Review Article | Impact Factor: 0.621 | ISSN: 2319-507X
Sarika Gorantiwar, IJPRET, 2014; Volume 2 (8): 324-330 | IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

SPEECH CODING TECHNIQUES: A REVIEW

SARIKA R. GORANTIWAR[1], NARESH P. JAWARKAR[2]

1. Department of Electronics & Telecommunication S.G.B.A.U. Amravati (India)
2. Babasaheb Naik College of Engg., Pusad.

Accepted Date: 15/02/2014 ; Published Date: 01/04/2014

Abstract: In human communication, voice is the preferred method. But in order to fit a band limited storage space or some transmission channel, speech signals have to be converted to formats using various techniques called as speech coding techniques. Today, speech coders have become an essential component in cellular communications, video conferencing, multimedia applications and computer based games. In the past decade, progress has been made towards the development of low-rate speech coders to be used in various applications which led to the development of new speech coders capable of producing high-quality speech reconstruction at low data rates. In order to optimize the performance, most of these coders employ the mechanisms to represent the spectral properties of speech, provided for speech waveform matching and hence achieve improved compression as compared to direct quantization. A number of these coders have already been adopted in national and international cellular telephony standards. This paper presents a review of various speech coders based on various parameters, their merits, demerits, types and applications.

Keywords: Vocoder, Excitation signal, Codebook, Encoder, coding technique

PAPER-QR CODE

Corresponding Author: Ms. SARIKA R. GORANTIWAR

Access Online On:

www.ijpret.com

How to Cite This Article:

Sarika Gorantiwar, IJPRET, 2014; Volume 2 (8): 324-330

INTRODUCTION

Speech coding or compression is the field concerned with obtaining compact digital representations of voice signals for the purpose of efficient transmission or storage. Narrow Band (NB) coding refers to coding of speech signals whose bandwidth is less than 4 kHz (8 kHz sampling rate), while wideband coding refers to coding of signals with bandwidth7 kHz (14–16 kHz sampling rate). A speech-coding algorithm performance [1] is evaluated based on number of factors. Coding rate reflects how much the signal can be compressed. It is usually expressed in terms of bits/second. Speech quality of reconstructed speech is usually given in terms of Mean Opinion Score (MOS). Algorithm complexity is usually given in terms of Million Instructions Per Second (MIPS). Delay refers to the coding delay, which affect the normal communication of a speech and degrade the speech quality.

Review of Various Coders

Broadly speech coders are classified into waveform coders, parametric coders (vocoders) and hybrid coders.

Waveform coders:

This section describes the coders that produce close reconstruction of the original speech signal. Waveform coders are generally more robust as they work well with a wider class of signals; however, they also generally operate at higher data rates relative to vocoders. Some of them operate in time domain and others operate in frequency domain.

Time domain coders are Pulse Code Modulation (PCM), Adaptive Differential PCM (ADPCM) and Adaptive Predictive Coder (APC). PCM is a memoryless, scalar quantization method [2]. Uniform PCM is a process that quantizes amplitudes by rounding off each sample to one of the set of discrete values. Although uniform PCM is the simplest method for digital encoding, it is also the most expensive in terms of data rate. Non-uniform PCM uses a non-uniform step size i.e. fine quantizing step size for frequently occurring amplitudes and a coarse step size for rarely occurring amplitudes. PCM is used today in both private and public telecommunications networks.

ADPCM [3] is the differential coder in which the difference between the input speech samples and the corresponding prediction estimates is calculated and the difference is then quantized and transmitted. APC is also a kind of differential coder that uses both short term and long term prediction in a differential coding structure.

Frequency domain coders are the one which decompose the input signal into sinusoidal components. The major types are Adaptive Transform Coding (ATC), Sub-Band Coding (SBC), Harmonic Coding (HC) and Sinusoidal Transform Coding (STC).

In sub-band coding, the signal band is divided into frequency sub-bands using a bank of analysis filters. The output of each filter is then sampled and encoded. At the receiver, the signals are multiplexed, decoded, demodulated, and then summed to reconstruct the signal. The design of the filter bank is a very important consideration in the design of an SBC.

