DWT–Based Audio Watermarking

Ali Al-Haj[1], Ahmad Mohammad[1] and Lama Bata[2]
[1]College of Electrical Engineering, Princess Sumaya University, Jordan
[2]Faculty of Engineering and Technology, University of Jordan, Jordan

Abstract: *Many effective watermarking algorithms have been proposed and implemented for digital images and digital video, however, few algorithms have been proposed for audio watermarking. This is due to the fact that, the human audio system is far more complex and sensitive than the human visual system. In this paper, we describe an imperceptible and robust audio watermarking algorithm based on the discrete wavelet transform. Performance of the algorithm has been evaluated extensively, and simulation results are presented to demonstrate the imperceptibility and robustness of the proposed algorithm.*

Keywords: *Multimedia, copyright protection, audio watermarking, imperceptibility, robustness, transform-domain watermarking, and discrete wavelet transform.*

Received September 16, 2009; accepted January 3, 2010

1. Introduction

By virtue of the new advancements in computer and telecommunication networks, multimedia files are produced, stored and distributed easily across the globe. However, the ownership and copyright of multimedia files are not usually protected. Digital watermarking has been proposed in recent years as a means of protecting multimedia contents from intellectual piracy. This is achieved by modifying the original content, by inserting a signature which can be extracted, when necessary, as a proof of ownership. Indeed, many effective digital image and video watermarking algorithms have been proposed and implemented at a commercial scale [9]. However, and due to the fact that the human audio system is far more complex and sensitive than the human visual system, few algorithms have been proposed for audio watermarking [4].

Audio watermarking techniques reported in literature can be grouped into two types; time-domain techniques and frequency-transform domain techniques [1, 3]. The two domains have different characteristics, and thus performances of their techniques may vary with respect to the robustness and imperceptibility (inaudibility) requirements of audio watermarking. Inaudibility refers to the condition that the embedded watermark should not produce audible distortion to the sound quality of the original audio, in such a way that the watermarked marked version of the file is indistinguishable from the original one. Robustness determines the resistance of the watermark against removal or degradation. The watermark should survive malicious attacks such as random cropping and noise adding, and its removal should be impossible without perceptible signal alterations.

Time-domain techniques include the Least Significant Bit substitution (LSB) and echo hiding techniques, among many others [5, 7]. LSB embeds the watermark information in the least significant bits of the audio sample values by overwriting the original bits [7, 10]. It takes advantage of the quantization error that usually derives from the task of digitizing the audio signal. On the other hand, echo watermarking attempts to embed information into the original discrete audio signal by introducing a repeated version of a component of the audio signal with small offset, initial amplitude and decay rate to make it imperceptible [12, 15]. In general, time-domain audio watermarking is relatively easy to implement, and requires few computing resources, however, it is weak against signal processing attacks such as compression and filtering.

Frequency domain audio watermarking techniques employ human perceptual properties and frequency masking characteristics of the human auditory system for effective watermarking [14]. In these techniques, the phase and amplitude of the transform domain coefficients are modified in a certain way to carry the desired watermark information. Popular transforms include the Discrete Fourier Transform (DFT), the Discrete Cosine Transform (DCT), and the Discrete Wavelets Transform (DWT). In [22], the Fourier transform magnitude coefficients over the frequency range from 2.4 KHz to 6.4 KHz are replaced with the watermark sequence since human sensitivity declines compared to its peak around 1 KHz . Moreover, human ears are relatively insensitive to phase distortion, and especially lack the ability to perceive the absolute phase value, therefore in [23], the watermark is represented by the relative phase between selected coefficients and their neighbors. The problem withthese watermarking schemes that they are less robust to signal processing and malicious attacks, such as audio compression.

Other than time-domain and frequency domain techniques, spread-spectrum watermarking methods are becoming popular. These methods, embed a narrow-band signal (the watermark) into a wide-band channel (the audio file) to spread the watermark data across the large frequency band, namely the audible spectrum [8, 16, 21]. Watermark detection is done by calculating the correlation between the watermarked audio signal and the watermark signal. Finally, Patchwork methods [25] use pseudorandom processes to embed a certain statistics into a data set which is detected in the reading process with the help of numerical indexes, like the mean, describing the specific distribution. Computational complexity of these methods is very high, and synchronization is difficult to implement.

