Speech Watermarking using a Hybrid Strategy of both Empirical Mode Decomposition and Singular Value Decomposition

Safaa el-Gazar
Dept. of Electrical and Communication Eng., Faculty of Electronic Engineering, Menoufia University, Egypt.

Sami El-Dolil
Dept. of Electrical and Communication Eng., Faculty of Electronic Engineering, Menoufia University, Egypt.

Alaa M. Abbas
Dept. of Electrical and Communication Eng., Faculty of Electronic Engineering, Menoufia University, Egypt.

Moawad I. Dessouky
Dept. of Electrical and Communication Eng., Faculty of Electronic Engineering, Menoufia University, Egypt.

El-Sayed M. El-Rabaie
Dept. of Electrical and Communication Eng., Faculty of Electronic Engineering, Menoufia University, Egypt.

Ibrahim M. El-Dokany
Dept. of Electrical and Communication Eng., Faculty of Electronic Engineering, Menoufia University, Egypt.

Fathi E. Abd El-Samie
Dept. of Electrical and Communication Eng., Faculty of Electronic Engineering, Menoufia University, Egypt.

Abstract

This paper presents a proposed robust speech watermarking approach. This approach aims to increase the speech watermarking robustness against different attacks. The method is based on Empirical Mode Decomposition (EMD) and Singular Value Decomposition (SVD). The speech signal is decomposed by EMD into its Intrinsic Mode Functions (IMFs), the first IMF transform to a 2-D format. The watermark image embedded into the singular values (SVs) of the first IMF. After watermark embedding, the speech signal transformed back into a1-D format. The first IMF preserves the speech signal perceptual quality, which leads to preserving the watermarked signal imperceptibility. The singular values matrix is stable against any small perturbation happens to the original signal which provide more secure and robustness against attacks. The proposed approach can be implemented on the speech signal as a whole or as a blocks. Block-based SVD implementation allows embedding more than one watermark in the speech signal which increase the opportunities and efficiency of watermark extraction in the presence of attacks. Simulation results show that using of EMD with SVD enhance the watermark extraction especially in the presence of attacks. A block-based implementation of the proposed speech watermarking also realize a higher correlation coefficient in the presence of attacks.

Keywords: Speech watermarking, EMD, SVD and IMF.

1. Introduction

One of the most important means of communication between humans is the speech, which carries the information so that we can distinguish between speakers [1, 2].Speech signals is important in many applications such as VoIP [3, 4], military communications, telephonic recording and identifying the airplanes through VHF radio channels, online speaker recognition systems and many other computer science applications [5, 6 and 7]. In recent years, there is no limits to transfer or share digital speech signals through communication and computer technologies. And due to the ease of producing and transmitting speech signal, it can be easily edit or remove small parts of the speech signal which leads to change the meaning of the speech signal. Therefore, it is important to apply speech watermarking technology to solve security, privacy and production problems. Speech watermarking is one of the branches of the speech processing that specializes in speech protection by embedding a piece of information in the speech signal [1, 2].

There are many watermarking techniques applied to the speech signal, one of these techniques is phase modulation which based on modifying the speech signal phase according to the watermark bits, there is no difference between the watermarked and the original speech signals power spectrum. This technique is not robust enough against signal processing attacks [8, 9].

Other technique is speech watermarking based on quantization where, watermark embedded in the perceptually irrelevant and relevant components of the speech signal. Also it is easy to implement and low complexity but quantization is sensitive to amplitude scaling attack and the imperceptibility of the speech signal is degraded [10].

Other techniques based on transformation such as spread spectrum and bit stream domain. In spread spectrum, a pseudorandom noise (PN) sequence is embedded to the speech signal in frequency domain spreading its spectrum. This technique allows the spread of the watermarked signal on a wide range of spectrum and thus, removal attacks cannot destroy the watermarked signal and thus realized a robust watermark. But this requires one of the decomposition functions such as DFT, which needs more time and therefore not suitable for online watermarking system [11, 12].

The bit stream domain depends on watermarked data embedding into speech signal relevant parts during or after compression thus, compression attacks cannot destroy the watermarked signal. It is not complicated and effective for online watermarking system but it very low payload and fragile to other attacks such as digital to analog conversion [13].

All previous techniques used a random sequence of bits to be embedded into the speech signal as a watermark. Which gives low data payload and so low capacity. Another trend to embed a meaningful information into the speech signal such as image. This process implemented by a robust algorithm such as SVD. The SVD is very robust to any to small perturbations in the signal. In this approach, the SVD will be utilized to embed the watermark image into the speech signal [14-20].

For more secure system information, watermark is embedded in EMD domain. The speech signal decomposes into components called IMFs using EMD. Then embedding the watermark image in selected IMF using SVD or block-based SVD. This process execute as follow, the speech signal decomposed using EMD to extract the first IMF which decomposed using SVD to extract the singular values (SVs)then, embedding the watermark into the SVs matrix. The choice to insert the watermark in the first IMF is because it is the closest mode of the original speech signal and thus embedding the watermark does not change or degraded the speech signal and make the watermark imperceptible to the listener. Also the choice of the SVs due to its stability against any variations in its values which make it more robust against different attacks.

This technique can be implemented in the speech signal as blocks by block-based SVD by first dividing the speech signal into equaled blocks and applying the method on each block individually then reconstruct the watermarked signal. This gives the chance to more than one watermark to be embedded besides realizing more chance to have more than one watermark to escape from attacks. A comparison study between SVD and block-based SVD after using EMD introduced later in the paper simulation results.

The rest of the paper is organized as follows: Section 2, briefly explains the EMD method. Section 3 presents the proposed watermarking algorithm using both EMD and SVD. Section 4, presents the simulation results after using EMD with SVD compared to using EMD with block-based SVD in the presence of different attacks. And section 5, gives the concluding remarks followed by the more relevant references.

