

Presenting a Novel Audio Watermarking Based on Discrete Wavelet Transform

Mahdi Mosleh, Saeed Setayeshi, and Mohammad Mosleh

Abstract—Digital watermarking techniques are attracting attention as a proper solution to protect copyright for multimedia data. In this paper, we propose a new audio watermarking method, in wavelet coefficients in the way that human auditory system is not sensitive to it. The proposed method can embed the watermark data into levels 3-4 approximation coefficients of wavelet transform using echo and discrete Fourier transform. The extraction algorithm of the proposed method extracts two watermark sequences from watermarked data. The final watermark sequence is obtained by extracted sequence using a hidden key. The experimental results show that the proposed method is very robust to noise. In order to evaluate, the proposed method is compared with the methods based on FFT, Echo and AM

Index Terms— Audio Watermarking, Discrete Fourier Transform (DFT) Discrete Wavelet, Transform (DWT)

I. INTRODUCTION

With the recent development of digital information technology, all personal computer owners enjoy a multimedia processor, a vast band having access to the entire world and an interchangeable memory for digital information. Therefore, digital information is accessible to everyone, and they are distributed easily. With broadcasting information transfer by internet network, the firms can present their digital products by internet network. Although this technology development causes many works to do easy, it has caused some problem as next technologies. Existing different devices such as printing- machines and copying media make the copying of these data easy[1]. Now there are two standard methods to protect copyright of digital documents such as coding and digital signature.

The major problem of coding method is that it does not protect the document placing in the stage after coding and is useable. In other words, the user, after applying certain key and decoding the document, can change document in any way or make unauthorized copies of it in numbers. So, document owner is not able to claims his right. If it is made a little change intentionally or not in digital document even one bit of it which uses digital signature to protect copyright, it

will be impossible for receiver to retrieve digital signature. As a result, the document will lose its credit easily[2].

Studies show that the only solution to this problem is to add security information to document in such a way that it doesn't separate from document during its life time and also remains undetectable to document user as much as possible. One of the proper solutions is digital watermarking. Watermarking is a technique that embeds imperceptible and statistically undetectable information into digital data (e.g. video, images and audio signals). This embedded information contains certain information (signature, logo, ID number, etc) which uniquely related to the owner or distributor.

Digital watermarking was invented by one of the Muzac firm engineers, Emil Hem brook, in 1954. In this invention an ID code was added to digital music file such a way that it was undetectable and imperceptible to proof the ownership[3]. After that, digital watermarking was applied in many ways, but this technique did not attract scientists' attention as an important research issue until 1990. At the beginning of 1990, it was attracting attention as an interesting research subject which has retained its importance until now[4].

During last decade, different applications of digital watermarking have been proposed in various fields such as text[5], audio[6-9], image[10, 11] and video[12, 13]. These applications can be classified from different aspects. From viewpoint the perceptible watermark into digital medium, two methods can be proposed: perceptible watermark into digital medium [14, 15] and imperceptible watermark into digital medium[12, 16]. Based on robustness against various attacks, watermarking systems can be classified into three categories consist of robust watermarking systems[17, 18], semi-robust watermarking systems[19] and fragile watermarking systems[20].

In this paper, we focus on audio watermarking. Some of the desirable characteristics of a watermarking algorithm are as follows: inaudibility, robustness against signal processing attacks and being undetectable by unauthorized persons. Although various video and image watermarking methods have been proposed, since Human Auditory System (HAS) is more sensitive than Human Visual System (HVS), developing a good watermarking system for embedding watermark into audio signal is more difficult than image one. Human auditory system is able to sense audio distortions with power effect and low frequency and also it is very sensitive to white noise mixing with signal[21]. One of the human auditory system features is masking, it means that one strong audio removes the auditory effect of another audio. Human auditory system is not sensitive to phase changes and phase change distortions are inaudible by man. These two features (making and not to be sensitive to phase changes) are

Manuscript received February 14, 2011; revised July 11, 2011.

Mahdi Mosleh is with Department of Computer Engineering, Islamic Azad University, Dezful Branch, Dezful, Iran (email:mahdi_mosleh@yahoo.com).

Saeed Setayeshi is with Faculty of Nuclear Engineering and Physics, Amirkabir University, Tehran, Iran (email: setayesh@aut.ac.ir).

