harm the intelligibility and quality of the transmitted signal. To accomplish voice coding, a number of voice coders or vocoder algorithms have been developed like waveform coders, source coders and hybrid coders. [1]

Here, the modified versions of Linear Predictive Coder, and Qualitative Multi- Band Excitation (MBE) vocoder. These speech vocoders are 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. These analysis have been carried out both at Hindi and English spoken language. The voice signals have been recorded in the recording room of Manav Rachna Radio.

II. DIFFERENCE BETWEEN SPOKEN HINDI AND ENGLISH LANGUAGE 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.

III. LINEAR PREDICTIVE CODING VOCODER

In the standard Linear Predictive vocoder, the speech signals are analyzed and synthesized using LPC algorithm. This algorithm is used to estimate the basic parameters like pitch, formants and spectra of a speech signal. The LPC vocoder comprises of three stages of algorithms comprising of LPC analyzer, Pitch analyzer and a coder as shown in Fig.1.

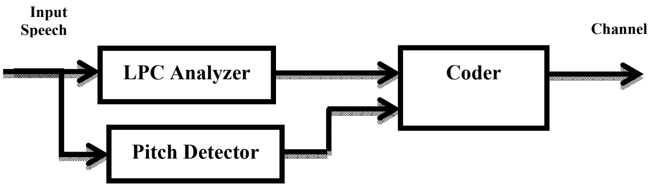

Fig 1. LPC Vocoder.

LPC vocoder uses the principle to minimize the sum of mean squared difference between original speech signal and reconstructed speech signal over a finite duration. [3] This results in a unique set of predictor coefficients. These coefficients have been estimated by dividing the signal into frames of 20ms each. These predictor coefficients are denoted by a_k and the gain is denoted by G, thus the transfer function of time-varying digital filter is given by the expression in equation 1.

$$H(z) = \frac{G}{1-\sum_{k=1}^{p} a_k z^{-1}} \quad (1)$$

In equation (1), p denotes the number of coefficients considered in LPC analyzer. To compute the predictor coefficients, autocorrelation is used. It ensures the stability of the system by making the poles of the system fall inside the unit circle. The parameters of autocorrelation are computed using Levinson- Durbin Recursion algorithm. The LPC analysis decides the category of the sound is voiced or unvoiced in each frame. If sound is voiced than impulse train with non-zero pitch period, T is used to represent it and if sound is unvoiced than white noise with pitch period, T=0. Thus, either impulse train or white noise is the excitation of LPC filter. A pre-emphasis filter is used to boost up the high frequencies in order to flatten the spectrum. Pre-emphasis filter is given by equation 2:

$$y(n) = x(n) - \alpha x(n-1) \quad (2)$$

A. *Quantization of LPC Coefficients*

The values of predictor coefficients when varied by a small value lead to relatively large changes in the pole positions. Thus, to ensure stability of the coefficients that is the poles and zeros lie within the unit circle in the z-plane, a relatively high accuracy is required. Thus, 8-10 bits per coefficient should be used for coding. Partial reflection coefficients (PARCOR) are used for quantization. These are intermediate values during the calculation of the well-known Levinson-Durbin recursion. Quantizing the intermediate values, Line Spectral Frequencies (LSFs) is less sensitive to quantization noise and thus ensure more stability. The quantization of LSFs results in 25-30% more reduction in number of bits of input speech signal without degrading its quality. This gives a necessary and sufficient condition for the PARCOR values k*i*, given by equation 3. [4]

$$|k_i| < 1 \quad (3)$$

B. *Voice- Excited LPC vocoder*

Voice-excited LPC vocoder is used to retain the intelligibility of the coded signal. [5]
The Voice-excited LPC avoids the detection of pitch and use of impulse train for synthesizing the speech signal. It is superior to estimate the excitation signal. In Voice excited LPC, the input signal is filtered with the estimated transfer function of LPC analyzer. This filtered signal is called the residual signal. It ensures the quality of the signal when transmitted to the receiver. The high compression ratio has been achieved by computing discrete cosine transform (DCT) of residual signal as in DCT; most of the energy is contained in first few coefficients. [6]

IV. QUALITATIVE MULTI-BAND EXCITATION (QMBE) VOCODER

QMBE vocoders remove buzziness from the signal. QMBE divides the spectrum into a number of frequency bands. It uses 20 or more frequency bands as compared to their frequency bands of the other vocoders. A binary voiced/unvoiced parameter is used to define each band.