In this paper, we propose an effective audio watermarking algorithm that is based on the DWT. The DWT transform decomposes the host audio signal into several multi-resolution sub-bands, enabling algorithm developers to locate the most appropriate sub-bands for embedding the watermark bits. In the proposed algorithm, the watermark bits are embedded in the high-resolution sub-bands of the audio signal, so that satisfactory robustness and imperceptibility (inaudibility) performances are obtained.

The rest of the paper is organized as follows. In section two, we give a brief description of the discrete wavelets transform. In section three, we describe in details the watermarking embedding and extraction procedures of the proposed algorithm. In section four, we evaluate the performance of the algorithm and present simulation results with respect to inaudibility and robustness. We conclude in section five with some remarks.

2. The Discrete Wavelets Transform

Wavelets are special functions which, in a form analogous to sines and cosines in Fourier analysis, are used as basal functions for representing signals. They provide powerful multi-resolution tool for the analysis of non-stationary signals with good time localization information. The coefficients of the discrete wavelet transform can be calculated recursively and in a straight forward manner using the well-known Mallat's pyramid algorithm [17]. Based on this algorithm, the one-dimensional discrete wavelet coefficients of any stage can be computed from the coefficients of the previous stage using the following iterative equations:

$$WL(n,j) = \sum_m WL(m,j-1)h_0(m-2n) \quad (1)$$

$$WH(n,j) = \sum_m WL(m,j-1)h_1(m-2n) \quad (2)$$

where $W_L(n,j)$ is the n^{th} scaling coefficient at the j^{th} stage, $W_H(n,j)$ is the n^{th} wavelet coefficient at the j^{th} stage, and $h_0(n)$ and $h_1(n)$ are he dilation coefficients corresponding to the scaling and wavelet functions, respectively. Equation 1 can then be used for obtaining the wavelet coefficients of subsequent stages. In practice this decomposition is performed only for a few stages.

In order to reconstruct the original data, the discrete wavelet transform coefficients are up-sampled and passed through another set of low pass and high pass filters, which is expressed as:

$$WL(n,j) = \sum_k WL(k,j+1)g_0(n-2k) + \sum_l WH(l,j+1)g_1(n-2l) \quad (3)$$

where $g_0(n)$ and $g_1(n)$ are respectively the low-pass and high-pass synthesis filters corresponding to the mother wavelet. It is observed from Equation 3 that the j^{th} level coefficients can be obtained from the $(j+1)^{th}$ level coefficients.

Due to its excellent spatio-frequency localization properties, the DWT is very suitable to identify areas in an audio signal where a watermark can be embedded effectively. Some DWT-based audio watermarking techniques can be found in literature [11, 18, 21].

3. The DWT-Based Algorithm

The algorithm we propose here is based on applying the Discrete Wavelet Transform (DWT) on the digital audio signal in which a watermark is to be embedded. The algorithm consists of two procedures; watermarking embedding procedure and watermarking extraction procedure.

3.1. Watermark Embedding Procedure

The embedding procedure performs three major operations; watermark pre-processing, DWT-based frequency decomposition of the audio signal, and watermark embedding in the DWT-transformed audio signal. The operations are described in the following steps.

1. Express the grey-scale image watermark as a two-dimensional matrix whose size is *M1 x M2*.

$$Img = \{Img(k,j), \quad 0 \le k \le M1, \quad 0 \le j \le M2\} \quad (4)$$

Convert the two-dimensional image matrix *Img* into a one-dimensional vector *W* of length *M1 x M2*.

$$W = \{W(i) = Img(k,j), i = k \, x \, M2 + j, \; 0 \le k \le M1, \; 0 \le j \le M2\} \quad (5)$$

2. Normalize the one-dimensional vector *W* by dividing each element by 255.

$$Wn_i = Wi/255; \quad (M2+1) \le i \le ((M1 \times M2) + M2) \quad (6)$$

3. Apply a two-level DWT to the left-channel of the stereo audio signal. The DWT operation produces the one-dimensional sub-bands shown below in Figure 1, where *A2* is the approximation sub-band, and *D1* and *D2* are the first-level and second-level details sub-bands, respectively. Its important to note here that the decision was made to adopt the two-level DWT since it gave better results than higher DWT levels.