2. The Empirical Mode Decomposition (EMD)

EMD method was proposed in the ocean waves study and found immediate applications in the signal processing. EMD can be applied in time domain on nonlinear and non stationary signals and so it can be applied on speech signal processing. The basic idea of EMD is that it decomposes the signal itself directly into a sum of functions called intrinsic mode functions (IMFs) and a residue by iterative sifting process. The IMFs are signal dependent. IMF must satisfy two conditions: (1) the number of extremea and the number of zero crossing are either equal or differ at most by one; and (2) the mean of its upper and lower envelopes equal zeros. For a speech signal S(t), n IMFs and a residual are obtained at the end of the decomposition. The signal could be then decomposed as follows [21-24]:

$$S(t) = \sum_{i=1}^{n} IMF_i + r_n \tag{1}$$

3. The Proposed Speech Signal Watermarking

The main idea of the proposed method depends on hybrid algorithm using SVD and EMD for speech watermarking. This is done by embedding the watermark into the first IMF of the speech signal using SVD technique. The first IMF contains the highest energy than other IMFs and this increase the robustness against different signal processing attacks. In the proposed method, at first the speech signal decomposed using EMD into its IMFs. Then, the first IMF decomposed again using SVD to extract SVs matrix. It characterized by its stability against any small variation caused by different attacks, making it more suitable for embedding the watermark in it. This can be considered the second level of robustness against different attacks. The watermark embedded by a certain weight which does not allow the speech signal to be destroyed and maintains an acceptable level of inaudibility. The proposed algorithm can be summarized as follows:

1. First decompose the speech signal by EMD into its IMFs, the first IMF converted to its 2-Dmatrix (***B***).
2. Perform SVD on the ***B*** matrix.

$$\boldsymbol{B = USV^T} \tag{2}$$

3. dd the watermark (***W***) to the singular values (SVs) matrix (***S***).

$$\boldsymbol{E= S+}k\boldsymbol{W} \tag{3}$$

K is a small value of about 0.01 required to preserve the speech signal quality.

4. Perform SVD on the new modified matrix (***E***).

$$\boldsymbol{E = U_w S_w V_w^T} \tag{4}$$

5. Obtain the watermarked speech signal in 2-D format ($\boldsymbol{B}_w$) using the modified matrix of SVs ($\boldsymbol{S}_w$).

$$B_w = US_wV^T \quad (5)$$

6. Transform the 2-D (B_w) matrix again into a 1-D signal.
7. Reconstruct the 1-D speech signal using EMD^{-1}and obtain the watermarked signal. By giving U_w, S, V_w matrices and the possibly distorted watermarked signal ($\mathbf{B}_w^*$),the possibly corrupted watermark can be extracted as follows:

- First decompose the possibly distorted watermarked signal by EMD into its IMFs, the first IMF converted to its 2-Dmatrix ($\mathbf{B}_w^*$).
- Perform the SVD on the distorted watermarked signal.

$$B_w^* = U^* S_w^* V_w^* \quad (6)$$

- Computed the watermark matrix.

$$E^* = U_w S_w^* V_w^* \quad (7)$$

- Obtain the extracted corrupted watermark.

$$W^* = (E^* - S)/k \quad (8)$$

12. Estimate the correlation coefficient between the extracted and the original watermarks. The watermark is present if this coefficient is higher than a certain threshold.

Figures (4) and (5) illustrate steps of applying the EMD with the block-based SVD where, more than one watermark can be embedded into the speech signal. This increase robustness against the different attacks especially for cropping attack because it gives a chance that at least one of the watermarks escape from the attacks.

4. Simulation Results

This section introduces the simulations based on the proposed watermarking method to test its performance. Simulations carried out using a computer with Intel 2.5GHz processor, 6.00 GB RAM and MATLAB R2008a. Simulations carried out by EMD with SVD compared with EMD with block-based SVD. Simulation Data used, Performance measurement parameters and assumed system attacks will be illustrated in this section.

4.1 Dataset Description

The used watermark image is the **CS** image and the speech signal for the sentences "Rice is often served in round balls. A large size in stockings is hard to sell" for a female. Table (1) illustrates the size and number of samples of the watermark image and the speech signal used for SVD and block-based SVD.

4.2 Performance Measurement Parameters of the Proposed Method

4.2.1 Correlation Coefficient

The quality of watermarking evaluated by estimated the correlation coefficient C_r between the extracted and the original watermark images. The high value of the C_r indicates the high quality of the watermarking. It can be defined using the following equation [25]:

$$c_r\left(W, \hat{W}\right) = \frac{\sum_i W(i)\hat{W}(i)}{\sqrt{\sum_i W^2(i) \sum_i \hat{W}^2(i)}} \quad (9)$$

$$W(i) = W_i - W_m \qquad W(i) = \hat{W}_i - W_m$$

where W_i and W_i are are intensity values of ith pixel in original and extracted watermark images respectively. W_m and $W_m$are mean intensity values oforiginal and extracted watermark images respectively.

4.2.2 Signal to Noise Ratio:

The effect of the watermarking on the original speech signal is evaluated by estimating the signal to noise ratio (SNR). The high value of SNR indicates the more improved the perceptual quality of the watermarked speech signal and its inaudibility [26].

$$SNR = 10\log_{10} \frac{\sum_{i=1}^{N} S_o^2(i)}{\sum_{i=1}^{N} (S_o(i) - S_w(i))^2} \quad (10)$$

where, where $S_o(i)$ is the original speech signal, $S_w(i)$ is the distorted speech signal, N is the samples total number in both speech signal, and i is the sample index.

4.3 The Assumed System Attacks

Simulations assumed presence of different types of attacks. The types of attacks applied on the watermarked speech signal were, an additive white Gaussian noise (AWGN) attack with SNR=10 dB, a low pass filtering attack with BW =8kHz, a cropping attack and a wavelet compression attack with threshold level T = 0.8, where the detailed coefficient below the threshold level will be removed.

4.4 Simulation Results and Discussion

Figure (6) shows the original speech signal and its spectrogram, first IMF and its spectrogram, SVD watermark image and block-based SVD watermark image. Figures (7) and (8) show the watermarked speech signal in the absence and presence of different attacks and its spectrogram respectively. Figure (9) shows the extracted watermarks, using EMD and SVD in the absence and presence of different attacks. Figures (10) and (11)show the extracted watermarks and magnification of watermark which achieves maximum correlation coefficient with the original watermark, respectively, using EMD and block-based SVD in the absence and presence of different attacks.

Table (2) compare the proposed approach with the previous methods in terms of SNR between the original and watermarked speech signals, computer execution time and the correlation coefficient of the extraction watermarks in the absence of any attacks.