QMBE algorithm models the speech spectrum as a product of excitation spectrum and spectral envelope. QMBE vocoder comprises of two processes namely Speech Analysis and Speech Synthesis. In Speech Analysis, estimation of all of the model parameters like the pitch period and spectral envelope parameters are estimated. These parameters are used to minimize the error, $\in$ between the original spectrum $|s_w(w)|$ and the synthetic spectrum $|s'_w(w)|$, as represented by equation 4. Then, the V/UV decisions are made based on the closeness of fit between the original and the synthetic spectrum at each harmonic of the estimated fundamental.

$$\in = \frac{1}{2\pi}\int_{-\pi}^{\pi} G(w)\,[|s_w(w)| - |s'_w(w)|]^2\,dw \quad (4)$$

Where,

$$|s'_w(w)| = |H_w(w)||E_w(w)| \quad (5)$$

G(w) is a frequency dependent weighting function. $|H_w(w)|$is a spectral envelope and $|E_w(w)|$ is an excitation spectrum.The voiced/unvoiced decision for each harmonic is computed by comparing the normalized error over each harmonic of the estimated fundamental to a threshold. If the normalized error is below the threshold, this region of the spectrum matches that of a periodic spectrum well else, region of the spectrum is assumed to contain noise-like energy.

A. *Synthesis in QMBE*

The process of Speech Synthesis includes categorization of samples into voiced or unvoiced samples depending on whether they are declared voiced (include both magnitude and phase) or unvoiced (include only magnitude) as in Fig. 2. The outputs of the bank of sinusoidal oscillators running at the harmonics of the fundamental frequency are then summed together. Then, the Voiced speech is synthesized from the voiced envelope samples as shown in Fig. 3. By synthesizing a white noise sequence, unvoiced speech is synthesized from the unvoiced envelope samples as shown in Fig. 4. Then the synthesized speech is obtained by summing the voiced and unvoiced synthesized speech signals as shown in Fig. 5.

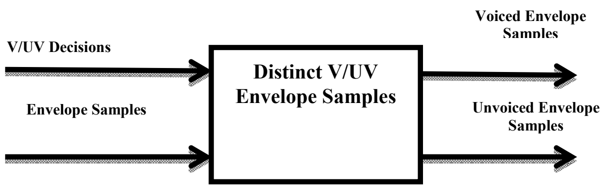

Fig. 2. Separation of Envelope Samples

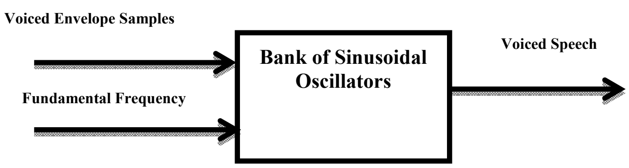

Fig. 3. Voiced Speech Synthesis

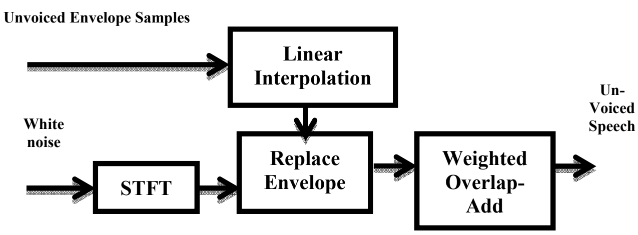

Fig. 4. Unvoiced Speech Synthesis

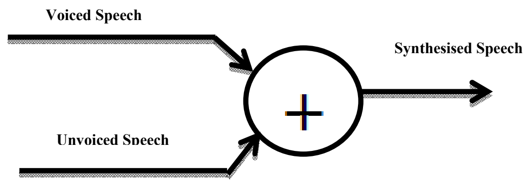

Fig. 5. Speech Synthesis

V. Implementation and Results

The speech coding techniques described above were implemented and tested on the Hindi and English spoken signals. The text of speech is:

Female voice (English): My name is Sukriti Sharma, I am from M.Tech ECE.

Female voice (Hindi): Mera naam Sukriti Sharma hai, Mai M.Tech ECE ki chatraa hun.

Fig. 6 represents the waveforms of Hindi speech signal, “Mera naam Sukriti Sharma hai, Mai M.Tech ECE ki chatraa hun” 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 and QMBE.

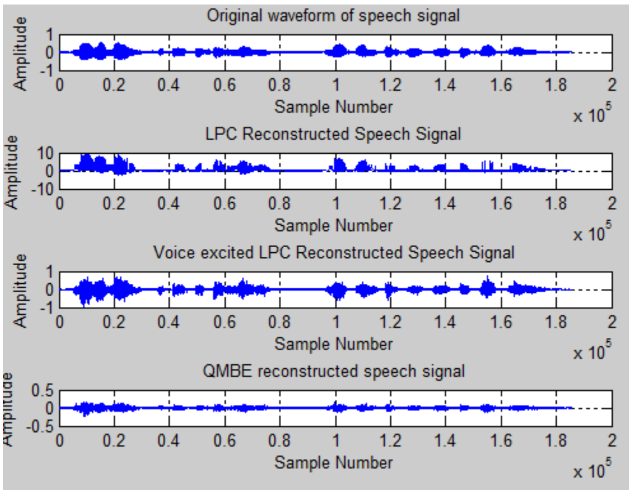

Fig. 6. Waveforms of Hindi speech signal (a) original speech signal, (b) plain LPC reconstructed speech signal and (c) voice-excited LPC reconstructed speech signal.

