Robust Audio Data Hiding Using Correlated Quantization With Histogram-Based Detector

Mohammad A. Akhaee, *Student Member, IEEE*, Mohammad J. Saberian, Soheil Feizi, and Farokh Marvasti, *Senior Member, IEEE*

***Abstract*—In this paper, two blind audio watermarking methods using correlated quantization for data embedding with histogram-based detector have been proposed. First, a novel mapping called the point-to-point graph (PPG) is introduced. In this mapping, the value of samples is important as well as the correlation among them. As this mapping increases the dimension of the signal, the data embedding procedure (quantization) will be diversified more securely than that of the 1-D domains such as the time or frequency domains. Hence, two watermarking techniques coined as hard and soft quantization methods based on the quantization of the PPG point radii are suggested. The performance of both techniques is analyzed by obtaining the radii distribution of PPG points after watermarking. Experimental results against AWGN attack confirm the validity of theoretical analysis. Moreover, the robustness of the proposed methods against other common attacks such as echo, low pass, resampling, and MP3 are investigated through extensive simulations.**

***Index Terms*—Audio watermarking, hard quantization, point-to-point graph (PPG), soft quantization.**

I. Introduction

DATA hiding has various applications such as authentication, fingerprinting, copyright protection, and broadcasting monitoring [1]–[6]. It is necessary that the watermark data are readily extracted from the watermarked signal in order to represent the copyright owner. Three important and basic issues that should be considered in watermarking systems are: 1) perceptual transparency, 2) robustness against various attacks, and 3) data rate of the watermark. There are trade-offs between the capacity of the watermarked data and robustness against watermark attacks, while keeping the perceptual quality of the watermarked audio signal at acceptable levels. The trade-offs among the embedding rate, distortion, and robustness have been examined from an information-theoretic perspective [7], [8]. In [7], a model of the watermarking scheme is introduced and upper and lower bounds on the data embedding capacity are determined. In [8], Cohen and Lapidoth investigated information rates for Gaussian host signals and the squared-error distortion measure.

Manuscript received November 20, 2007; revised September 13, 2008. First published April 28, 2009; current version published July 17, 2009. This work was supported by Iran Telecommunication Research Center (ITRC). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Wen Gao.

The authors are with the Advanced Communication Research Institute (ACRI), Department of Electrical Engineering, Sharif University of Technology, Tehran 1458889694, Iran (e-mail: akhaee@ee.sharif.edu; mjsaberian@ee.sharif.edu; sfeiz@ee.sharif.edu; marvasti@sharif.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2009.2012923

In comparison with digital image watermarking, less attention has been paid to audio watermarking because it is more challenging. Audio data hiding schemes exploits human auditory system (HAS) including spectral and temporal masking, inaudibility of phase distortion for making the hidden message transparent. Least significant bit (LSB) [9], echo hiding [10], and phase coding approaches [11] are some examples of these perception based audio watermarking techniques. The main drawback of these algorithms is their low embedding capacities.

In this framework, the quantization index modulation (QIM) method is an important blind watermarking class performing close to the system capacity [13], [14]. In this technique, data embedding is attained by quantizing the host feature stream with a quantizer that is selected among a set of quantizers which are associated with different messages. For example, the selection of a scalar quantizer results in a method named dither modulation (DM) with bit repetition or a distortion compensated version of DM (DC-DM) [13]–[15]. The main problem of this approach is the development of quantizer codebooks which are used for data embedding. This problem is solved by using lattice-based quantizers, which are efficient in data embedding and decoding [16], [17]. Recently, it has been shown that the channel capacity may be achievable using lattice codebooks [18].

Although QIM method is commonly applied to random Gaussian signals for information theoretic discussions, it can be employed as a powerful approach in audio watermarking. Based on QIM technique, an efficient self-synchronization scheme is proposed for audio data transmission by Wu *et al.* [19]. In this case, the imperceptibility of the algorithm should be considered. In [19], the imposed distortion is related to signal-to-noise ratio (SNR) which does not seem to be an appropriate idea because the mean opinion score (MOS) in audio signals do not necessarily match with SNR.