A2	D2	D1

Figure 1. Output sub-bands of the two-level DWT operation.

4. Embed the normalized watermark vector W_n, expressed by Equation 6, into the second-level details sub-band D2 of the left channel. The embedding procedure produces the watermarked second-level details sub-band (WD2) as given in equations 7 and 8.

$$WD2=\begin{cases}(D2_i+Wt_j); \quad i=\{1,120,240...L2\}, \text{ and} \\ \qquad (M2+1)\le j\le((M1\times M2)+M2) \\ D2_i; \qquad i=\{(2:219),(121:239)...(L2-119:L2\}, \\ \qquad and(M2+1)\le j\le((M1\times M2)+M2)\end{cases} \tag{7}$$

where $Wt_i = Wn_i\times\alpha;\ (M2+1)\le i\le((M1\times M2)+M2)$ (8)

3.2. Watermark Extraction Procedure

The watermark extraction procedure enables the owner of the audio clip to extract the embedded watermark. The procedure requires knowledge of the original audio file, the watermark intensity, and the size of the watermark, in order to extract the watermark. The watermark extraction steps are a direct reversal of steps carried out in the embedding procedure. Extraction steps are described as follows:

1. Apply a two-level DWT operation on both the original and watermarked audio signals, each producing one approximate and two details sub-bands.
2. Compute the watermarked vector W_t according to the following formula:

$$Wti = WD2_i - D2_i;\ i=\{1,120,240...L2\}, \text{and} \\ (M2+1)\le j\le((M1\times M2)+M2) \tag{9}$$

3. Divide W_t by the watermark amplification factor α to obtain the normalized watermark vector W_n.

$$Wn_i = Wti/\alpha;\quad (M2+1)\le i\le((M1\times M2)+M2) \tag{10}$$

4. Multiply W_n by 255 to reconstruct the original gray-scale image watermark.

$$W_i = Wn_i \text{ x } 255;\quad (M2+1)\le i\le((M1\times M2)+M2) \tag{11}$$

5. Convert W back into a two dimensional matrix which represents the gray- scale image watermark.

4. Simulation Results

In this research, we used instrumental and pop audio signals to measure and evaluate the performance of the proposed audio watermarking algorithm. A WAV file of 11 seconds in length was used for each type. Each file was sampled at 44.1 kHz and quantized to 16 bits per sample (CD quality). The files are of the stereo type, having left and right channels, and therefore the embedding procedure may embed a separate watermark into each channel. A 108 x 57 gray-scale image, shown in Figure 2, was used as a watermark for embedding in the audio files. The size of the image watermark was chosen after taking into consideration the length of the audio files.

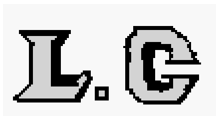


Figure 2. The original watermark image.

Performance of audio watermarking algorithms are commonly evaluated with respect to two common metrics: imperceptibility (inaudibility) and robustness [1]. Other metrics such as data payload and computation time, are less commonly used, and their importance varies from application to another. In what follows we present performance results of the proposed DWT-based algorithm.

4.1. Imperceptibility (Inaudibility)

Imperceptibility is related to the perceptual quality of the embedded watermark data within the original audio signal. It ensures that the quality of the signal is not perceivably distorted and the watermark is imperceptible to a listener. To measure imperceptibility, we use Signal-to-Noise Ration (SNR) as an objective measure, and a listening test as a subjective measure.

Signal to Noise Ratio (SNR) is a statistical difference metric which is used to measure the similitude between the undistorted original audio signal and the distorted watermarked audio signal. The SNR computation is done according to equation (12), where A corresponds to the original pop signal, and A' corresponds to the watermarked pop signal.

