Comparative Study of PCM, LPC, and CELP Speech Coders Used for VoIP Applications

Mahesh Chandra and Manas Ray

Abstract The quality of the speech signal in a voice over internet protocol (VoIP) is governed by the speech coding technique employed. Currently, various standard coders such as FS-1015 (LPC-10), ITU-G.711, and FS-1016 (ITU-G.728) are used to digitize the speech signal. This paper analyzes the performance of the above coders by comparing the coding capabilities of the coders on two Hindi and two English language sentences. The performance is then evaluated in terms of compression ratio (CR), peak signal-to-noise ratio (PSNR), and normalized root mean square error (NRMSE).

Keywords VoIP · LPC-10 · G.711 · G.728 · CR · PSNR · NRMSE

1 Introduction

The exponential growth and wide acceptance of voice over IP (VoIP) in various applications of communication technology is the driving force behind the evaluation of telephony technologies in the recent years and is fast becoming a dominant service for the overall telephony industry [1]. This had led to the rapid deployment of VoIP services over the recent years and had opened up various quality-of-service (QoS) challenges in the deployment, especially in the microwave domain [2]. The central component of the VoIP services delivery is the voice coder or the codec deployed for the application. There are different types of codec used, based on the application, complexity, bandwidth requirement, etc. Currently,

M. Chandra (✉) · M. Ray
Department of Electronics and Communication Engineering,
Birla Institute of Technology, Mesra, Ranchi 835215,
Jharkhand, India
e-mail: shrotriya69@rediffmail.com

M. Ray
e-mail: manas_ray.me@rediffmail.com

various ITU/FS standard codecs are deployed for the VoIP applications. These codecs were initially designed and deployed in the legacy public switched telephone network (PSTN); however, due to the inherent qualities of the VoIP application to adapt to the existing infrastructure, these codecs have been seamlessly integrated to VoIP systems and have become the central theme of VoIP, dictating the QoS requirements. The most recent codec deployments in the VoIP setup are FS-1015/LPC-10, ITU-G.711, ITU-G.728, ITU-G.729.1, etc. The data rate generated by VoIP codecs differs based on the engineering trade-off between voice quality, available bandwidth, and complexity of the codec [3].

This paper attempts to analyze the working of the existing codecs by testing their characteristics on various sentences of a different speaker's speech in Hindi and English languages. The paper is organized into five sections. Section 1 introduces the motivation and nature of experimental work. The different types of codecs deployed in the VoIP applications along with a brief background on their signal processing techniques are presented in Sect. 2. The parameters measured for the performance evaluation of the codecs are defined in Sect. 3. The simulation results of the codecs and their response to the various Hindi and English sentences are presented in Sect. 4, and finally, the conclusions are drawn in Sect. 5.

2 Standard Codecs for VoIP Applications

There are various standard codecs that are used for speech coding applications. Some of them are pulse code modulation (PCM), differential pulse code modulation (DPCM), adaptive delta pulse code modulation (ADPCM), linear predictive coding (LPC), code excited linear predictive coding (CELP), etc. In this paper, three different speech coders are discussed for VoIP applications.

2.1 Pulse Code Modulation (ITU-G.711)

PCM is speech codec where the input speech is passed through a low-pass filter of bandwidth 4 kHz, and the resulting output is sampled at a rate of 8 kHz to generate a train of pulses as samples. These samples are then quantized, and each sample is encoded with 8-bit binary code. These binary codes are then transmitted from the transmitter to the receiver. At the receiving end, the binary codes are converted into the original speech with the help of digital-to-analog converter and a low-pass filter. It is a waveform-type speech coding technique where the waveform of the synthesized speech signal agrees with that of the original speech signal waveform. The operating bandwidth of the coder is 64 kbps. The reproduced speech signal has the highest quality, called as toll quality.

2.2 Linear Predictive Coding (LPC-10/FS-1015)

LPC is a digital method of speech coding in which the speech signal at a particular instant is represented by a linear sum of the 'p' previous samples [4]. This process is called as autoregressive process of speech coding. It is a lossy form of speech compression which operates at a data rate of 2.4 kbps. LPC is based on the concept of parametric coding, where the human vocal tract is mathematically approximated as a tube with varying diameter. Figure 1 provides an analogy of LPC voice coder with that of the human speech production. LPC is used to estimate the basic speech parameters such as voice/unvoiced decision, pitch, and formant. In the analysis stage of the LPC coding, the speech samples are broken down into segments or blocks, and each segment is then analyzed to determine the following:

(a) Nature of signal as voiced or unvoiced?
(b) Pitch of the signal
(c) The vocal tract filter is analyzed as per the equation

$$y_n = \sum a_i y_{i-1} + G\varepsilon_n \tag{1}$$

where y_n is the output and ε_n is the input and a_i and G are the parameters of the vocal tract filter which need to be estimated. These parameters are passed by the encoder to the decoder of the synthesis stage to reconstruct the signal. At the decoder end, the vocal tract filter is estimated from the parameters received from the analysis stage. The voice quality generated by the codec is of robotic nature, and the trade-off is the low bit rate employed for the coding purpose.