Here, we propose two blind techniques based on the quantization in a novel transform domain. This transformation, called PPG mapping, increases the dimension of the signal and utilizes the correlation among the samples as well as their magnitudes. In the first proposed method, hard quantization, code-words are selected as two circles in a PPG domain, while for more robustness a guard margin is inserted around these quantization levels. In the second method, soft quantization method code-words are specific rings with inherent dither, which are periodically repeated in all regions in the PPG domain. In fact, these methods are based on forming different patterns in the radii of points according to the embedded bit. At the decoder, instead of the QIM method, a histogram based de-watermarking procedure has been used for both proposed techniques. In fact, the number of PPG points around "zero" or "one" quantization levels is

counted and the detection is performed by comparing the histogram around these levels. Simulation results and analytical discussion demonstrate that these kinds of code-words and decoders yield better performance in comparison with some recent improved version of the QIM methods reported so far.

The rest of the paper is organized as follows: In Section II, major characteristics of the PPG mapping and the quantization in this domain are introduced. Two novel methods, hard and soft quantization, are presented in Section III; while error probabilities of these methods are calculated in Section IV. The robustness of the suggested techniques against various common attacks is evaluated by simulations in Section V. Finally, Section VI concludes the paper.

II. Point-to-Point Graph (PPG) Mapping

This mapping is based on putting N samples of the signal together in a specific order to convert a 1-D signal to an n-dimensional counterpart. For example, for the case of $n = 2$, this mapping is such that sample one and sample $k+1$, sample two and sample $k+2, \ldots$, sample k and sample $2k$ are taken as pairs. This process is performed on the following sets of $2k$ samples as well, where k is the index of the PPG mapping. Using this method, a 1-D signal can be converted to a 2-D one. It is worthwhile to mention that the index, k, of the PPG mapping can be chosen arbitrarily depending on the desired characteristics. For example, in the case of small k's, because of the strong correlation between samples, two components of each PPG point is close to each other and thus, the PPG points form an elliptical curve about the inclination line with 45. This distribution along this line is related by its inherent low-pass property. On the other hand, choosing larger values of k results in more dispersion in the PPG points due to decreased correlation among the samples. In an extreme case, when k tends to infinity, two components of each point will be uncorrelated and therefore, they form a rough circle instead of an ellipse.

For convenience, we will use the notation $PPG\ (n, k)$, in which n is the dimension of the domain and k is the PPG mapping index. Furthermore, in this paper we only concentrate on the 2-D case of $n = 2$ where by convention $PPG(k) = PPG(2, k)$. In the following, the basic concepts of data embedding in this domain will be explained.

A. Quantization in the PPG Mapping Domain

Quantization in the PPG mapping domain can be performed on both radius and angle. However, in this paper, in order to have better performance, we just quantize the radii of PPG points while keeping their angles intact. Quantization levels are concentric equidistant circles. In order to quantize the points, each point is scaled in the direction of the line which connects it to the origin, either upward or downward depending on the embedded bit.

Quantization of the radii of PPG points has three major advantages. 1) More transparency in the time domain, 2) more robustness against lowpass attacks such as filtering and resampling, and 3) higher resistance in noisy environments. These three advantages are further discussed in the following.