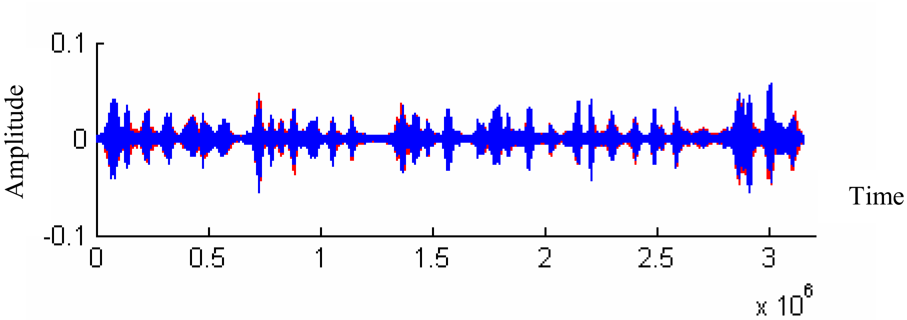
$$SNRL(dB) = 10\log_{10}\frac{\sum_n A_n^2}{\sum_n (A_n - A'_n)^2} \tag{12}$$

Although SNR is a simple way to measure the noise introduced by the embedded watermark and can give a general idea of imperceptibility, it does not take into account the specific characteristics of the human auditory system. Therefore, we also employed the Perceptual Audio Quality Measure (PAQM) [4]. Its has been shown in [6] that the correlation between PAQM and the mean opinion score (MOS) is 0.98. Therefore, in our experiments the PAQM scores will be mapped to the grading scale of MOS shown in Table 1.

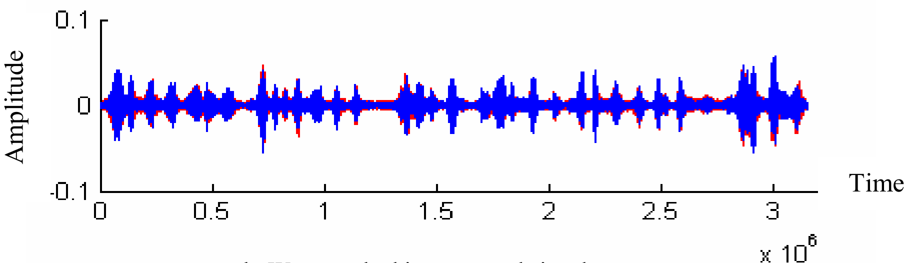
Table 1. MOS grading scale.

MOS Grade	Description
5	Imperceptible
4	Perceptible, but not annoying
3	Slightly annoying
2	Annoying
1	Very Annoying

Applying equation 12 on the original and watermarked instrumental and pop signals shown below in Figures 3 and 4, we obtained the SNR_{dB} values shown in Table 2. The high SNR_{dB} values reflect the obvious similarity between the waveforms in the Figures.

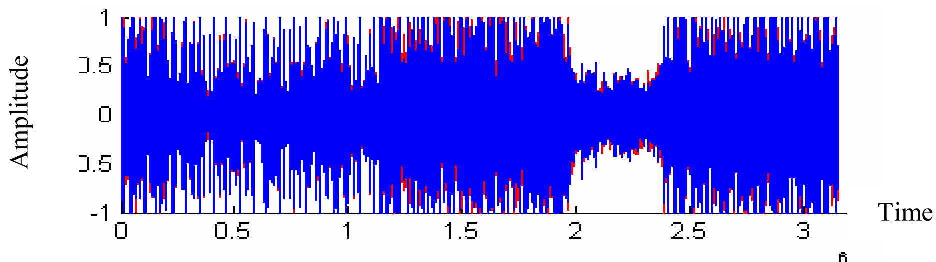


a. The original instrumental signal.

