Penerbit Akademia Baru

Journal of Advanced Research in Computing and Applications

Journal homepage: www.akademiabaru.com/arca.html
ISSN: 2462-1927

Chaos based secure image hiding in variable bit rate CELP speech coding systems

Open Access

Said E El-Khamy[1,*], Noha O Korany[1], Marwa H El-Sherif[1]

[1] Department of Electrical Engineering, Alexandria University, Alexandria 21544, Egypt

ARTICLE INFO

Article history:
Received 6 June 2017
Received in revised form 10 July 2017
Accepted 15 July 2017
Available online 4 December 2017

Keywords:
Audio Steganography, speech codec, chaotic maps, least significant bit, CELP, PESQ

ABSTRACT

Audio Steganography is a powerful tool for protecting important data during transmission. Inaudibility and robustness are two substantial requirements for an effective audio steganography algorithm. In this paper, a simple and effective hiding approach in the bit stream of a Code Excited Linear Prediction (CELP) speech codec using random sequences generated by chaotic maps, is proposed. Chaotic maps were used in our scheme for two important purposes: Encryption of the secret image using hopping between multiple random sequences generated by different logistic maps, and selecting of random bit locations in LSB of linear prediction coefficients of speech for hiding image bits. This approach boosts algorithm security. The technique has been evaluated extensively in terms of PESQ and PSNR, and the simulation results prove the effectiveness of information transmission with a fiddling effect on the quality of the coded speech.

1. Introduction

Low bit rate coders have been extensively used in nowadays speech communication systems due to its low communication costs [1]. Speech coding techniques have many applications in public switched telephone network (PSTN), videoconferencing systems, digital cellular communications, and emerging voice over Internet protocol (VoIP) applications. There is always a trade-off between voice quality and lower bitrates, however many other components affect speech coding systems like: complexity, delay and packet losses.Therefore information hiding in low bit speech streams had a great interest between researchers, as it gives large scale transmission of secret data, and also most speech streams in communication systems are transmitted in real time, which gives short time to eavesdroppers to perform steganalysis.

Although embedding data into low bit-rate speech streams is a challenging work because of little redundancy in low bit-rate codecs and the real-time requirement of communication systems, there have been some efforts by researchers for data hiding without affecting speech quality [2-9].

* *Corresponding author.*
E-mail address: elkhamy@ieee.org (Said E El-Khamy)

Peng Liu [2] proposed an information hiding method for low bitrate VoIP speech codecs based on matrix embedding during the VQ process of linear predictive coefficients by code word replacements. Shufan Yan [4] proposed a method where modification directions (adding or subtracting one) of the pitch parameter are selected adaptively in order to achieve a high embedding efficiency, and matrix encoding method was used twice to increase the hiding capacity. In The technique used by Tamer Rabie [5], low-pass spectral properties of the speech magnitude spectrum were exploited to embed a secret speech signal in the low-amplitude-high-frequency regions of the host speech signal's spectral magnitude. Xiao Bo [6] introduced a novel codebook partition algorithm for Quantization Index Modulation (QIM), which is applied to information hiding in instant low bit-rate speech stream. Hui Tian, [7], presented a novel model of real-time steganography based on m-sequence in VoIP to eliminate the correlation among secret messages to resist the statistical detection and to provide a short-term secure protection of secret messages.

In a paper by Rekik *et al*. [8], a new method to secure speech communication using the discrete wavelet transforms (DWT) and the fast Fourier transform was presented. In the first phase of the hiding technique, the speech high-frequency components were separated from the low-frequency components using the DWT. In a second phase, the low-pass spectral proprieties of the speech spectrum were exploited to hide another secret speech signal in the low-amplitude high-frequency regions of the cover speech signal. A paper by Ahlam [9], presents a new lossless speech steganography approach based on Integer-to-Integer Lifting Wavelet Transform (Int2Int LWT) and Least Significant Bits (LSBs) substitution. Chaotic key has been proposed to increase security.