In transform coding, the transform components of a unitary transform are quantized and transmitted. At the receiver end they are decoded and inverse-transformed. The bit-rate reduction in transform coding [4] is achieved due to the fact that unitary transforms generates near-uncorrelated transform components which can be coded independently. Also, the variances of these components [5] usually show consistent patterns which can be used for redundancy removal using some bit-allocation rules. ATC is a transform coder that employs the Discrete Cosine Transform (DCT) and encodes the transform components using adaptive quantization and bit assignment rules. STC is a technique, in which finite segments of speech signal are represented by linear combination of sinusoids with time-varying amplitudes, phases, and frequencies. The STC produces reconstructed speech of high quality at data rates below 10 kbps. Speech is analysed by first segmenting the signal using a finite duration analysis window. Then, for each segment, the excitation parameters are determined, which include voiced/unvoiced decision and a pitch frequency. In Multi Band Excitation (MBE) approach, voicing is estimated with mean squared error of original speech and synthetic speech. Improved MBE (IMBE) incorporates better quantizing methods compared to MBE and hence proved to be very good for background noise and channel errors.

Parametric coders (vocoders):

In order to reduce the number of bits required to represent the speech signal, vocoders extracts the feature parameters of the speech signals and then encodes them before transmitting. Coding of the source parameters needs significant computational complexity and memory to preserve the original speech. The system is usually coded using a codebook that needs only less number of bits. Unlike the waveform coders, the performance of vocoders [6] generally degrades for non-speech signals. They operate at very-low rates but produces speech of synthetic quality. Vocoders are classified into low rate (channel and Linear Predictive Coder (LPC) vocoder) and medium/high rate vocoders (formant, homomorphic vocoders). In channel vocoders, the speech is low pass filtered to create a baseband signal and then transmitted by some waveform coding technique. At the receiver, the baseband signal is decoded, spectrally

flattened, applied as excitation to vocoder for frequencies higher than the baseband, and also added to the vocoder speech to produce the output. Improvement in this technique is achieved by increasing the number of channels, introducing spectral flattening techniques on the excitation signal, and exploiting the correlation of the channel signals in the time and frequency domain.

LPC predictive coding has proved to be an efficient method as the coding efficiency is achieved by removing any redundant information of the speech signal. The predictor and the past values of the speech signals from the estimate of the current sample of the input signal. The difference between the current values of the input signal and its predicted value is quantized and stored. The properties of speech signals are different from one sound to another. Thus, it is necessary for efficient coding that both the predictor and the quantizer be adaptive. It has also been shown that the quantization noise appearing in the output speech signal is identical to the quantizer error. One of the most powerful predictive speech analysis techniques is the linear predictive method. The importance of this method lies in its ability to provide extremely accurate estimates of speech parameters, and its speed of computation. The major advantage is that they can be coded using perceptual quantization rules. The main drawback lies in the complexity associated with their computation.

In homomorphic vocoders, the vocal tract and the excitation log-magnitude spectra is combined to produce the speech log-magnitude spectrum. A good-quality speech can be produced by combining homomorphic deconvolution with analysis-by-synthesis excitation modelling [7]. The inverse fast Fourier transform (IFFT) of the log-magnitude spectrum of speech is used to produce the spectral sequence. The synthesizer takes the FFT of the cepstrum and the resulting frequency components are exponentiated. The IFFT of these components gives the impulse response of the vocal tract which is convolved with the excitation to produce synthetic speech.

Formant vocoders are implemented by using a cascade and parallel resonator configurations. The transfer function for voiced-speech synthesis consists of a cascade of three second-order all-pole resonators. For unvoiced speech, it consists of a cascade of a second-order all-zero function and a second-order all-pole function. The fixed spectral compensation function accommodates the effects of the glottal pulse and the lip radiation. The major difficulty in formant vocoders lies in the computation of the formants and their bandwidths.

Hybrid coders:

Hybrid coders combine the coding efficiency of vocoders and providing waveform matching similar to waveform coders at the same time. To get the accurate excitation model, hybrid

coders form their excitation signal directly from the input signal. The resulting signal is then coded, with the filter information parameters, and transmitted to the receiver where speech is synthesized. Modern hybrid coders [8] can achieve communication quality speech at 8 Kbits/s and below at the expense of increased complexity. These coders can be classified as Multi Pulse Excitation LPC (MPE-LPC), Code Excited Linear Predictive Coding (CELP), Mixed Excitation Linear Prediction (MELP) and Backward Excitation Recovery (BER-LPC).

CELP encodes the excitation using a codebook of Gaussian sequences. The excitation vector is scaled by a gain factor and excitation samples are filtered by long and short term synthesis filters. The synthesis filter process is carried out in three steps namely LPC analysis to compute the LPC parameters; pitch analysis to compute the long-term predictor parameters; and a codebook search to determine the shape and gain of the excitation vector. Many national and international communication standards use CELP [9] algorithms. The main drawback of CELP is computational complexity, so for wideband coding, Algebric CELP (ACELP) [10] can be used having much optimized codebook searches.