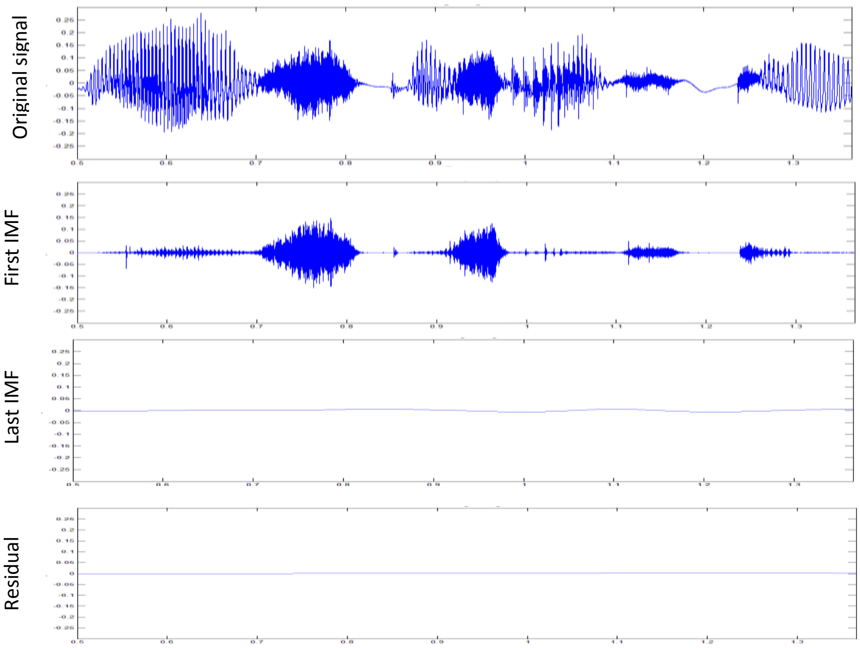

Figure 1 The original speech signal, the first IMF, the last IMF and the residual.

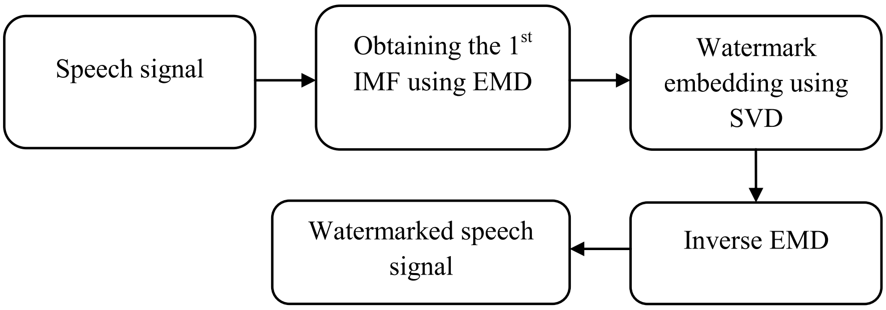

Figure 2 Proposed watermark embedding using SVD block diagram.

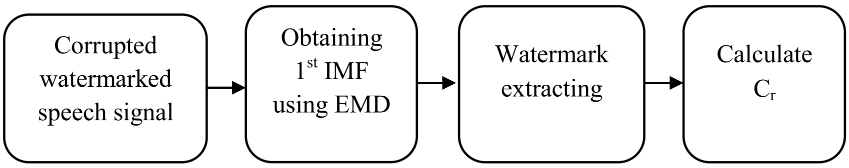

Figure 3 Proposed watermark extraction block diagram.

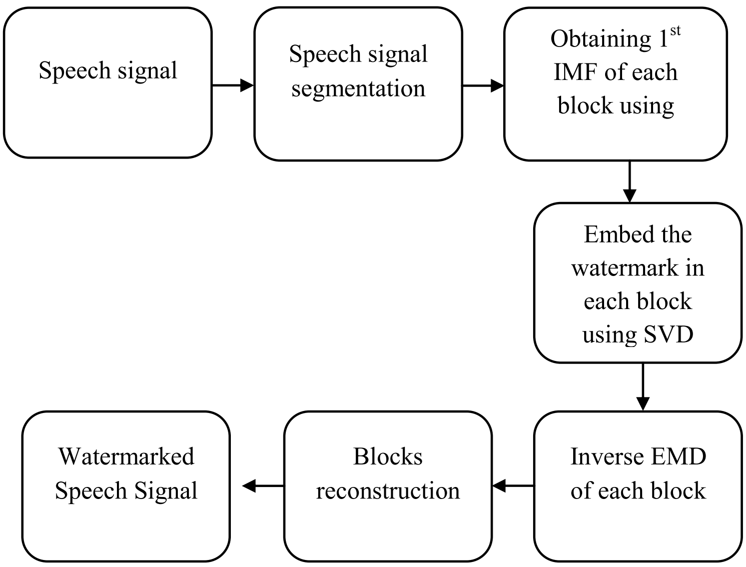

Figure 4 Proposed watermark embedding using block-based SVD block diagram.

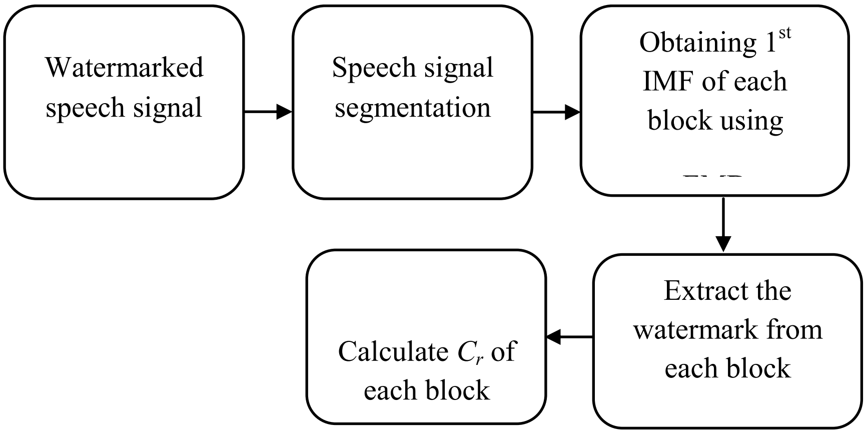

Figure 5 Proposed watermark extraction block diagram.

Table 1 dataset description

Embedding Method	Number of Speech Signal Samples	Watermark Image Dimension	Watermark Image Size
SVD	65536	256 × 256	63.5 KB
Block-based SVD	256 per block	16 ×16	1.15 KB