2.3 Code Excited Linear Prediction (ITU-G.728/FS-1016)

CELP is a hybrid coding technique which utilizes the features of the parametric coding as well as that of the waveform coding techniques in order to provide a robust low-bit speech coder [5]. In the analysis stage, the speech signal is passed through a cascade of formant predictor filter and pitch predictor filter. The formant predictor filter removes sample-to-sample correlation, and pitch predictor filter removes the long-term correlations. The residual signal $r(n)$ noise like signal is compared with the entries of the code book and the index of the best-matched entry is selected. The parameters of the filters along with the index value of the codebook representing the residual filter are passed on to the synthesizer. Figure 2 provides the block diagram of the CELP analysis. In the synthesis stage, the excitation waveform is chosen from a dictionary of waveforms which drives a
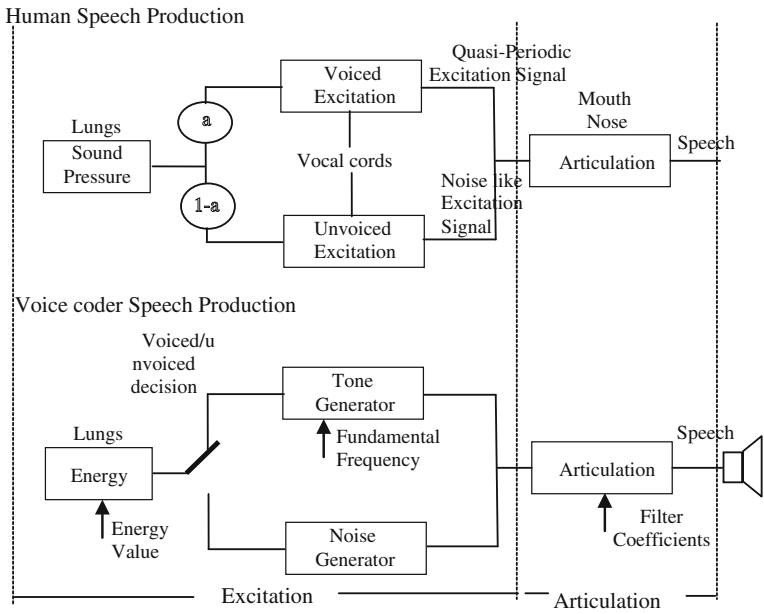
Human Speech Production

Fig. 1 Block diagram of LPC speech coder
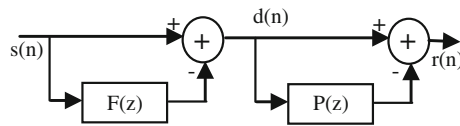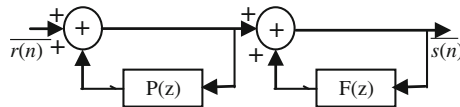

Fig. 2 CELP analysis stage

Fig. 3 CELP synthesis stage

cascade of filters synthesized from the parameters received from the analysis stage to approximate the input speech signal. Figure 3 provides the block diagram of the CELP synthesis stage.

3 Performance Evaluation of Speech Codecs

The performance evaluation of the codecs is carried out by objective testing for six speech samples containing both male and female voices of Hindi and English languages. The subjective test for each sample can also be carried out by mean opinion score, and the objective tests were carried out by evaluating the performance in terms of compression ratio (CR), SNR, PSNR, and NRMSE [6, 7]. The expressions of these parameters are given below:

$$CR = \frac{\text{Length of } (x(n))}{\text{Length of } (r(n))} \tag{2}$$

$$SNR = 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) \tag{3}$$

$x(n)$ and $r(n)$ are the original and reconstructed signals, respectively. σ_x^2 and σ_e^2 are mean square of the speech signal and the mean square difference between the original and reconstructed signals, respectively. Peak signal-to-noise ratio (PSNR) and normalized root mean square error (NRMSE) can be given by

$$PSNR = 10\log_{10}\frac{NX^2}{\|X - r\|} \tag{4}$$

$$NRMSE = \sqrt{\frac{(x(n) - r(n))^2}{(x(n) - \mu x(n))^2}} \tag{5}$$

In Eq. 4, N is the length of the reconstructed signal, X is the maximum absolute square value of the signal x, and $\|X-r\|^2$ is the energy difference between the original and reconstructed signals. Where $x(n)$ is the speech signal, $r(n)$ is the reconstructed signal and $\mu x(n)$ is the mean of the speech signal.