1) Quantization in the PPG domain produces correlated noise on the components of each point. In other words, it creates specific patterns in the distribution of PPG points which can be utilized in the de-watermarking procedure. By adjusting the PPG mapping index, the time domain traces of quantization can be controlled. For example, as illustrated in Fig. 1, for $k = 1$, adjacent samples are quantized similarly and therefore has a great impact on the time domain shape, whereas, for $k = 10$, this effect is absolutely transparent in the time domain, guaranteeing the security of the watermarking process.
2) Radii quantization imposes lowpass distortion on the watermarked signal. In other words, the hidden data are inserted in the low-frequency components of the host signal without producing disturbing high frequency components. As an extreme example, when the host signal is multiplied by a real factor, its bandwidth does not change. In such a case, just the radii of PPG points are multiplied by that factor and their angles remain unchanged. Therefore, because of the low-frequency characteristic of the hidden data, it is robust against lowpass attacks.
3) In order to describe better robustness against AWGN, we compare quantization in the 1-D and 2-D domains when the radius is just quantized. To have a fair comparison, the bit rate and distortion in both cases must be equal. Thus, we duplicate the same watermark data bit in the 1-D case and the quantization distance Δ is kept the same for both cases. Since the watermark distortion should not be noticeable, the value of Δ is kept small; this results in uniform distribution for the PPG points. Therefore, the overall distortion (variance) of both quantizations will be $\Delta^2/12$. Fig. 2(a) demonstrates the decision making region for quantization in the 1-D case while Fig. 2(b) shows the decision making region in the 2-D domain for the radius only quantization.

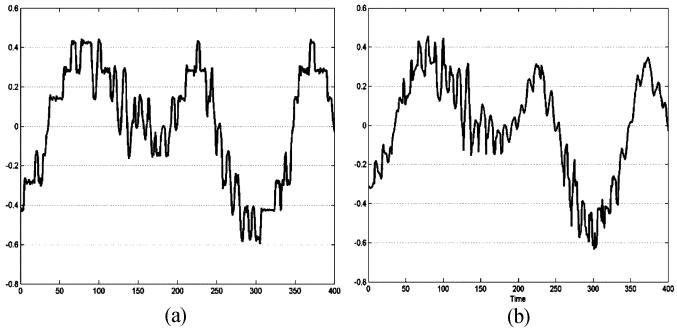


Fig. 1. Time domain shape of a signal with 400 samples for (a) $k = 1$ and (b) $k = 10$.

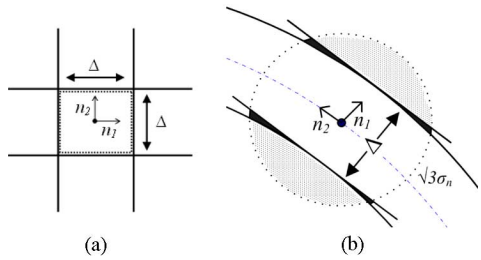


Fig. 2. Decision making region of two samples when (a) they are separately quantized and (b) the PPG radius is quantized. The circle is the effect of noise with a variance of $3\sigma^2$.

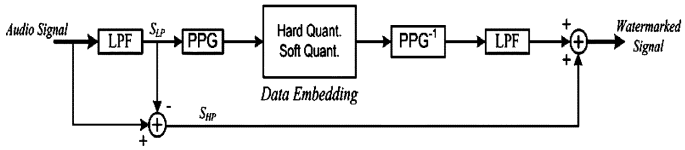


Fig. 3. Block diagram of the watermark embedding using the PPG mapping.

In the presence of noise, according to Fig. 2, error occurs when a typical point stands outside its corresponding square region as in Fig. 2(a), or outside its corresponding ring as in Fig. 2(b).

Since the noise is assumed to be Gaussian with zero mean and variance σ^2 and it is independent in both directions, according to [20], the probability of error, for the 1-D case, can be written as

$$P_{e\_1-D} = 1 - \left(1 - Q\left(\frac{\Delta}{2\sigma}\right)\right)^2 \tag{1}$$

where $Q(\alpha) = 1/2\pi \int_{\alpha}^{\infty} e^{-u^2/2} du$ is the error function.