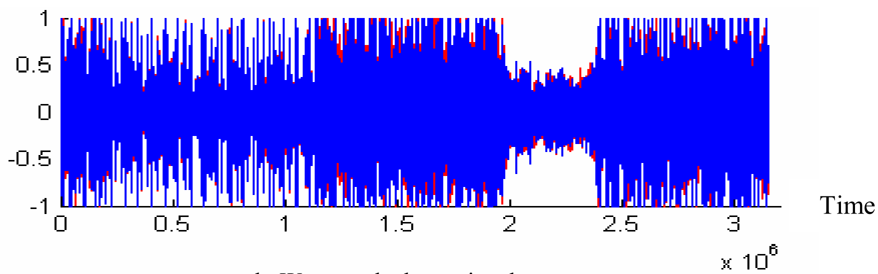


b. Watermarked instrumental signal.

Figure 3.



a. The original pop signal.



b. Watermarked pop signal.

Figure 4.

A listening test was actually performed with five listeners to estimate the subjective grade of the watermarked signals. Each listener was presented with the pairs of original signal and the watermarked signal and was asked to report whether any difference could be detected between the two signals. The five people listed to each pair for 10 times, and they gave a grade for this pair. The average grade for of each pair from all listeners is the final grade for this pair. Imperceptibility results (SNR and MOS grades) are summarized in Table 2.

Table 2. Imperceptibility (SNR and MOS) values.

Audio Type	SNR	MOS
Instrumental	28.5525	5.00
Pop	25.0314	5.00

4.2. Robustness

Watermarked audio digital signals may undergo common signal processing operations such as linear filtering, lossy compression, among many others. Although these operations may not affect the perceived quality of the host signal, they may corrupt the watermark image embedded within the signal. To evaluate robustness of the proposed algorithm, we implemented a set of attacks that commonly affect audio signals. Most of these attacks have been defined by Stirmark® watermarking benchmark [20]. Additional attacks were adopted form Adobe® Audition® 3 software which is a popular tool set for professional audio production [2]. The extracted watermarks after applying the attacks, one-at-a-time, are shown in Figures 5 & 6 for the instrumental and pop audio signals, respectively. The images in the figures indicate that the impact of the attacks on the watermarked signals vary, however, it was possible to extract the embedded watermarks after all attacks. Other than the subjective evaluation (inspection) of the extracted watermarks, we measured the similarity between the original watermark and the watermark extracted from the attacked watermarked images using the correlation factor ρ, which is computed as shown in equation 13 below.

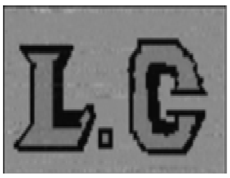
$$\rho(w,\hat{w}) = \frac{\sum_{i=1}^{N} w_i \hat{w}_i}{\sqrt{\sum_{i=1}^{N} w_i^2}\sqrt{\sum_{i=1}^{N} \hat{w}_i^2}} \tag{13}$$

where N is the number of pixels in watermark, w and $w^{\wedge}$ are the original and extracted watermarks respectively. The correlation factor $\boldsymbol{\rho}$ may take values between 0 (random relationship) to 1 (perfect linear relationship).

Add Brumm
($\boldsymbol{\rho}$ = 0.963)

Add Noise-20dB
($\boldsymbol{\rho}$ = 0.968)

Echo Attack
($\boldsymbol{\rho}$ = 0.966)



Amplify
($\boldsymbol{\rho}$ = 0.915)

Smooth
($\boldsymbol{\rho}$ = 0.939)

Stat
($\boldsymbol{\rho}$ = 0.950)

Exchange
($\boldsymbol{\rho}$ = 0.973)

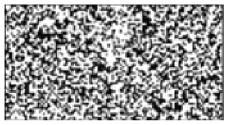
Extra Stereo
($\boldsymbol{\rho}$ = 0.971)

Zero Cross
($\boldsymbol{\rho}$ = 0.970)

Figure 5. Extracted watermarks after various StirMark attacks on the watermarked instrumental signal.