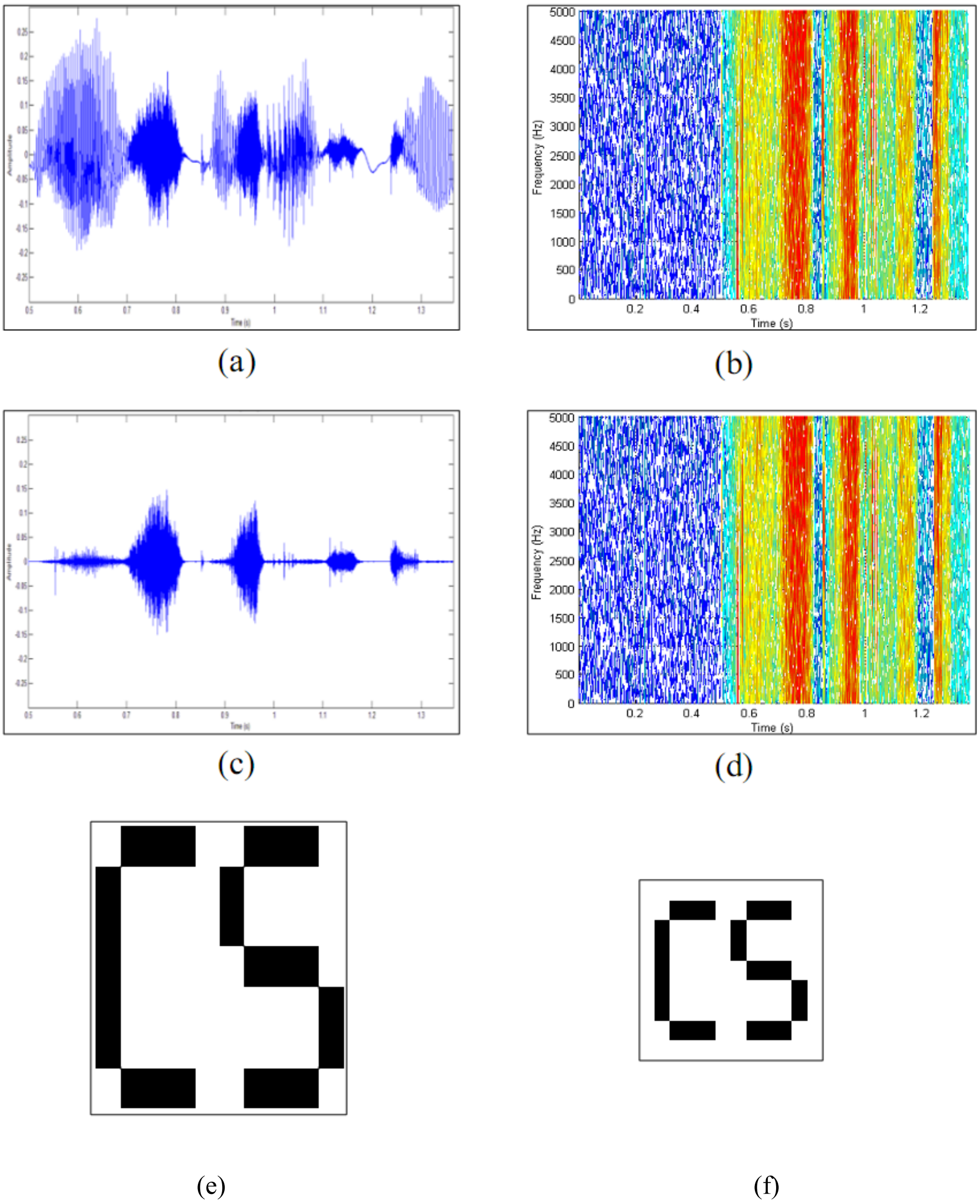

Figure 6 (a) Original speech signal. (b) Spectrogram of the original speech signal. (c) The first IMF. (d) Spectrogram of the first IMF. (e) The SVD method watermark image. (f) The block-based SVD method watermark image.

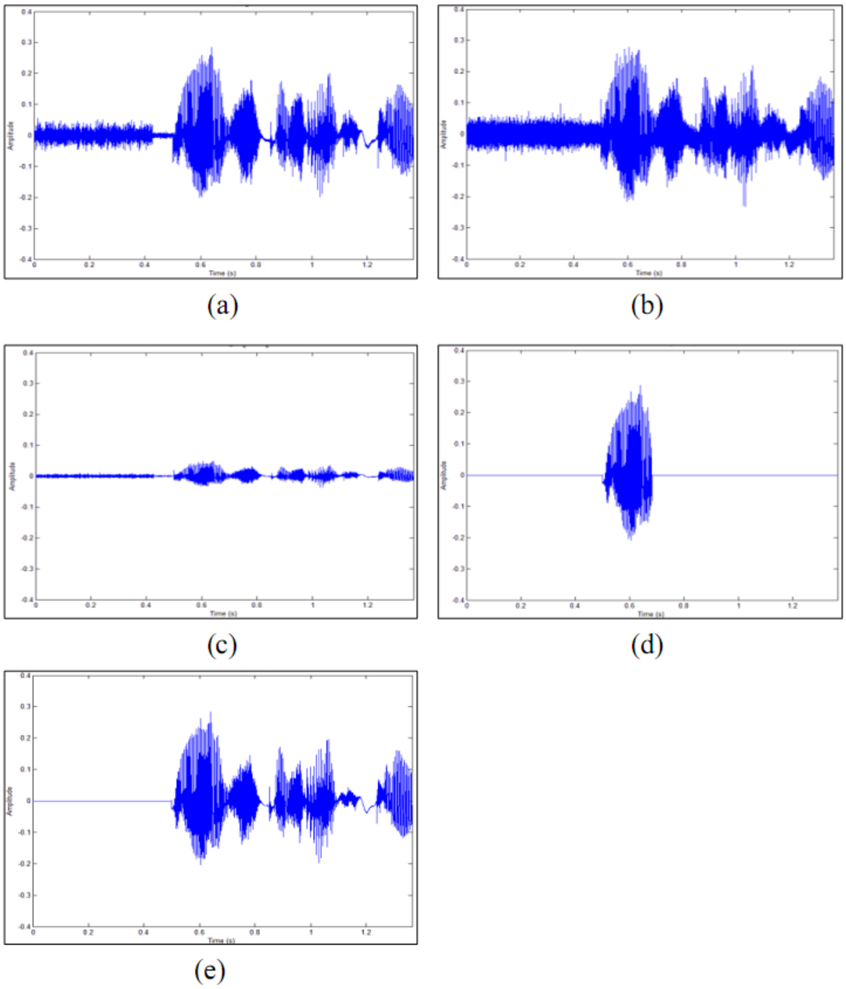

Figure 7 The watermarked signal in the (a) absence of attacks, (b) presence of WGN with SNR = 10dB, (c) presence of filtering attack with BW = 8kHz, (d) presence of cropping attack and (e) presence of wavelet compression attack with threshold level T = 0.8.