In the 2-D case, the problem is more sophisticated than the 1-D case because the tangential noise component, n_2, affects the error less than the one in radial direction especially when the quantization levels have large distances from the origin. In Fig. 2(b), the effect of noise with a variance of $3\sigma^2$ is depicted with a circle. With a high probability, the error region is inside the noise circle and outside the ring. These regions are shown with dots. The tangents of the quantization levels are also shown. We assume black filling regions have the same area and this is the only approximation we have used. As a consequence, the region of interest that produces correct decisions is between these two parallel tangents. In this case, the tangential noise component, n_2, does not cause error and the probability of error, similar to [20], can be written as

$$P_{e\_2-D} = 1 - \left(1 - Q\left(\frac{\Delta}{2\sigma}\right)\right). \tag{2}$$

As we can see from (1) and (2), the probability of error is smaller in the 2-D case because the tangential noise component has been eliminated.

III. Proposed Methods

In this section, we introduce the idea of blind quantization based watermarking method in the PPG transform domain. Then, by using this idea in two special cases, two novel watermarking algorithms are suggested. First of all, in order to have more transparency, just low-pass components of the input audio signal are extracted and watermarked while its high-pass components are kept intact. To remove the high-pass components of the quantization process, we filter the quantized signal with a similar low-pass filter and then add the high-pass components of the original signal to obtain the watermarked audio signal. The block diagram of the data embedding algorithm is demonstrated in Fig. 3. In the de-watermarking process, first the watermarked audio is passed through the low-pass filter and then the data extracting algorithm is performed. This procedure makes the watermarking algorithm more robust against lowpass attacks including resampling, MP3, and filtering.

A. Blind Quantization Watermarking Approach

Here, the problem of blind watermarking techniques based on quantization is investigated. A typical watermarking method can be expressed in the following manner.

Assume $S = (s_1, s_2, \ldots, s_n)$, $s_i \in X$ and $Q = \{q_1, q_2, \ldots, q_M\}$ to be the input signal and its quantization levels, respectively. Also $m = 0, 1$ is the message bit. Define functions $L_m(s) : X \to Q$ which map a specific sample s to a quantization q level according to the message bit. One conventional example of $L(s)$ assigns the nearest quantization level to the input sample. Moreover, consider functions $f_m(s, q) : X \times Q \to X$, which can be applied to the input signal during the watermarking procedure.

In the quantization based watermarking process, for a typical sample s_i, the watermarked sample s_i' is achieved in the following manner:

$$P_m(s_i) = s_i' = f_m(s_i, L_m(s_i)) \tag{3}$$

where P_m represents the process of embedding bit m.

Since the process should be blind and inaudible, we look for the functions f_m and L_m to satisfy the following requirements.

1) In the absence of the original signal, the algorithm should be able to recognize whether the processes P_0 or P_1 is applied to the host signal S. In fact, the main idea is to design monotone processes (P_0 and P_1) in such a way that their corresponding ranges are separated.
2) Besides, the distortion imposed on the host signal should be controllable within an acceptable range.

Any set of functions that satisfies these requirements is a solution to this problem. Here, we present a suitable set of functions $f(.)$ and $L(.)$ to handle this problem as follows:

$L_0(s)$ largest quantization level smaller than s;

$L_1(s)$ smallest quantization level greater than s;

$$f_m(s_i, L_m(s_i)) = \frac{s_i + dL_m(s_i)}{d+1} \quad m = 0, 1 \tag{4}$$

where d is a design parameter. Since f_0 and f_1 are convex functions, it can be easily deduced that

$$L_0(s) \le f_0(s, L_0(s)) \le \frac{L_1(s) + dL_0(s)}{d+1}$$
$$\frac{L_0(s) + dL_1(s)}{d+1} \le f_1(s, L_1(s)) \le L_1(s).$$

Thus, the range of the process P_0 is $[L_0(s), L_1(s) + dL_0(s)/d + 1]$ and the range of P_1 is $[L_0(s) + dL_1(s)/d + 1, L_1(s)]$. As discussed earlier, the watermarking algorithm is blind if these intervals do not overlap. Consequently, we should have

$$\frac{L_1(s) + dL_0(s)}{d+1} \le \frac{L_0(s) + dL_1(s)}{d+1} \Rightarrow d \ge 1. \tag{5}$$