4 Experimental Setup

MATLAB simulation model is prepared for PCM [8], CELP [9], and LPC [10] as per the standard available. A total of four sentences were used for testing the performance of PCM, LPC, and CELP codecs as given in Table 1.

Table 1 Details of the sample sentences used in the experiment

| Sample No.	Sentence	Language	Speaker
1	धोबिन जब सोकर उठती तब देखती कि चौंका साफ़ पढ़ा हैं और बर्तन मजे हुये हैं।	Hindi	Male
2	यहाँ से लगभग पाँच मील दक्षिण पश्चिम में कटघर गाँव हैं।	Hindi	Male
3	Welcome to the Internet my friend, How can I help you?	English	Male
4	I don't want to go out with mad people	English	Female
5	धोबिन जब सोकर उठती तब देखती कि चौंका साफ़ पढ़ा हैं और बर्तन मजे हुये हैं।	Hindi	Female
6	यहाँ से लगभग पाँच मील दक्षिण पश्चिम में कटघर गाँव हैं।	Hindi	Female

Fig. 4 Compression ratio comparison
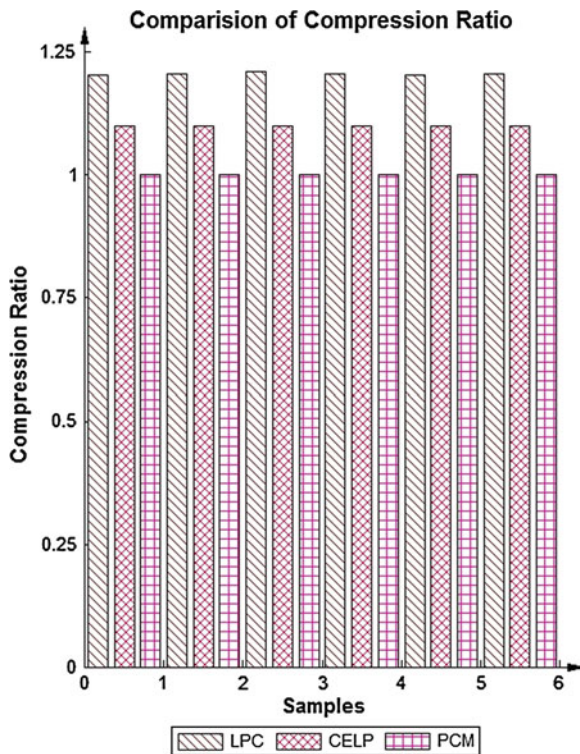
Fig. 5 SNR comparison
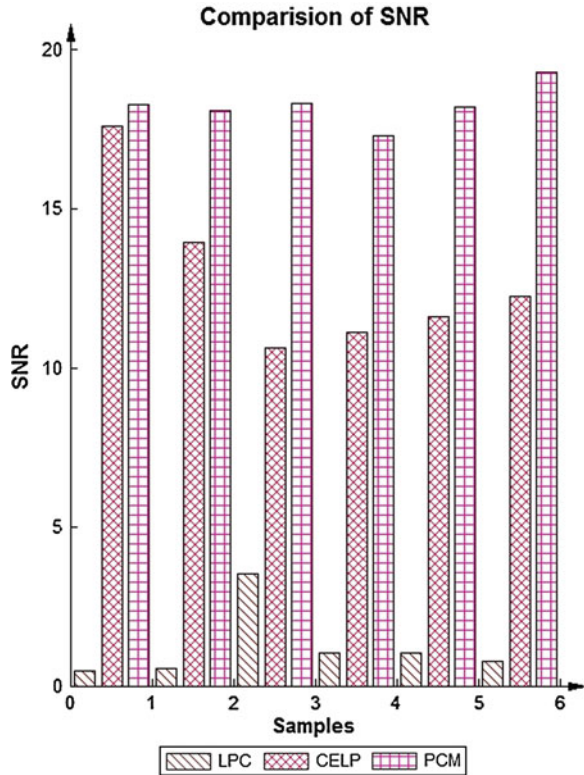
Comparative performance of LPC, CELP, and PCM is shown in Figs. 4, 5 and 6 in terms of compression ratio, SNR, and NRMSE. It is observed from results that LPC provides a greater degree of compression as compared to CELP and PCM for the entire sample sets. PCM provides excellent SNR, PSNR, and NRMSE measurements; however, it requires a huge bandwidth. CELP on the other hand provides an acceptable level of SNR, PSNR, and NRMSE with low bit rate requirements as compared to PCM. The quality of the reconstructed signal was tested and found to be in compliance with the MOS standard requirements [11, 12]. The MOS of PCM is best followed by CELP and LPC. Hence, it can be inferred from the above results that CELP provides a good alternative to PCM by conserving less bandwidth.