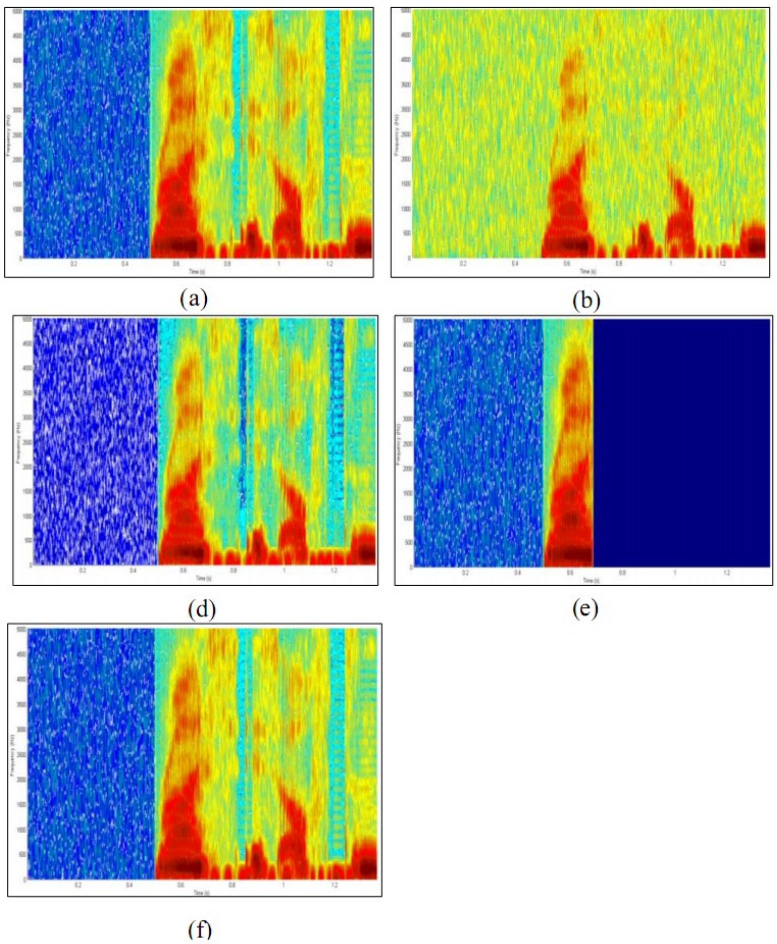

Figure 8 The spectrogram of the watermarked signal in the (a) absence of attacks, (b) presence of WGN with SNR = 10dB, (c) presence of filtering attack with BW = 8kHz, (d) presence of cropping attack and (e) presence of wavelet compression attack with threshold level T = 0.8.

Values in table indicate that, the proposed method takes longer time to execute especially for the whole signal. It is clear that using the proposed approach decrease the SNR than the previous method. Table (3) gives the correlation coefficient values using EMD and SVD compared with using EMD and block-based SVD in the absence and presence of attacks.

Figures 9, 10 and 11 and Table 3 indicate the perfect extraction of the watermark in case of using EMD and block-based SVD where C_r are close to 1 even in the presence of different attacks especially for filtering attack reverse what is in the case of EMD and SVD where the lowest value of C_r is achieved in the presence of filtering attack.

From the obtained results in figure (12), it is clear that the trend of the correlation coefficient between the original and the extracted watermarks is increasing with the SNR of the hybrid EMD and SVD method. On the other hand, this is trend is approximately constant in hybrid EMD and the block-based SVD watermarking at the level of approximately 0.9.

The reason for these results is that we work on the maximum value of correlation coefficients in the presence of different attacks. This means that some blocks in all cases can survive the attacks with large correlation coefficient.

Figure (13) reveals that low-bandwidth filtering destroys the correlation levels in the case of hybrid EMD and SVD watermarking, while the situation is better with hybrid EMD and block-based SVD. This is attributed to the survival of some of the blocks to attacks.

In compression attacks in the wavelet domain, figure (14) reflects the strength of hybrid EMD and the block-based scheme in surviving compression attacks. Some of the blocks can pass these attacks with large correlation values. On the other hand, the situation becomes worse with the hybrid EMD and SVD technique when it is subject to compression. The enhancement of watermark detects ability with hybrid EMD and block-based SVD watermarking is at the cost of speech quality deterioration. This is acceptable since the security issue is of major concern with moderate quality levels.