B. Hard Quantization Method

As an example of (5), consider there are just two levels for quantization and d tends to infinity. Thus, the points between these two levels q_1 and q_2 are quantized either to q_1 or q_2; this is called the hard quantization method. In practical implementations, it is better to put a guard interval near these two quantization levels for more robustness. Thus, for data embedding, four parameters g_1, q_1, q_2, and g_2 are selected between zero and $\sqrt{2}$ (for the case in which both samples have the maximum value of one) depending on the input signal characteristics (such as power) and assuming $g_1 < q_1 < q_2 < g_2$. The values of these parameters are set between the encoder and decoder. Then, the PPG points are quantized by the following procedure.

Step 1) The points with radii less than g_1 are remained unchanged in order to keep the transparency of the watermarking method.

Step 2) The points with radii between g_1 and q_1 are mapped to a circle with radius g_1 to increase the robustness of the method against noise by developing a guard band for q_1.

Step 3) The points between the levels q_1 and q_2 are mapped to level q_1 or q_2 depending on the embedded bit.

Step 4) The points between q_2 and g_2 levels are mapped to g_2 level to produce a guard band for q_2.

Step 5) The points greater than g_2 are kept unchanged.

In the de-watermarking procedure, decision making is based on the number of points close to the levels q_1 and q_2. In order to decrease the noise effects and obtain more robustness against attacks, the points with radii between g_1 and g_2 are prequantized to the nearest level. Then, by comparing the number of points quantized to levels q_1 and q_2, the embedded data can be extracted.

The steps of de-watermarking algorithm are explained completely as follows.

Step 1) The dc component of the signal is eliminated.

Step 2) The PPG points of each frame are obtained.

Step 3) Prequantization process is performed on the points between levels g_1 and g_2.

Step 4) By comparing the number of points quantized to the levels q_1 and q_2, the embedded data can be obtained.

It should be mentioned that by adjusting the quantization levels, the robustness of the method and its transparency can be controlled.

C. Soft Quantization Method

This method is obtained by (5) considering N quantization levels and $d = 1$. Thus, assume s is a typical PPG point whose radius R is located between two quantization levels q_m and q_{m+1}, and $q_i, 1 < i < N$, are equidistant levels. To embed bits "0" or "1", the point s' is mapped to the point with radius $R_0 = (R + q_m)/2$ or $R_1 = (R + q_{m+1})/2$, respectively. Since only the radius is quantized, the ranges of zero and one embedding processes are two different adjacent rings which are called *zero rings* and *unity rings*, respectively.

The watermarking procedure is described as follows.

Step 1) The dc component of the signal is eliminated.

Step 2) The PPG mapping with index k is applied in each frame.

Step 3) To embed bit "0", a typical point with radius R between levels q_m and q_{m+1} is mapped to the point R_0 and to embed bit "1", it is mapped to the point R_1.

Step 4) The inverse of the PPG mapping is applied to the resultant points.

In the de-watermarking process, the number of PPG points in the zero rings is compared with that of the unity rings. It is worthwhile to mention that in addition to less complexity and more robustness, this approach creates less distortion in the host signal and improves its transparency.

D. Synchronization Approach

Since the quantization is performed in the PPG transform domain, the number of samples and their relations are important for data extraction. Generally, we can use other synchronization techniques such as [21] and [22] to make the proposed techniques robust against de-synchronization attacks. In this framework, we should decrease the payload and instead insert the synchronization code in each frame [19]. Here, we propose a method adapted to our data embedding algorithm which can be used for finding the beginning of the frames. To this aim, suppose the received signal is shifted by k_0 samples. Thus, the frame starting point is lost and the framing process should be modified.