The proposed algorithm performs a simple and robust data embedding technique by using adapted LSB technique to hide data in LPC parameters of coded speech with a specified chaotic maps sequence. Instead of its relatively low resistibility to interference, LSB method can hide comparatively large amount of information with little modification of original data, which makes it a practical steganography technique. Besides, random insertion can scatter information in the carrier in a relatively balanced manner, thus greatly improve the security of steganography. The remainder of the paper is arranged as follows: in section 2 the basic concept of (CELP) speech coder is discussed, section 3 presents proposed model for embedding and extracting data in details. Section 4 discusses the objective tests results and related analysis. Section 5 concludes the paper and discusses the future work.

2. Code Excited Linear Prediction (CELP)

The basic concept of all speech coders depends on the reality that speech signals are highly correlated waveforms [10-11]. Speech can be represented using a linear combination of the previous p samples plus a white noise, as in equation (1).

$$X(m) = \sum_{k=1}^{p} a_k x(m-k) + e(m) \quad (1)$$

The weighting coefficients a_1, a_2, …., a_p are called linear prediction coefficients (LPC'S) .A block diagram of CELP analysis-by-synthesis coder is shown in the Fig.1. The samples of the input speech are divided into frames of N samples each. Each frame is typically 10-20 ms long. Each frame is divided into smaller sub-frames, of L samples each. It is called analysis by synthesis because speech is encoded and decoded in the encoder and then the parameters that minimize the energy of the error signal are found. First, LP analysis is used to estimate the vocal system impulse response in each frame. A codebook of size M and dimension L, is available to both the encoder and the

decoder. The code vectors have components that are all independently chosen from N (0, 1) distribution so that each code vector has an approximately 'white' spectrum. The code vector whose output best matches the input speech (least MSE) is selected to represent the sub frame. In our implementation frame duration of 20ms is used for the vocal-tract analysis (160 samples of an 8 kHz sampling rate) and 5ms sub-frame duration (40 samples) for determining the excitation.

3. Proposed Algorithm PCA

In embedding phase, the secret image is divided into 16 blocks each of size 16×16, and then each block samples are encrypted by sample XORing with different logistic maps [13-14]. The equation of logistic map is given in equation (2), where Xn is a number between zero and one, n is the iteration number, X0 is the initial value and r is a positive number between 3.57 and 4 where the chaotic behaviour occurs [15-16].

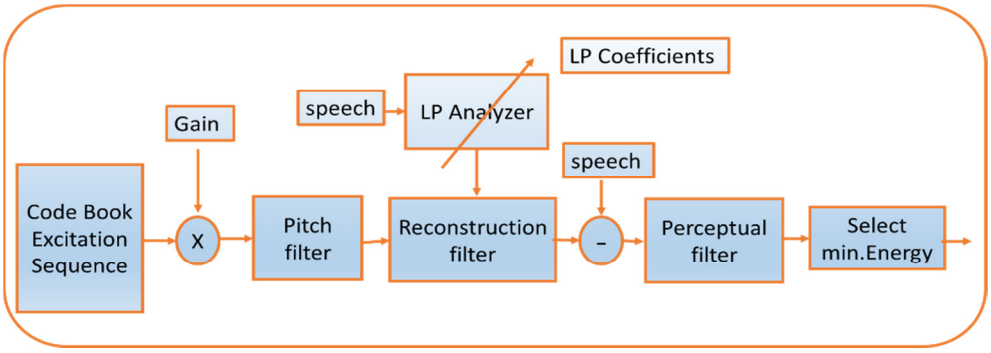

Fig. 1. Block Diagram of CELP

$$X_{n+1} = rX_n(1 - X_n) \tag{2}$$

The hopping between logistic maps is done by generating other map responsible for selecting the bifurcation parameter (r), after adjusting its values to be in the chaos region. Then final encrypted image of size 64×64 is obtained and then image samples are converted into binary form. Speech stream is divided into frames of 160 sample and encoded parameters are computed for each frame. Then image bits are embedded in the LSB'S of LPC's by a predetermined sequence using logistic map. Subsequently, all computed parameters for each frame are encoded and transmitted.