Mohammad Mosleh is with Department of Computer Engineering, Islamic Azad University, Dezful Branch, Dezful, Iran (email: mosleh@iaud.ac.ir).

used to develop watermarking systems[22].

Developing a proper watermarking algorithm is very difficult. Audio watermarking methods can be classified into two main categories consist of time domain based methods and transform domain based methods. Time domain based technique is easier to implement in comparison with transform domain based method and needs less accounting resources. In other words, time domain based watermarking systems are not robust enough against signal processing attacks in comparison with transform domain based systems. AM[23], Echo[24], methods are some kinds of techniques based on time domain. These methods are easy to implement but are not robust enough against attacks. The second category is methods based on transform domain which are more robust than time domain methods. Some of them include methods based on Fast Fourier Transform (FFT)[25], methods based on Discrete Cosine Transform (DCT)[26] and Discrete Wavelet Transform (DWT) base methods [27]

In this paper, we propose a new audio watermarking method for embedding watermark sequence into approximation coefficients 3-4 levels of DWT, by echo and FFT methods, in order to decrease audibility and increase robustness against attacks.

II. DISCRETE WAVELET TRANSFORM

DWT is utilized in a wide range of DSP applications including audio/image/video compression, data communication over the Internet, pattern recognition and numerical analysis[28]. The basic idea of DWT for one-dimensional signals is briefly described. A signal is split into two parts, usually the high frequency (Detail coefficient) and the low frequency part (Approximation coefficients). This splitting is called decomposition. The edge components of the signal are largely confined to the high frequencies part. The signal is passed through a series of high pass filters to analyze the high frequencies, and it is passed through a series of low pass filters to analyze the low frequencies. Filters of different cutoff frequencies are used to analyze the signal at different resolutions. Let us suppose that $x[n]$ is the original signal, spanning a frequency band of 0 to π rad/s. The original signal $x[n]$ is first passed through a half band high pass filter $g[n]$ and a low pass filter $h[n]$. After the filtering, half of the samples can be eliminated according to the Nyquist's rule, since the signal now has the highest frequency of $\pi/2$ radians instead of π . The signal can therefore be subsample by 2, simply by discarding every second sample. This constitutes one level of decomposition and can mathematically be expressed as follows:

$$y_{high}[k] = \sum_n x[k].g[2k - n] \quad (1)$$

$$y_{low}[k] = \sum_n x[k].h[2k - n] \quad (2)$$

where $y_{high}[k]$ and $y_{low}[k]$ are the outputs of the high pass and low pass filters, respectively, after sub sampling by 2.

The above procedure can be repeated for further decomposition. The outputs of the high pass and low pass filters are called DWT coefficients and by these DWT coefficients the original image can be reconstructed. The reconstructed process is called the Inverse Discrete Wavelet

Transform (IDWT). The above procedure is followed in reverse order for the reconstruction. The signals at every level are up sampled by two, passed through the synthesis filters $g'[k]$ and $h'[k]$ (high pass and low pass, respectively), and then added. The analysis and synthesis filters are identical to each other, except for a time reversal. Therefore, the reconstruction formula becomes (for each layer).

$$x[n] = \sum_n (y_{high}[k].g[-n + 2k] + y_{low}[k].h[-n + 2k]) \quad (3)$$

To ensure the above IDWT and DWT relationship the following orthogonality condition on the filters $H(\omega)$ and $G(\omega)$ must hold:

$$|H(\omega)|^2 + |G(\omega)|^2 = 1 \quad (4)$$

where

$$H(\omega) = \sum_n h[n]e^{-j\omega n}$$

and

$$G(\omega) = \sum_n g[n]e^{-j\omega n}$$

III. THE PROPOSED METHOD

In the following of the paper we describe the embedding and extraction algorithms of the proposed method. As usual, audio signals are analyzed in short-time form because the audio signals are non-stationary and their properties change during the time. Therefore, the embedding and extraction algorithms are implemented for each frame.

A. Watermark Embedding Procedure

Our proposed embedding scheme has been shown in Fig. 1.