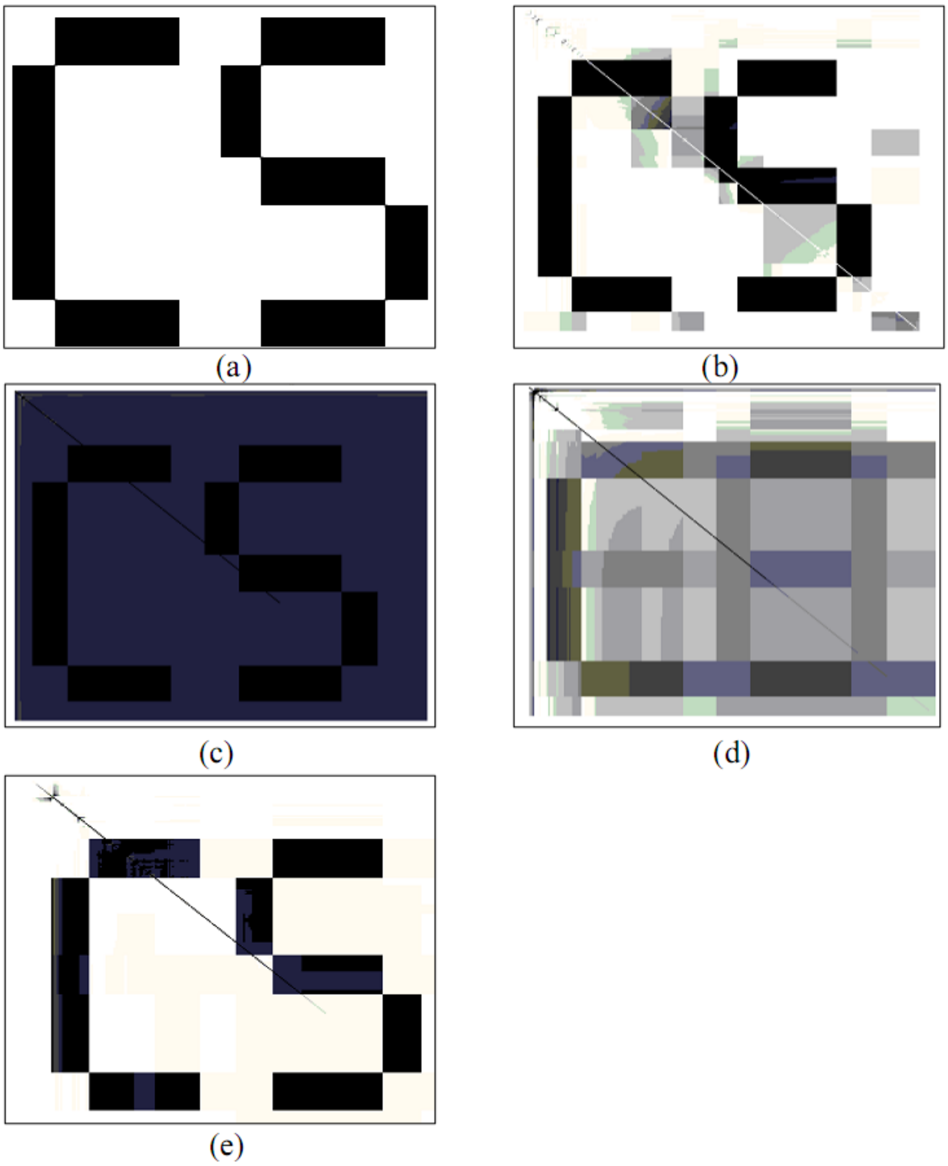

Figure 9 The extracted watermark image using the method of EMD and SVD in the (a) absence of attacks, (b) presence of WGN with SNR = 10dB, (c) presence of filtering attack with BW = 8kHz, (d) presence of cropping attack and (e) presence of wavelet compression attack with threshold level T = 0.8.

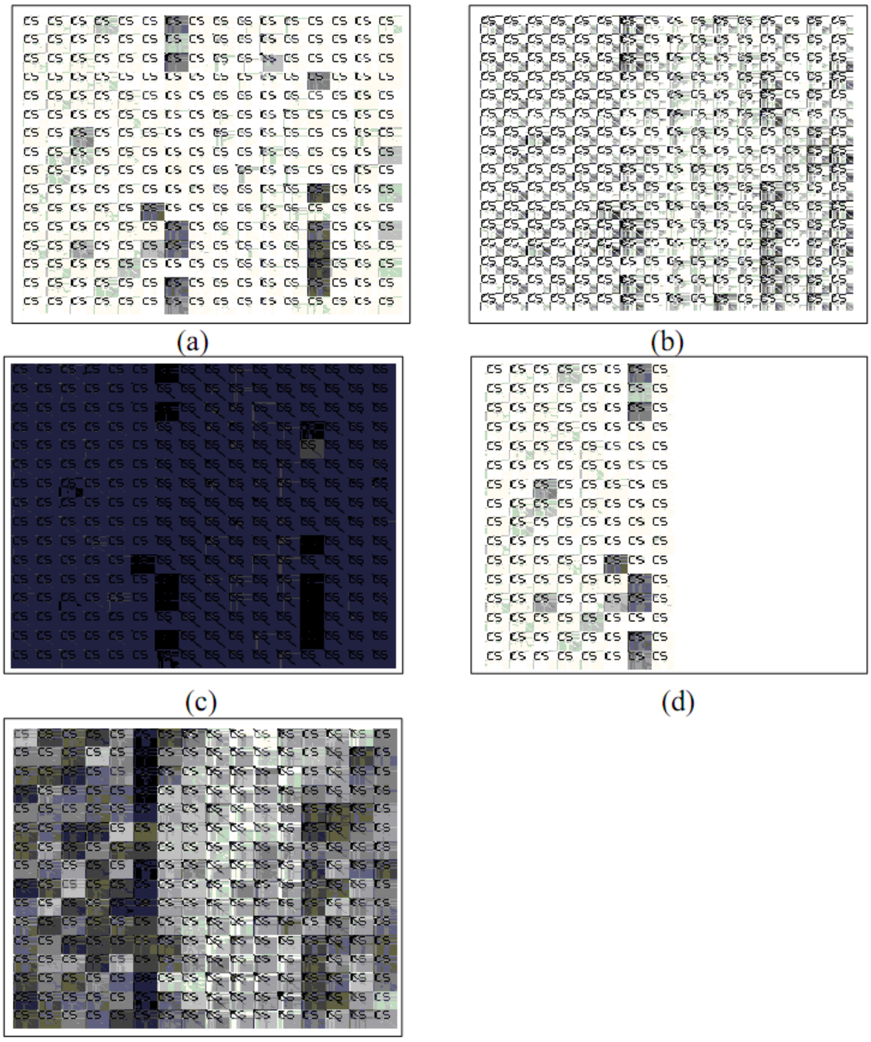

Figure (10) The extracted watermarks image using the method of EMD and block-based SVD in the (a) absence of attacks, (b) presence of WGN with SNR = 10dB, (c) presence of filtering attack with BW = 8kHz, (d) presence of cropping attack and (e) presence of wavelet compression attack with threshold level T = 0.8.

Figure (11) The magnification of the extracted watermark which achieves maximum correlation coefficient with the original watermark using the method of EMD and block-based SVD in the (a) absence of attacks, (b) presence of WGN with SNR = 10dB, (c) presence of filtering attack with BW = 8kHz, (d) presence of cropping attack and (e) presence of wavelet compression attack with threshold level T = 0.8.

Table (2) The SNR values between the original and the watermarked signals, Computer execution time and the correlation coefficient between the original and extracted watermarks in the absence of attacks in case of using EMD and SVD, EMD and block-based SVD, SVD and block-based SVD.

Speech Watermarking Method	SNR (dB)	Computer Execution Time (sec)	Correlation Coefficient
EMD SVD	20.1152	12.4921	1
EMD block-based SVD	14.8463	6.4520	1
SVD	28.0106	1.2249	1
Block-based SVD	15.9457	1.2909	1

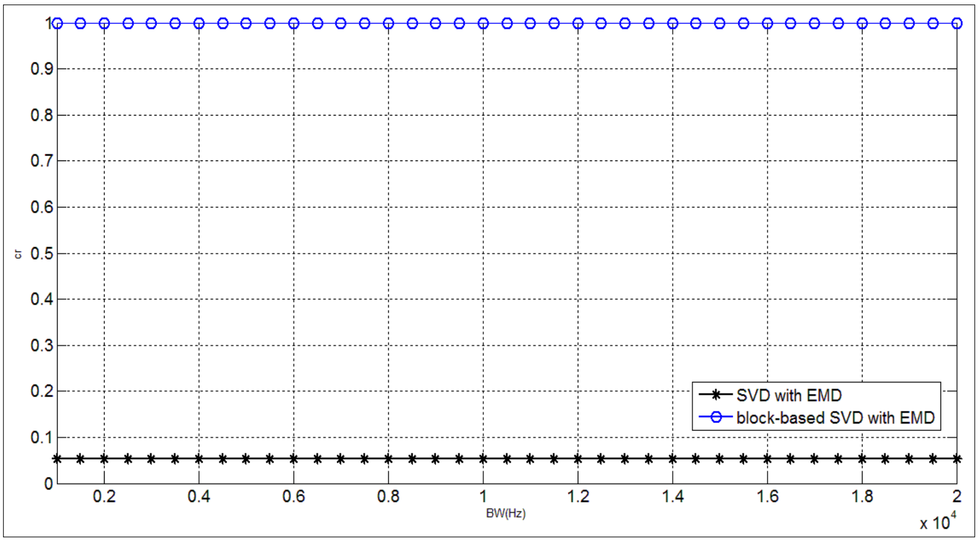

Figure (13)The variation of C_r with the BW when using EMD and SVD compared with EMD and block-based in the effect of filtering attack.