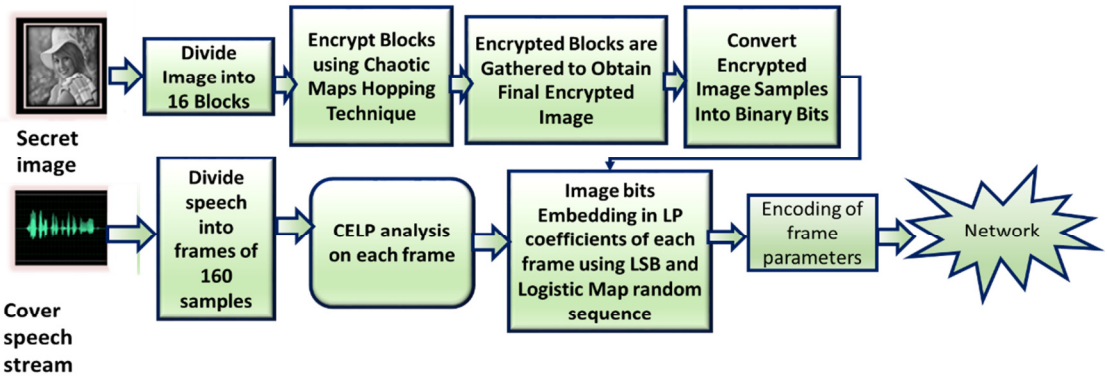

Fig. 2. The proposed steps for embedding hidden data

In extraction phase, the received LPC parameters are decoded and image bits are extracted from each frame, then the resulted bits are decrypted to recover the extracted image. The proposed method for embedding and extraction is shown in Figs. 2 and 3.

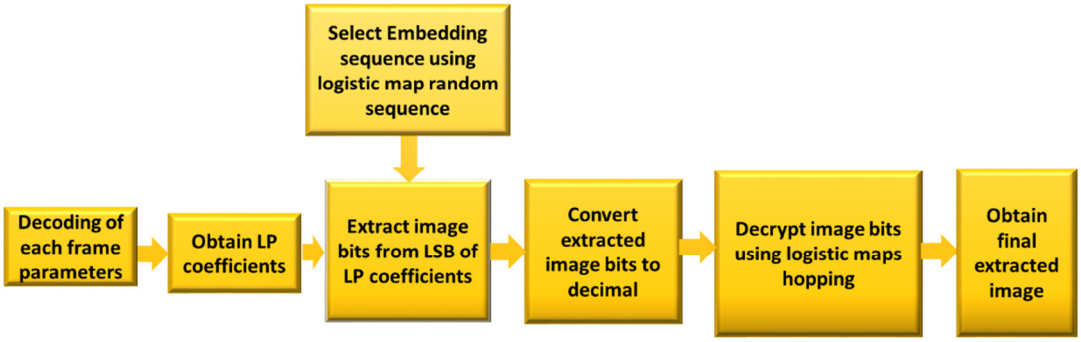

Fig. 3. The proposed steps for extracting hidden data

5. Results and Discussion

The speech quality is measured in terms of Peak Signal to Noise Ratio (PSNR) and Perceptual Evaluation of Speech Quality (PESQ) using the test speech data in Table. 1 and the degradation of extracted image is measured using Mean–structure similarity-index-measure (MSSIM). The high MSSIM values of the extracted images denote better imperceptibility and secrecy of the hidden data. MATLAB R2014a program has been used to generate the code, and four speech samples of lengths between (1-3) mins' were considered with a sampling frequency of 8000 Hz. An image of size 64×64 is used as a hidden image.