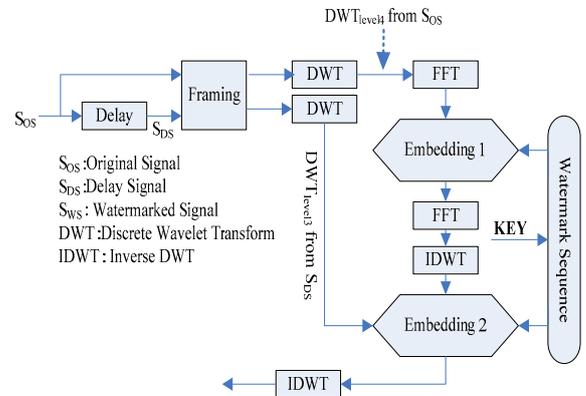


Fig.1. Block diagram of watermarked embedding scheme

Watermark sequence embedding algorithm is explained in several steps as follows:

Step 1: The calculation of delay signal by original signal

$$S_{DS} = S_{OS}[n - d] \quad (5)$$

Step 2: Farming both original and delay signals.

Step3: Levels 3-4 DWT are obtained from each frame of original and delay signal. Approximation and detail coefficients of level 3 for both original and delay signal are named respectively as (CAOS3, CDOS3) and (CADS3, CDDS3). Also, Approximation and detail coefficients of level 4 for original signal are presented as (CAOS4, CDOS4).

Step4: Regarding to the watermark data, approximation coefficients watermark signal in level 4 is calculated according to Eq. (6).

$$CA_{WS4} = \begin{cases} FFT(CA_{OS4}) + coeff & w = 1 \\ FFT(CA_{OS4}) - coeff & w = 0 \end{cases} \quad (6)$$

where *coeff* term is a coefficient between [0 1].

Also, it is mentionable which watermark sequence is shaped by a random number generator using a hidden key.

Step 5: Inverse wavelet transform is obtained from the wavelet coefficients in former step.

The approximation coefficients level 3 of watermark signal are calculated by using the calculated approximation coefficients in the former step and approximation coefficient level 3 of delay signal according to Eq.(7) :

$$CA_{WS3} = \begin{cases} CA_{WS3} + coeff.CA_{DS3} & w = 1 \\ CA_{WS3} - coeff & w = 0 \end{cases} \quad (7)$$

Step 6: The inverse wavelet transform is obtained from the wavelet coefficients of watermark signal. Finally, watermarked signal frames are calculated.

B. Watermark extraction procedure

Fig.(2) illustrates the proposed block diagram for extracting the watermark sequence from watermark signal.

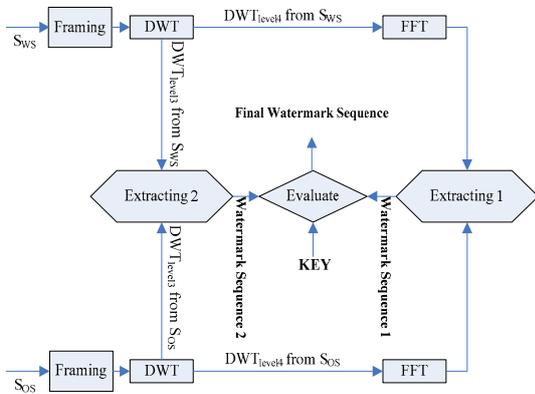


Fig.2. Block diagram of watermarked extraction scheme

The watermark sequence extraction process is mentioned in some steps as follows:

Step1: Framing the watermarked signal and original signal.

Step2: Level 3 DWT is obtained from each frames of watermarked and original signal. Both original signal and watermarked signal approximation coefficients in level 3 are compared with each other according to Eq.(8) and so the first watermark sequence is extracted.

$$w1_{extracted} = \begin{cases} 1 & \text{if } Abs[Sum(Abs(CA_{WS3})) - Sum(Abs(CA_{OS3}))] < -0.01 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Step 3: level 4 DWT is obtained from each of watermarked and original signal frames. Regarding to the sum of coefficients FFT difference, according to Eq.(9), the second watermark sequence is extracted.

$$w2_{extracted} = \begin{cases} 1 & \text{if } Sum[real(FFT(CA_{WS4}))] - Sum[real(FFT(CA_{OS4}))] > 0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

Step 4: The final watermark sequence is formed through two extracted watermark sequences using user hidden key. In other word, the final watermark sequence can be computed

using the hidden key by two extracted sequences. Notice that only those who have access to this hidden key can extract the final watermark sequence

IV. EXPERIMENTAL RESULTS

In order to evaluate the proposed method, two speech and one music audio files with wave format were used. These files were 16- bits mono with sampling frequency 16 KHZ. Watermark sequence was formed by a pseudorandom number generator using a hidden key. In all experiments, every audio file was divided into frames with 512 samples. All implementations were done by Matlab software. The waveforms of an example original signal file with its watermarked and attacked signals have been shown in Fig.3.