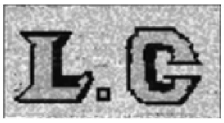
Add Brumm
(ρ = 0.944)

Add Noise-20dB
(ρ = 0.938)

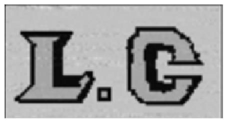


Echo Attack
(ρ = 0.06)

Amplify
(ρ = 0.865)



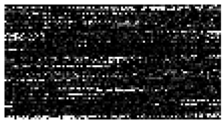
Smooth
(ρ = 0.910)



Stat
(ρ = 0.940)

Exchange
(ρ = 0.732)

Extra Stereo
(ρ = 0.870)



Zero Cross
(ρ = 0.020)

Figure 6. Extracted watermarks after various StirMark attacks on the watermarked pop signal.

5. Discussion

Many digital audio watermarking have been developed, and claims about their performances are made public. However, many of such algorithms are not evaluated with respect to imperceptibility (SNR, MOS), and robustness (StirMark Attacks, Correlation), as we have done in this paper. Nonetheless, and for the sake of completion, we present in Table the SNR and MOS results of with some traditional techniques as they were reported in [19]. Comparing the results with those in Table 2, we conclude that the proposed algorithm performs better than most traditional techniques. Moreover, our results fulfills optimal audio watermarking requirements set by the International Federation of the Phonographic Industry [13].

Table 3. Comparison with SNR and MOS values of traditional techniques.

Reference	Algorithm	SNR	MOS
Uludag [24]	DC-level Shifting	21.24	3.35
Bender [7]	Echo	21.47	3.60
Bender [7]	Phase	12.20	2.44
Bender [7]	LSB	67.91	4.90
Cox [8]	Spread Spectrum	28.59	4.46
Swanson [19]	Frequency Masking	12.87	2.93

6. Conclusions

Audio watermarking is an active research area that has been driven by the need to solve the copyright protection problem of digital audio products. Many promising audio watermarking techniques have been proposed and proved to be effective, however, and due to the challenging nature of audio signal processing, there remains much to do. In this paper, we proposed an effective audio signal watermarking Algorithm based on the discrete wavelets transform. The spectrum of the host audio signal was decomposed to locate the most appropriate regions to embed the watermark bits, imperceptibly and robustly. Indeed, our simulation results demonstrated the audibility and robustness of the proposed audio watermarking algorithm.

References

[1] Acevedo A., *Digital Watermarking for Audio Data in Techniques and Applications of Digital Watermarking and Content Protection*, Artech House, USA, 2003.

[2] ADOBE® Audition, http://www.adobe.com/products/audition/, 2009.

[3] Arnold M., "Audio Watermarking: Features, Applications and algorithms," *in Processings of the IEEE International Conference on Multimedia and Expo*, pp. 1013-1016, 2000.

[4] Arnold M., Wolthusen S., and Schmucker M., "Techniques and Applications of Digital Watermarking and Content Protection," *Artech*

House, Psychoacoustics: Facts and models, Springer-Verlag, 2003.

[5] Bassia P. and Pitas I., "Robust Audio Watermarking in the Time Domain," *IEEE Transactions on Multimedia*, vol. 3, no. 2, pp. 232-241, 2001.

[6] Beerends J. and Stemerdink J., "A Perceptual Audio Quality Measurement Based on a Psychoacoustic Sound Representation," *Journal of the Audio Engineering Society*, vol. 40, no. 12, pp. 963-978, 1992.

[7] Bender W., Gruhl D., Morimoto N., and Lu A., "Techniques for Data Hiding," *IBM System Journal*, vol. 35, pp. 313-336, 1996.

[8] Cox I., Kilian J., Leighton T., and Shamoon T., "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673-1687. 1997.

[9] Cox I., Miller M., and Bloom J., "Digital Watermarking," Academic Pressing, USA, 2002.

[10] Cvejic N. and Seppänen T., "Increasing the Capacity of LSB-Based Audio Steganography," *in Proceedings of the IEEE International Workshop on Multimedia Signal Processing*, pp. 336-338, 2002.

[11] Erelebi E. and Bataki L., "Audio Watermarking Scheme Based on Embedding Strategy in Low Frequency Components with a Binary Image," *Digital Signal Processing*, vol. 19, no. 2, pp. 265-277, 2009.