PSNR can intuitively compare the differences between the original signal x and the synthetic coded signal y:

$$PSNR(dB) = 10\log_{10}\left(\frac{R^2}{MSE(x,y)}\right), \tag{3}$$

where R is the maximum signal value in the original audio signal and Mean Squared Error (MSE) gives a measure of the average of squared errors between the two signals [16].

Mean–structure similarity-index-measure (MSSIM) is an estimate for the deformity of extracted image. SSIM depends on the human eye perception and so it's preferable than other methods like PSNR or MSE in expressing the amount of distortion. MSSIM is expressed as

$$MSSIM(P,\overline{P}) = \frac{1}{M'}\sum_{J=1}^{M'} SSIM(P_J,\overline{P}_J), \tag{4}$$

where $SSIM(P,\overline{P}) = [l(P,\overline{P})]^{\delta}.[c(P,\overline{P})]^{\beta}.[s(P,\overline{P})]^{\gamma}$.The functions $c(P,\overline{P})$, $l(P,\overline{P})$ and $s(P,\overline{P})$ are the contrast comparison, luminance comparison and the structure comparison functions, respectively [16]. The symbols δ, β and γ {δ, β, γ > 0} regulate the relative prominence of the components.

Perceptual evaluation of speech quality (PESQ) is an objective method, used in speech quality assessment of speech codecs. A computer model of the subject, consisting of a perceptual and a cognitive modeling process, is used to compare the output of the device under test (e.g. a speech or a music codec) with the input (= reference), using any audio signal (speech, music or artificial test signal) [17]. A perceptual model, applying a time alignment for delay estimates, models original and degraded speech. The output is a single value in -1 to 4.5 range. A simplest variable bit rate Vocoder having a codebook containing 1024 sequences of length 40 was used and operated in two modes:

- Medium bit rate (16 Kbps) CELP.
- Low bit rate (9.6 Kbps) CELP.

Table1 shows bit allocation for the specific bit rate and table2 presents the results of objective tests.

Table 1
Bit allocation for the specific bit rate

parameter	Bits/parameter		Bits/frame	
	9.6 kbps	16kbps	9.6 kbps	16kbps
Code book index k	10	10	40	40
12 LPC coeff	6	12	60	144
Gain	7	13	28	52
Pitch filter coeff,b	8	13	32	52
Lag of pitch filter	8	8	32	32
			∑192	∑320

Table 2
Tests of quality of coded speech and extracted image

Speech samples	Bit rate	PSNR(dB)	PESQ	MSSIM
Male 1	16 kbps	21.85	1.54	1
	9.6 kbps	21.76	1.51	0.88
Male 2	16 kbps	17.4	1.82	1
	9.6 kbps	17.2	1.81	1
Female 1	16 kbps	16.3	1.56	1
	9.6 kbps	16.2	1.57	0.6
Female 2	16 kbps	14.6	1.42	1
	9.6 kbps	12.5	1.43	0.7

5. Conclusion

This paper gave a simple and effective hiding technique for low and medium bit rate speech stream depending on analyzed parameters from CELP coding technique. CELP coders can maintain some phase information from original signal, so it's able to simulate the original envelope of speech more strictly. Hence, CELP is preferable to Linear Predictive Coding (LPC) in speech synthesis purposes. The results showed that changes in LSB of LP coefficients have a little influence on speech quality, which can be used effectively for embedding. The security of proposed method has been improved by using a dynamic encryption algorithm with chaotic maps. Moreover, quality tests proved that synthesized speech with embedded secret data still has acceptable levels of imperceptibility, and with high embedding efficiency of 12 bits per frame. Future work is to try to improve hiding capacity and speech audibility of the proposed algorithm. Also, future work may include design and implementation of a real time speech secure communication system with different modulation techniques and test the behavior of the proposed models and algorithms in information hiding.

References

[1] R. Ram, H.K. Palo, M. N. Mohanty and B.N. Sahu," Speech Coding Techniques: A Comparative Study", International Journal of Electronics & Communication Technology, IJECT Vol. 6, Issue 3, July - Sept 2015 ISSN : 2230-7109 (Online) | ISSN : 2230-9543 (Print).