In order to evaluate the robustness of the proposed method, seven attacks were applied including Additional Noise (time), Additional Noise (frequency), Cropping, Re-quantization, MP3, Re-sampling and Filtering. Bit error rate is calculated by Eq.(10):

$$BER(\%) = \frac{\text{Number of Erroneously decoded bits}}{\text{Number of embedding bits for the clip}} \times 100 \quad (10)$$

The proposed method was compared with AM, Echo and FFT methods. Table (I) shows the experimental results for two speech audio files (a female and a male) and one music audio file. Also, error rate percentage of AM, Echo and FFT methods against mentioned attacks have been shown in Fig. 4. It should be mentioned that S/N ratio, after inserting watermark sequence, in all methods is about 20dB.

As seen from Fig.4, performance of the proposed method against four attacks Cropping, MP3, Resampling and Filtering is too better than other algorithms.

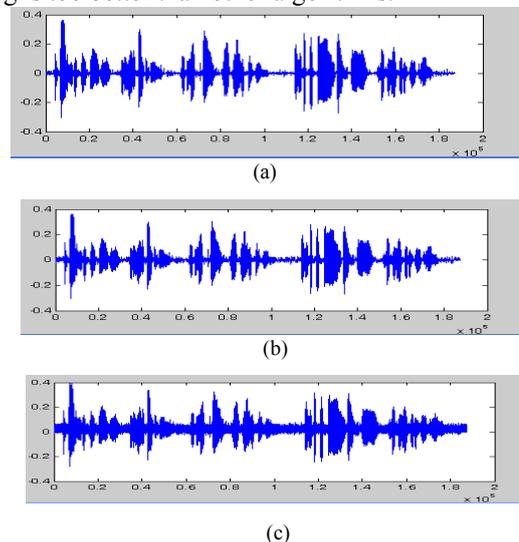


Fig.3. The wave form of (a) original signal (b) watermarked signal (c) attacked signal with additional noise(time)

TABLE I. THE OBTAINED RESULTS OF APPLYING DIFFERENT ATTACKS OVER

(a) An example male speech (b) An example female speech (c) An example music signal

Attacks	AM method	Ech o method	FFT method	Propo sed method
Additional	45.8	51/7	0	2/72

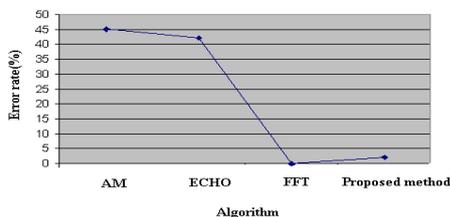
noise(time)		7		
Additional noise (frequency)	47/9	50/68	0	0
Cropping	29/72	34/03	29/7	12/8
Quantization	48/05	28/05	0	2/44
MP3	48/95	53/13	49/5	21/19
Resampling	48/65	22/88	9/53	0
Filtering	48/20	36/51	25/88	10/89

(b)

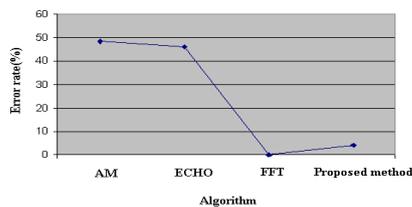
Attacks	AM method	Echo method	FFT method	Proposed method
Additional noise(time)	40/69	28/68	0	1/02
Additional noise (frequency)	50	45/83	0/387	0/387
Cropping	31/78	29/4	27/9	12/4
Quantization	50/38	22/86	0	0
MP3	50/45	46/89	49/61	24/8
Resampling	50/22	24/41	10/07	0
Filtering	50/35	37/59	32/55	13/56

(c)

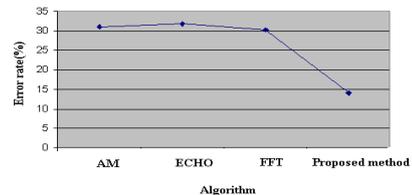
Attacks	AM method	Echo method	FFT method	Proposed method
Additional noise(time)	48/84	46/08	0	2/73
Additional noise (frequency)	47/01	41/93	0	12/9
Cropping	31/79	31/79	33/17	17/05
Quantization	48/84	16/58	0	12/44
MP3	48/38	47/46	51/61	21/19
Resampling	48/44	14/74	15/2	13/36
Filtering	48/84	24/88	33/64	17/51