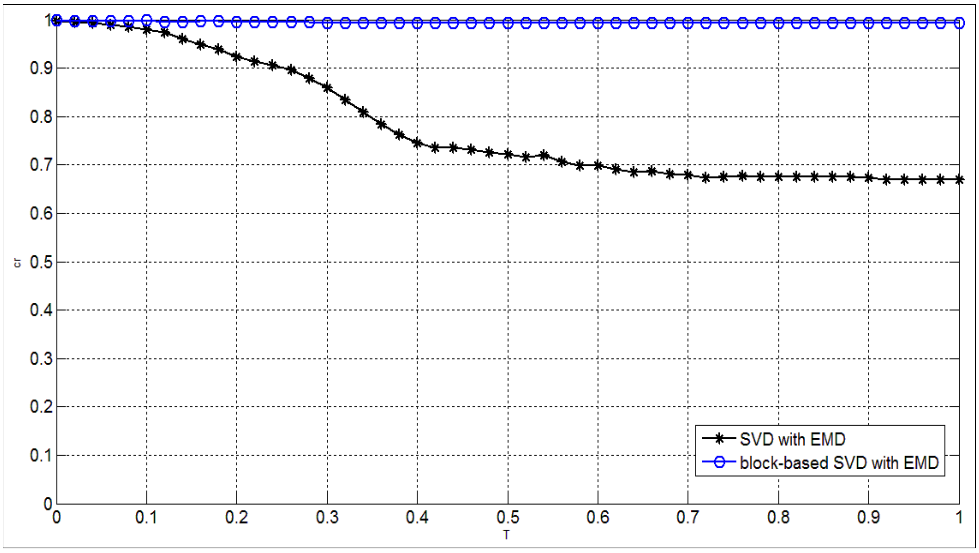

Figure (14) The variation of C_r with the compression threshold level when use EMD and SVD compared with EMD and block-based in the effect of wavelet compression attack.

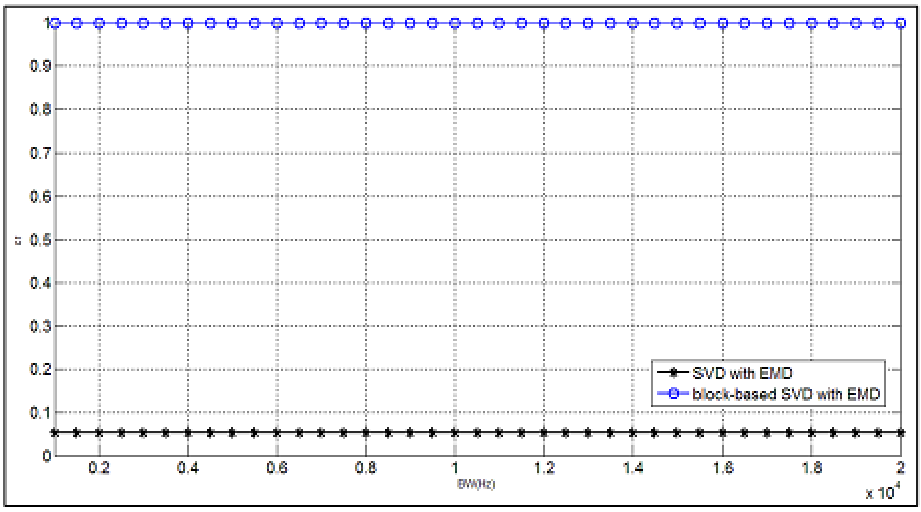

Figure (13) The variation of Cr with the BW when using EMD and SVD compared with EMD and block-based in the effect of filtering attack.

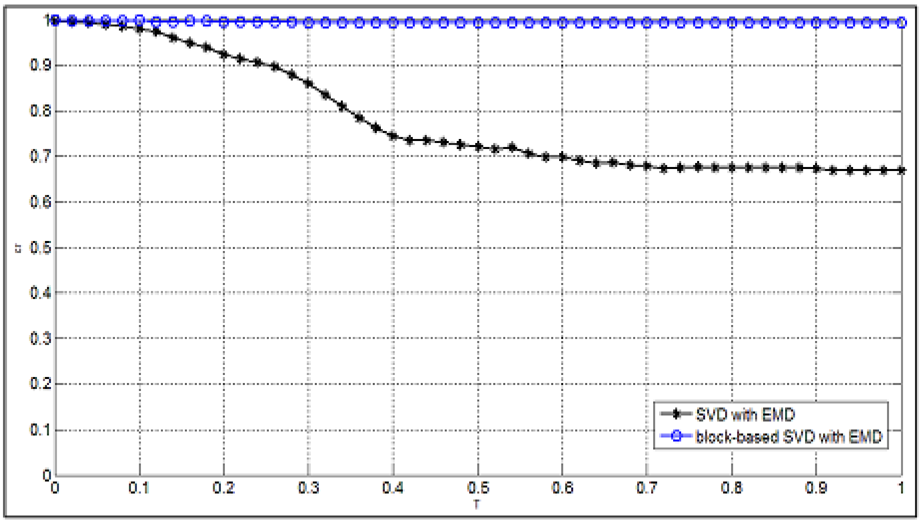

Figure (14) The variation of Cr with the compression threshold level when use EMD and SVD compared with EMD and block-based in the effect of wavelet compression attack.

Table (3) The correlation coefficient values in case of using the method of EMD and SVD compared with the method of EMD and block-based.

Attacks Type	EMD and SVD	EMD and block-based SVD
Without attacks	1	1
White Gaussian noise	0.5160	0.9589
Filtering	0.0536	0.9984
Cropping	0.4186	0.9975
Wavelet compression	0.6757	0.9930

5. Conclusions

This paper proposed a robust speech signal watermarking method with an efficient watermark extraction based on a hybrid EMD and SVD. The proposed method applied to the whole speech signal and to the block- based speech signal. The speech signal is decomposed into several Intrinsic Mode Functions (IMFs) using EMD then, the watermark is embedded into the singular values of thefirst IMF of the speech signal. From simulation results, it is clear that using EMD enhances the values of the correlation coefficient between original and extracted watermark,especially in the presence of attacks. It is clear that the values of C_r using EMD and block-based SVDare higher than those using EMD and SVD, especially in the presence of different attacks, and thus we used the EMD to get more robust watermark. This perfect watermark extraction comes on the expense of the speech signal quality, but this acceptable deterioration where it is not noticeable by subjective quality metrics.