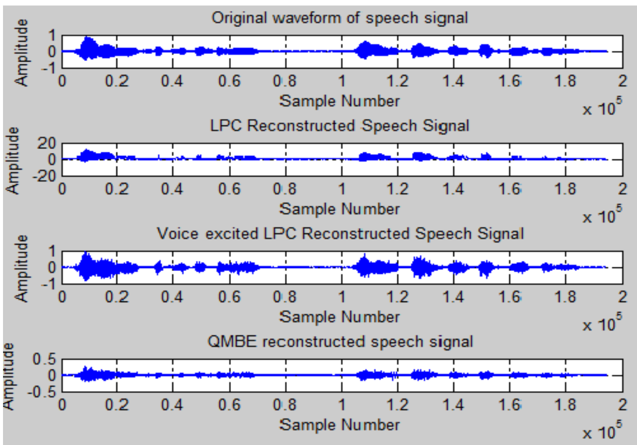

Fig. 7. Waveforms of English speech signal (a) Original speech signal, (b) Plain LPC reconstructed speech signal and (c) Voice-excited LPC reconstructed speech signal.

Subjective and objective analysis of both the languages has been done. The original Hindi and English speech signals are compared with the plain LPC, voice-excited LPC and QMBE reconstructed speech signals. In all 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. And, the reconstructed speech signal of QMBE is more intelligent than both of the LPC algorithms and similar to input speech signals. On other hand, the objective analysis includes following mentioned parameters.

A. *Bit Rates*

Bit rates in all the three cases are lower than the original speech signal as shown by TABLE I, TABEL II and TABLE 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 I. 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	6
Voiced/unvoiced switch	1
Synchronization	1
Total	54
Overall bit rate	(54bits/frame)*(50frames/second) = 2700 bits/second

TABLE II. Bit rate for voice-excited LPC vocoder with DCT

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

TABLE III. Bit rate for QMBE

Parameters	**Number of bits per frame**
Fundamental Frequency	9
Harmonic Amplitudes	139-94
Harmonic Phases	0-45
V/UV Bits	12
Total	160
Overall bit rate	(160bits/frame)*(50frames/second) = 8000 bits/ second

Thus, it is clear that from the tables 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. But, QMBE needs lower bandwidth and sounds better than voice- excited LPC. [6]

B. *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 6:

$$MSE = \{\sum err^2\} \tag{6}$$

TABLE IV. Comparison of MSE for Plain LPC and Voice- Excited LPC using Hindi and English Speech Signals.

Vocoder Type	**Hindi Speech Signal MSE**	**English Speech Signal MSE**
Plain LPC	1.0529	1.3623
Voice- Excited LPC	0.0080	0.0081
QMBE	0.0050	0.0051

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

C. *Power Signal to Noise Ratio*

It is given by equation (7).

$$PSNR = 10\log_{10}\{[\max(A)]/MSE\} \tag{7}$$

In equation (7), 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 and QMBE, PSNR for both the signals is positive that means it is better but still does not sounds exactly like original speech signal. And, while comparing Hindi and English speech signals, Hindi speech signal has lower PSNR values than English speech signal. This is represented by TABLE V.

TABLE V. Comparison of PSNR for Plain LPC and Voice- Excited LPC using Hindi and English Speech Signals.

Vocoder Type	**Hindi Speech Signal PSNR**	**English Speech Signal PSNR**
Plain LPC	-5.8592	-2.6750
Voice- Excited LPC	16.2433	19.5781
QMBE	18.1933	21.5750

VI. CONCLUSION

Speech coding has been achieved both for Hindi and English Language signals. After implementing different speech coding techniques, it has been found that the results obtained from QMBE have high quality and are more spoken and intelligent in noisy and clean in both English and Hindi speech signals as compared to both LPC algorithms which are having the disadvantage of "BUZZINESS". But, as compared to English speech signal, Hindi speech signal in all the three algorithms have low PSNR i.e. more whispered and noisy. Thus, we can conclude that the implemented vocoders well for English language as compared to Hindi language.

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] John Makhoul, S. Roucos, "Vector Quantization in Speech Coding" In IEEE Transactions on 28 (1), 27-34, 1980.

[5] M. H Johnson and A. Alwan, "Speech Coding: Fundamentals and Applications", to appear as a chapter in the *encyclopedia of telecommunications*, Wiley, December 2002.

[6] 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.

[7] 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.

[8] 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.