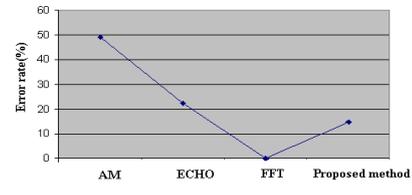
(a)



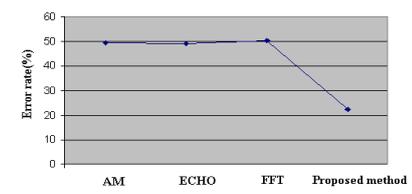
(b)



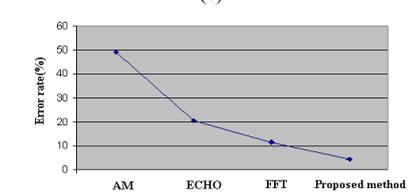
(c)



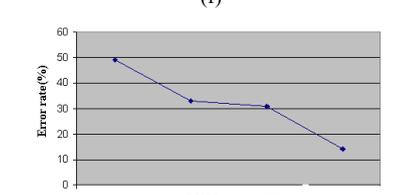
(d)



(e)



(f)



(g)

Fig.4. Attacks Influence over AM, Echo, FFT and Proposed methods (a) Additional Noise(time) (b) Additional Noise(frequency) (c) Cropping (d) Quantization (e) MP3 (f) Re sampling (g) Filtering

V. CONCLUSION

Due to dynamics of audio signal and also sensitivity of Human Auditory System (HAS), audio watermarking is more complex in comparison with image and video watermarking. By now, various techniques have been reported in this field but most of them are not enough robust against signal processing attacks. In this paper, we propose a new method for audio signal watermarking in wavelet domain. The proposed method is able to embed a watermark sequence into two different levels of wavelet transform, based on Echo and FFT methods, and therefore is a proper robustness against several signal processing attacks. The experimental results show that the robustness of the proposed method is more proper in comparison with AM, Echo and FFT methods.

REFERENCES

[1] S. J. Lee, and S. H. Jung, "A survey of watermarking techniques applied to multimedia," in ISIE 2001, 2001, pp. 272-277 vol. 1.