References

[1] Mohammad Ali Nematollahi, ChaleeVorakulpipat andHamurabiGamboaRosale, "*Speech Watermaking*,"Springer Topics in Signal Processing, pp. 39-53, August 2016.

[2] Mohammad Ali Nematollahi and S. A. R. Al-Haddad, "*Distant Speaker Recognition: An Overview*," International Journal of Humanoid Robotics, vol. 13, pp. 1-45, issue. 02, June2016.

[3] Dong Hu and Ping Lee, "*Security Research on VoIP with Watermarking*," AIP conference proceedings, vol. 1060,pp. 98-101, issue. 1, November 2008.

[4] WojciechMazurczyk and ZbigniewKotulski, "*New VoIP Traffic Security Schema with Digital Watermarking*," International Conference of Computer Safety, Reliability and Security, pp. 170 – 181, 2006.

[5] Faundez – Zanuy M, Lucena – Molina JJ and Hagmuller M, "*Speech Watermarking: An Approach for the Forensic Analysis of Digital Telephonic Recordings*," Journal of Forensic Sciences, vol.55, pp. 1080-1087, July 2010.

[6] Shervin Shoki, Mohamod Ismail, NasharuddinZainal, "*Voice Quality in Speech Watermarking using Spread Spectrum Technique*," International Journal on Computer

and Communication Engineering, pp. 169-173, July 2012.

[7] Mohammad Ali Nematollahi, Mohammad Ali Akhaee, S.A.R. Al-Haddad and HamurabiGamboa – Rosales, " *Semi-fragile Digital Speech Watermarking for online Speaker Recognition,*" EURASIP Journal on Audio, Speech and Music Processing, pp. 1-15, December 2015.

[8] Doh.Suk Kim, "*Perceptual phase Quantization of Speech*," IEEE Transaction on Speech and Audio Processing, vol. 11, pp. 355-364, issue. 4, July 2003.

[9] Michael Arnold, Peter G.Baum and Walter VoeBing, "*A phase Modulation Audio Watermarking Technique*," International Workshop on Information Hiding, pp. 102-116, 2009.

[10] I. D. Shterev, "*Quantization-based Watermarking: Methods for Amplitude Scale Estimation Security and Linear Filtering Invariance,"* Delft Univ. of Technology, pp.14-25, 2007.

[11] Akira Nishimura, "*Reversible and Robust Audio Watermarking Based on Spread Spectrum and Amplitude Expansion*," International Workshop on Digital Watermarking, pp. 215-229, 2014.

[12] Qiang Cheng and J.SorenSen, "*Spread Spectrum Signaling for Speech Watermarking*," IEEE International Conference on Acoustic, Speech and Signal Processing, vol. 6,pp.1337-1340, May 2001.

[13] Frank Hartung and Bernd Girod, "*Digital Watermarking of MPEG-2 Coded Video in the Bitstream Domain*," Proceedings International Conference on Acoustics, Speech and Signal Processing, vol. 4, pp. 2621-2624, April 1997.

[14] Abd El-Samie FE," *Information Security for Automatic Speaker Identification*", 1st edition. Springer, pp. 66-92, 2011.

[15] Chin-Chen Chang, Yih-Shin Hu and Chia-Chen Lin, "*A Digital Watermarking Scheme Based on Singular Value Decomposition*," *Digital Signal Process*, vol. 21, pp. 522-527, 2011.

[16] Fathi E. Abd El-Samie, "*An Efficient Singular Value Decomposition Algorithm for Digital Audio Watermarking*," International Journal of Speech Technology, vol. 12, pp. 27–45, 2009.

[17] Ahmad A. Mohammad, Ali Alhaj and Sameer Shaltaf, "*An Improved SVD-based Watermarking Scheme for Protecting Rightful Ownership,*" Signal Processing, vol. 88, issue. 9, pp. 2158-2180, September 2008.

[18] Vivekananda Bhat K, IndranilSengupta andAbhijit Das, "*An Adaptive Audio Watermarking based on the Singular Value Decomposition in the Wavelet Domain,*" Digital Signal Processing, vol. 20, issue 6, pp.1547-1558, December 2010.

[19] Bai Ying Lei, IngYann Soon and Zhen Li, "*Blind and Robust Audio Watermarking Scheme based on SVD-DCT*", Signal Processing, vol. 91, issue 8, pp. 1973-1984, August 2011.

[20] Pranab Kumar Dhar and Tetsuya Shimamura, "*Blind SVD-based Audio Watermarking using Entropy and log-polar Transformation*," Journal of Information Security and Applications, vol. 20, pp. 74-83, February 2015,

[21] Zeiler A and et al, "*Empirical Mode Decomposition-an Introduction*," IEEE International Joint Conference on Neural Networks, pp.1-8, 2010.

[22] KaisKhaldi and Abdel-OuahabBoudraa, "*Audio Watermarking Via EMD*," IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, pp. 675-680, March 2013.

[23] Donghoh Kim andHee-Soek Oh, "*EMD: A Package for Empirical Mode Decomposition and Hilbert Spectrum,"*The R Journal, vol. 1/1, May 2009.

[24] NAWAL EL HAMDOUNI, ABDELLAHADIB, SONIA DJAZIRILARBI ANDMONIATURKI, "*HYBRID EMBEDDING STRATEGY FOR A BLIND AUDIO WATERMARKING SYSTEM USING EMD AND ISA TECHNIQUES*," 4TH INTERNATIONAL SYMPOSIUM ON COMMUNICATIONS, CONTROL AND SIGNAL PROCESSING (ISCCSP), PP.1-4, MARCH 2010.

[25] S. WANG, A. SEKEY AND A. GERSHO, "*AN OBJECTIVE MEASURE FOR PREDICTING SUBJECTIVE QUALITY OF SPEECH CODERS,*" IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 10, NO. 5, PP. 819–829, 1992.

[26] VOLODYA GRANCHAROV AND W. BASTIAAN KLEIJN, "*SPEECH QUALITY ASSESSMENT*," PROCEEDINGS OF THE IEEE PACIFIC RIM CONF. ON COMMUNICATIONS, COMPUTERS AND SIGNAL PROCESSING, PP. 125-128. 1993.