[12] Gruhl D., Lu A., and Bender W., "Echo Hiding," *in Proceedings of the International Conference on Info Hiding*, pp. 295-315, 1996.

[13] IFPI (International Federation of the Phonographic Industry), http://www,ifpi.org, 2009.

[14] Katzenbeisser S. and Petitcolsd F., "Information Hiding Techniques for Steganography and Digital Watermarking," Artech House, Boston, USA, 2000.

[15] Kim H. and Choi Y., "A Novel Echo-hiding Scheme with backward and Forward Kernels," IEEE Transactions on Circuits and Systems for Video Technology; vol. 13, no. 8, pp. 885-889, 2003.

[16] Malik H., Ansari R., and Khokhar A., Robust "Audio Watermarking Using Frequency-Selective Spread Spectrum," *IET Information Security*, vol. 2, no. 4, pp. 129-150, 2008.

[17] Mallat S., "A Theory for Multi-Resolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674-693, 1989.

[18] Ozer H., Sankur B., and Memon N., "An SVD-Based Audio Watermarking Technique," *in Proceedings of the Multimedia and Security Workshop*, New York, pp. 51-56, 2005.

[19] Sehirli M., Gurgen F., and Ikizoglu S., "Performance Evaluation of Digital Audio Watermarking Techniques Designed in Time, Frequency and Cepstrum Domains," *in Proceedings of the International Conference on Advances in Information Systems*, pp. 430-440, 2005.

[20] Steinebach M., Petitcolas F., Raynal F., Dittmann J., Fontaine *et al*, "Stirmark Benchmark: Audio Watermarking Attacks," *in Proceedings of the International Conference on Information Technology: Coding and Computing*, pp. 49-54, 2001.

[21] Swanson M., Zhu B., and Tewfik A., Boney L. "Robust Audio Watermarking Using Perceptual Masking," *Signal Processing*, vol. 66, no. 3, pp. 337-355, 1998.

[22] Tilki J. and Beex A., "Encoding a Hidden Digital Signature Using Psychoacoustic Masking," *in Proceedings of the 7th International Conference on Signal Processing Applications and Technology*, pp. 476-480, 1996.

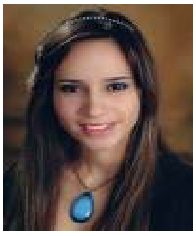
[23] Tilki J. and Beex A., "Encoding a Hidden Auxiliary Channel onto Digital Audio Signal Using Psychoacoustic Masking," *in Proceedings of IEEE Southeastern*, pp. 331-333, 1997.

[24] Uludag U. and Arslan L., Audio Watermarking Using DC-level Shifting, *Project Report*, Bogazici University, Available at http://busim.ee.boun.edu.tr/~speech/publications/audio_watermarking/uu_la_audio_wm2001.pdf, 2001.

[25] Yeo I. and Kim H., "Modified Patchwork Algorithm: A Novel Audio Watermarking scheme," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 4, pp. 381-386, 2003.

Ali Al-Haj received his first university degree in electrical engineering from Yarmouk University, Jordan, in 1985, the MSc degree in electronics engineering from Tottori University, Japan, in 1988 and the PhD degree in computer engineering from Osaka University, Japan, in 1993. He then worked as a research associate at ATR Advanced Telecommunications Research Laboratories in Kyoto, Japan, until 1995. He joined Princess Sumaya University, Jordan, in October 1995, where he is now an associate professor. Al-Haj has published papers in dataflow computing, information retrieval, VLSI digital signal processing, neural networks, and multimedia watermarking.

Ahmad Mohammad received his BSc in electrical engineering from Ein-Shams University, Egypt in 1981, and the MSc and PhD degrees in electrical engineering, from Akron University, USA, in 1989, and 1992, respectively. He has been an assistant professor at the Department of Computer Engineering, Princess Sumaya University, Jordan, since 2000. His research interests include control systems, image processing, and information security.



Lama Bata received her BSc in computer engineering from the University of Jordan, in 2007. Currently, she is a Web developer at Patterns Inc., Jordan.