- [2] S. Katzenbeisser, and F. A. P. Petitcolas, *Information hiding techniques for steganography and digital watermarking*: Artech house, 2000.
- [3] R. Chandramouli, and N. Memon, "Analysis of LSB based image steganography techniques," in 2001 International Conference on Image processing, 2001, pp. 1019-1022 vol. 3.
- [4] I. J. Cox, and M. L. Miller, "The first 50 years of electronic watermarking," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 1, pp. 126-132, 2002.
- [5] Y. W. Kim, K. A. Moon, and I. S. Oh, "A text watermarking algorithm based on word classification and inter-word space statistics," *Document Analysis and Recognition*, vol. 2, pp. 775, 2003.
- [6] V. Bhat K, I. Sengupta, and A. Das, "An audio watermarking scheme using singular value decomposition and dither-modulation quantization," *Multimedia Tools and Applications*, pp. 1-15.
- [7] H. Kim, Y. Choi, J. Seok *et al.*, "Audio watermarking techniques," *Intelligent Watermarking Techniques*, pp. 185-218, 2004.
- [8] A. Gurijala, and J. Deller, "Robust algorithm for watermark recovery from cropped speech," in IEEE international conference on acoustics speech and signal processing, 2001.
- [9] S. Shin, O. Kim, J. Kim *et al.*, "A robust audio watermarking algorithm using pitch scaling," in 4th International Conference on Digital Signal Processing, 2002, pp. 701-704.
- [10] H. Inoue, A. Miyazaki, and T. Katsura, "An image watermarking method based on the wavelet transform," in international Conference on Image Processing, 1999, pp. 296-300 vol. 1.
- [11] C. S. Lu, and H. Y. M. Liao, "Multipurpose watermarking for image authentication and protection," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1579-1592, 2001.
- [12] N. Checcacci, M. Barni, F. Bartolini *et al.*, "Robust video watermarking for wireless multimedia communications." pp. 1530-1535 vol. 3.
- [13] F. Hartung, and B. Girod, "Watermarking of uncompressed and compressed video," *Signal Processing*, vol. 66, no. 3, pp. 283-301, 1998.
- [14] M. D. Swanson, B. Zhu, B. Chau *et al.*, "Object-based transparent video watermarking," in IEEE First Workshop on Multimedia Signal Processing, 1997, pp. 369-374.
- [15] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Transparent robust image watermarking," in International Conference on Image Processing, 1996, pp. 211-214 vol. 3.
- [16] T. Furon, and P. Duhamel, "Robustness of asymmetric watermarking technique," in International Conference on Image Processing, 2000, pp. 21-24 vol. 3.
- [17] P. J. Lee, and M. J. Chen, "Robust error concealment algorithm for video decoder," *IEEE Transactions on Consumer Electronics*, vol. 45, no. 3, pp. 851-859, 1999.
- [18] M. Ramkumar, and A. N. Akansu, "A robust protocol for proving ownership of multimedia content," *IEEE Transactions on Multimedia*, vol. 6, no. 3, pp. 469-478, 2004.
- [19] D. He, Q. Sun, and Q. Tian, "A semi-fragile object based video authentication system," in International Symposium on Circuits and Systems, 2003, pp. III-814-III-817 vol. 3.
- [20] J. Fridrich, M. Goljan, and A. C. Baldoza, "New fragile authentication watermark for images," in International Conference on Image Processing, 2000, pp. 446-449 vol. 1.
- [21] F. Hartung, and M. Kutter, "Multimedia watermarking techniques," in Proceedings of the IEEE, 1999, pp. 1079-1107.
- [22] L. Boney, A. H. Tewfik, and K. N. Hamdy, "Digital watermarks for audio signals." p. 0473.
- [23] W. N. Lie, and L. C. Chang, "Robust and high-quality time-domain audio watermarking subject to psychoacoustic masking." pp. 45-48 vol. 2.
- [24] W. Bender, D. Gruhl, N. Morimoto *et al.*, "Techniques for data hiding," *IBM systems journal*, vol. 35, no. 3/4, pp. 313-336, 1996.
- [25] M. Fallahpour, and D. Megias, "High capacity audio watermarking using FFT amplitude interpolation," *IEICE Electronics Express*, vol. 6, no. 14, pp. 1057-1063, 2009.
- [26] X. Y. Wang, and H. Zhao, "A novel synchronization invariant audio watermarking scheme based on DWT and DCT," *Signal Processing, IEEE Transactions on*, vol. 54, no. 12, pp. 4835-4840, 2006.
- [27] M. Pooyan, and A. Delforouzi, "Adaptive and robust audio watermarking in wavelet domain," in Third International Conference on International Information Hiding and Multimedia Signal Processing, 2007, pp. 287-290.
- [28] S. G. Mallat, *A wavelet tour of signal processing*: Academic Pr, 1999.



Mahdi Mosleh, received his B.S. in computer engineering from Islamic Azad University, Dezful Branch, in 2008 and the M.S. in computer engineering from Islamic Azad University, Dezful, in 2011 in computer engineering. His research interests are Audio and Image watermarking.



Saeed Setayeshi, B.Sc., M.Sc., M.A., M.A.Sc., Ph.D. (Electrical & Computer Eng., TUNNS, Canada, 1993) is an Associate Professor and teaching in the Faculty of Nuclear Engineering and Physics, Amirkabir University of Technology (Tehran Polytechnics). He is also presenting some courses in Computer Eng. Department of Res. & Sc. Branch of IAU as an Invited Professor. He has founded AL and Complex System Researches for first time there and supervised many Graduate students in the areas of using CA and LA. He has published more than 100 papers in ISI Journals and Conferences. His research interests are in the areas of AI & AL, Intelligent Control (Neural-Fuzzy-Expert-GA-CA-LA), Adaptive Signal Processing, Agent Based Modeling, Artificial Society, Social Evolution, Wealth Distribution, Knowledge Based Systems, and Dynamics of Complex Systems.



Mohammad Mosleh, received his B.S. in computer engineering from Islamic Azad University, Dezful Branch, in 2003, the M.S. in computer engineering from Islamic Azad University, Tehran, in 2006 and the PhD degree in computer engineering at the Islamic Azad University, Tehran, in 2010. He is assistant professor in the Department of Technical and Engineering at the Islamic Azad University, Dezful Branch. His research interests are in the areas of Speech Processing, Machine Learning and Parallel Processing.