Fig. 6 NRMSE comparison
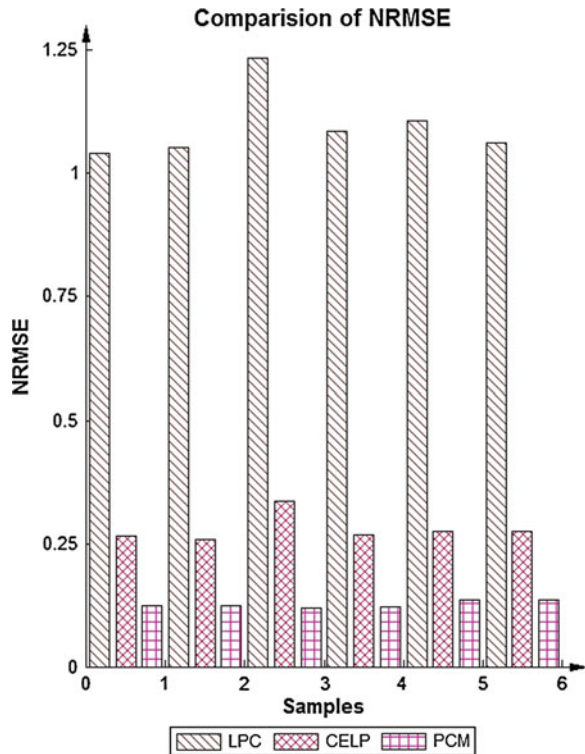
5 Conclusions

The performance of various standard codecs used for VoIP applications has been evaluated in this paper. It is observed that code excited linear prediction provides a comparable performance as compared to PCM with lower bandwidth requirements than that of PCM. LPC provides the good compression at lower bit rate; however, it lags in SNR as compared to the other two codecs. The results reveal that the performance of codecs under study remains unaffected by change in language or speakers.

References

1. Ogunfunmi, T., Narasimha, M.J.: Speech over VoIP networks: Advanced signal processing and system implementation. IEEE Circ. Syst. Mag. **2**(2), 35–55 (2012)
2. Ray, M., Chandra, M., Patil, B.P.: Evaluation of CDMA microwave links at different environments for VoIP applications. Int. J. Adv. Res. Comput. Commun. Eng. **1**(8), 508–512 (2012)

3. Ali, A.A., Vassilaras, S., Ntagkounakis, K.: A comparative study of bandwidth requirements of VoIP codecs over WiMax access networks. IEEE 3rd International Conference on Next Generation Mobile Applications, Services and Technologies, pp. 197–203 (2009)
4. Sunny, S., David, S.P., Jacob, K.P.: Recognition of speech signals: an experiment comparison of linear predictive coding and discrete wavelet transforms. Int. J. Eng. Sci. Technol. **4**, (2012)
5. Tandel, M., Shah, V., Patel, B.: Implementation of CELP CODER and to evaluate the performance in terms of bit rate, coding delay and quality of speech. In: 3rd IEEE International Conference on Electronics Computer Technology, pp. 86–89. Apr 2011
6. Ambika, D., Radha, V.: A comparative study between discrete wavelet transform and linear predictive coding. IEEE World Congress on Information and Communication Technologies, pp. 965–969 (2012)
7. Najih, A.M.M.A., Ramli, A.R., Ibrahim, A., Syed, A.R.: Comparing speech compression using wavelets with other speech compression schemes. In: IEEE Proceedings of Student Conference on Research and Development, pp. 55–58 (2003)
8. http://www.itu.int/ITU-T/recommendations/rec.aspx?rec=911
9. http://www.itu.int/ITU-T/recommendations/rec.aspx?rec=11674
10. Campbell, P.J., Tremain, T.E.: Voiced/unvoiced classification of speech with applications to the U.S. government LPC-10E algorithm. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 473–476 (1986)
11. http://www.cisco.com/c/en/us/support/docs/voice/h323/14069-codec-complexity.html#mos
12. Ray, A.K., Acharya, T.: Information Technology: Principles and Applications. Prentice-Hill of India Private Limited, New Delhi (2004)