[2] Liu, Peng, Songbin Li, and Haiqiang Wang. "Steganography integrated into linear predictive coding for low bit-rate speech codec." *Multimedia Tools and Applications* 76, no. 2 (2017): 2837-2859.

[3] Huang, Yongfeng, Chenghao Liu, Shanyu Tang, and Sen Bai. "Steganography integration into a low-bit rate speech codec." *IEEE transactions on information forensics and security* 7, no. 6 (2012): 1865-1875.

[4] Yan, Shufan, Guangming Tang, Yifeng Sun, Zhanzhan Gao, and Liuqing Shen. "A triple-layer steganography scheme for low bit-rate speech streams." *Multimedia Tools and Applications* 74, no. 24 (2015): 11763-11782.

[5] Rabie, Tamer, and Driss Guerchi. "Spectral Magnitude Speech Steganography." *International Journal of Computer Applications* 116, no. 5 (2015).

[6] Xiao, Bo, Yongfeng Huang, and Shanyu Tang. "An approach to information hiding in low bit-rate speech stream." In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pp. 1-5. IEEE, 2008.

[7] Tian, Hui, Ke Zhou, Hong Jiang, Jin Liu, Yongfeng Huang, and Dan Feng. "An M-sequence based steganography model for voice over IP." In *Communications, 2009. ICC'09. IEEE International Conference on*, pp. 1-5. IEEE, 2009.

[8] Rekik, Siwar, Driss Guerchi, Sid-Ahmed Selouani, and Habib Hamam. "Speech steganography using wavelet and Fourier transforms." *EURASIP Journal on Audio, Speech, and Music Processing* 2012, no. 1 (2012): 20.

[9] Kathum, Ahlam Majead, and Saad Najim Al-Saad. "Speech Steganography System Using Lifting Wavelet Transform." *International Information Institute (Tokyo). Information* 19, no. 10B (2016): 4633.

[10] Huang, Yongfeng, Bo Xiao, and Honghua Xiao. "Implementation of covert communication based on steganography." In *Intelligent Information Hiding and Multimedia Signal Processing, 2008. IIHMSP'08 International Conference on*, pp. 1512-1515. IEEE, 2008.

[11] Mulye, Minal, and Sonal Jagtap. "Overview of Code Excited Linear Predictive Coder."

[12] S. K.Jagtap, M.S.Mulye and M.D. Uplane, "Speech coding techniques", 4th International Conference on Advances in Computing, Communication and Control (ICAC3'15), DOI: 10.1016/j.procs.2015.04.251.

[13] S. E. El-Khamy, M. R. M. Rizk and A. E. Youssef, "Image Authentication and Watermarking using Histogram Matched Chaotic Maps." Computational Modeling of Objects Presented in Images: Fundamentals, Methods, and Modeling (CompIMAGE'10), Buffalo-Niagara, USA, May 5-7, 2010.

[14] El-Khamy, Said E., NohaO Korany, and Marwa H. El-Sherif. "Robust image hiding in audio based on integer wavelet transform and Chaotic maps hopping." In *Radio Science Conference (NRSC), 2017 34th National*, pp. 205-212. IEEE, 2017.

[15] Avval, Amir Houshang Arab, and Shahram Mohanna. "A New Robust Audio Signal Steganography Based On Multiple Chaotic Algorithms With Edge Detection Technique." *Advances in Computer Science: an International Journal* 3, no. 2 (2014): 97-106.

[16] Mondal, Bhaskar, and Tarni Mandal. "A multilevel security scheme using chaos based encryption and steganography for secure audio communication." *International Journal of Research in Engineering and technology* 2, no. 10 (2013): 399-403.
[17] Cernak, Milos, and Milan Rusko. "An evaluation of synthetic speech using the PESQ measure." In *Proc. European Congress on Acoustics*, pp. 2725-2728. 2005.