In MELP coder, the input frame is first split into sub- bands and long term correlation in every band is then used to classify it as voiced or unvoiced so as to improve the voicing errors.

In MPE-LPC, the excitation sequence is a sequence of irregularly spaced pulses and hence the name. The optimum pulse locations are found by computing the error for all possible pulse locations with their optimum amplitudes in a given analysis block and selecting the allowable number of locations and their amplitudes that result in the minimum error. MPE coders use a synthesis filter which consists of one or two autoregressive filters in series to have the smooth spectral envelope and to model the harmonic structure of the spectrum. MPE-LPC coders provide near network quality speech in the range 16 to 8 kbits/sec. Their performance deteriorates significantly below that. In Regular Pulse Excitation (RPE) coders, the pulses are spaced uniformly and hence their positions are determined by specifying the location of the first pulse.

BER-LPC defines the excitation signal from past information that is available at both the transmitter and the receiver. Table 1 summarizes the performance of various coders as given in [2].

Table 1: Comparison of some speech coding algorithms

[1] Encoder	[2] Bit Rate [3] (kbits/s)	[4] MOS	[5] Complexity [6] (MIPS)
[7] PCM	[8]64	[9]4.3	[10] 0.01
[11] SB-ADPCM	[12] 64,56,48	[13] 4.1	[14] 5
[15] ACELP	[16] 5.3	[17] 3.7	[18] 11
[19] ADPCM	[20] 40,32, [21] 24,16	[22] 4.1 [23] 4.0	[24] 5 [25]
[26] LD-CELP	[27] 16	[28] 4.0	[29] 30
[30] CS-ACELP	[31] 8	[32] 4.0	[33] 20
[34] LPC-10	[35] 2.4	[36] 2.3	[37] 7
[38] CELP	[39] 4.8	[40] 3.2	[41] 16
[42] MELP	[43] 2.4	[44] 3.2	[45] 40
[46] IMBE	[47] 4.15	[48] 3.4	[49] 40
[50] STC-1	[51] 4.8	[52] 3.52	[53] 13
[54] STC-2	[55] 2.4	[56] 2.9	[57] 13

CONCLUSION

This paper gives the review of various speech coding techniques. Various speech coding methods have been developed in the past decade to have the low bit rate coding which produce high quality speech. Many of them are incorporated in the national and international telephone standards. We note that waveform coders have relatively simpler algorithms, better adaptive capability and enhanced speech quality. However they operate at higher bit rates as compared to vocoders. Vocoders on the other hand provide improved quality at the cost of increased complexity. Hybrid coders are used according to specific application. The coder performance is studied in terms of bit rate, MOS and MIPS.

REFERENCES

1. Ming Yang, *"Low bit rate speech coding",* Potentials, IEEE, vol. 23,pp. 32- 36, Oct-Nov.2004.

2. A.S. Spanias, *"Speech coding: a tutorial review",* Proceedings of the IEEE vol. 82, pp. 1541 – 1582, Oct 1994.

3. J. D. Gibson, *"Speech coding methods, standards, and applications",* IEEE trans on Circuits and Systems Magazine, vol. 5, pp. 30-49, Dec 2005.

4. S. Ahmadi and A. S. Spanias, *"New techniques for sinusoidal coding of speech at 2400 bps",* vol.1, pp. 770 - 774, Nov 1996.

5. A. P. Berg and W. B. Michael, *"A survey of mixed transform techniques for speech and image coding",* Proceedings of the 1999 IEEE International Symposium on Circuits and Systems, vol. 4, pp. 106 – 109, Jul 1999.

6. A. Saranya, N. Sripriya and T. Nagarajan, *"Design of a VOCODER using instants of significant excitation",* International Conference on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), pp. 742 – 746, July 2011.

7. C. Xydeas, *"An overview of speech coding techniques",* IEEE trans. on Speech Coding - Techniques and Applications, pp. 111 – 125, Apr 1992.

8. A. M. Kondoz, *'Digital speech',* John Wiley & Sons, 2nd Edition, 2004.

9. E. Pryadi, K. Gandi and H. Y. Kanalebe, *"Speech compression using CELP speech coding technique in GSM AMR",* 5th IFIP International Conference on Wireless and Optical Communications Networks, pp. 1 - 4, May 2008.

10. C. Laflamme, J. P. Adoul, R. Salami, S. Morissette and P. Mabilleau, *"16 KBPS WIDEBAND SPEECH CODING TECHNIQUE BASED ON ALGEBRIC CELP",* International Conference on Acoustics, Speech, and Signal Processing, vol.1, pp. 13 - 16, Apr 1991.