As we know, each frame contains the message bit "0" or "1". Therefore, according to the hidden bit, concentration of PPG points near some specific quantization levels are more than the others. Thus, for a frame in which the bit "0" is hidden, the number of points near the zero rings, N_0, is by far higher than the number of points close to the unity rings, N_1. If we denote $\rho_0 = N_0/N_1$ and $\rho_1 = N_1/N_0$, the values of ρ_0 and ρ_1 will be maximized in accordance to the message bit. Hence, we slide a series of frames in a range of samples periodically, and then compute the value of ρ_0 and ρ_1 for each slide number in each frame. Consequently, each frame will introduce a slide value in which ρ_0 or ρ_1 is maximized. The desired slide number is the one presented by most of the frames. It must be noted that this synchronization approach suffers from its high complexity cost; however it can be reduced using wavelet transform and its properties in shifting [19], which is not in the scope of this paper.

IV. Performance Analysis

Consider a typical PPG point, (x, y). With AWGN on each component (n_1, n_2), with zero mean and variance σ_n^2, the PPG point will change to $(x+n_1, y+n_2)$. Therefore, the distribution of its radius $A = \sqrt{(x+n_1)^2 + (y+n_2)^2}$ is Rician, i.e.,

$$f_A(A, R) = \frac{A}{{\sigma_n}^2} e^{-(A^2+R^2)/2{\sigma_n}^2} I_0\left(\frac{AR}{{\sigma_n}^2}\right) U(A) \quad (6)$$

where $R = \sqrt{x^2 + y^2}$ and I_0 is the modified Bessel function of the first kind.

In order to analyze the error probability of both methods, the distribution of PPG point radius should be evaluated. As shown in Fig. 4, the distribution is Rayleigh with parameter σ that depends on both the signal energy and the PPG mapping index k:

$$f_R(R; k, p_{in}) = \frac{R}{\sigma^2} e^{-R^2/2\sigma^2} U(R). \quad (7)$$

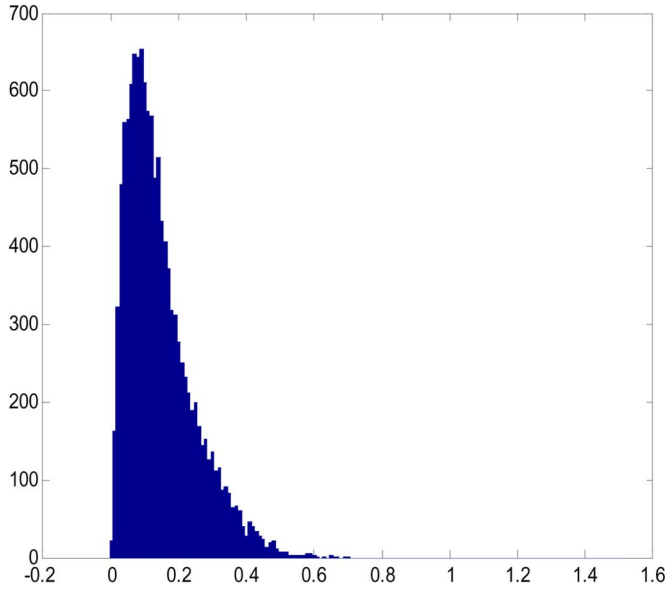


Fig. 4. Rayleigh distribution of the radius of PPG points.

To derive the Rayleigh distribution in (7), the parameter σ should be calculated with respect to k and the input power P_{in}. To achieve this aim for a specific frame, we fix the input power, and then fit a curve to the PPG point radius histogram for different values of k using MATLAB software. Fig. 5(a) illustrates the Rayleigh parameter σ versus the PPG mapping index. According to this figure, there is an approximately linear relationship between these two parameters which can be written as

$$\sigma^2 = -0.0026k + 0.1895. \tag{8}$$

Similarly, in order to find the relation between the input power and the parameter σ, for a specific k, we fit a curve to the PPG point radius histogram. By using different frames leading to various input powers, according to Fig. 5(b), we can see there is also a nearly linear relationship between these parameters. This relationship can be written as follows:

$$\sigma^2 = 2.4771P_{in} + 0.0025. \tag{9}$$

Merging (3) and (4), a general relation between these three parameters can be derived as

$$\sigma^2 = 2.47P_{in} - 0.397k - .295P_{in}k + 0.409. \tag{10}$$

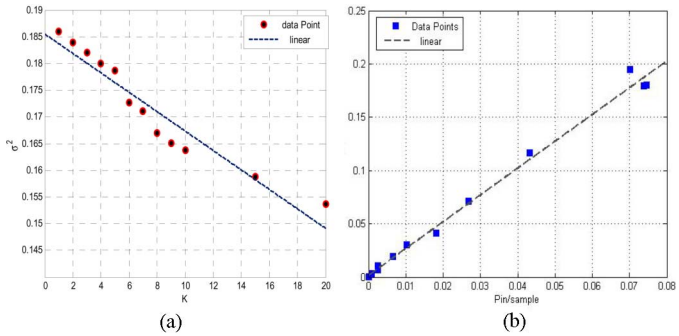
By considering (10), the relation (7) can be rewritten as

$$f_R(R;k,p_{in}) = \frac{R}{(2.47P_{in} - 0.397k - .295P_{in}k + 0.409)} \times e^{-R^2/2(2.47P_{in}-0.397k-.295P_{in}k+0.409)}U(R). \tag{11}$$

In the following, we discuss each method separately.

A. Error Probability of Hard Quantization

The decision making in the hard quantization method is based on the number of points close to the levels q_1 and q_2. However, considering the prequantization technique explained before, it leads to comparing the number of points within $[(g_1 + q_1)/2, (q_1 + q_2)/2]$ and $[(q_1 + q_2)/2, (q_2 + g_2)/2]$ which were denoted by N_0 and N_1, respectively.



(a) (b)

Fig. 5. Rayleigh distribution of PPG points radii versus. (a) PPG mapping index k. (b) Input power P_{in}.

The probability that a typical point with radius R is dropped, because of the noise, into $[(g_1 + q_1)/2, (q_1 + q_2)/2]$ is

$$\begin{aligned} P_{fst-reg}(R) &= \int_{(g_1+q_1)/2}^{(q_1+q_2)/2} f_A(A,R)dA \\ &= \int_{(g_1+q_1)/2}^{(q_1+q_2)/2} \frac{A}{{\sigma_n}^2} e^{-(A^2+R^2)/2{\sigma_n}^2}] \\ &\quad \times I_0\left(\frac{AR}{{\sigma_n}^2}\right)dA \end{aligned} \tag{12}$$

where $f(A,R)$ is the Rician distribution given in (6) and R is the radius of a typical PPG point. Then, the average number of points that fall into the first region can be derived as follows:

$$P_f = E\{P_{fst-reg}(R)\} = \int_0^\infty P_{fst-reg}(R) f_R(R;k,P_{in})dR$$

where $f_R(R,k,P_{in})$ is the radius distribution of a typical point. Similarly, for the second region

$$\begin{aligned} P_{snd-reg}(R) &= \int_{(q_1+q_2)/2}^{(q_2+g_2)/2} f_A(A,R)dA \\ &= \int_{(q_1+q_2)/2}^{(q_2+g_2)/2} \frac{A}{{\sigma_n}^2} e^{-(A^2+R^2)/2{\sigma_n}^2} I_0\left(\frac{AR}{{\sigma_n}^2}\right)dA \\ P_s &= E\{P_{snd-reg}(R)\} \\ &= \int_0^\infty P_{snd-reg}(R) f_R(R;k,P_{in})dR. \end{aligned} \tag{13}$$

It is worth mentioning that after quantization of the points, according to Fig. 6, the distribution of R changes to $f(R;k,P_{in}|m = 1)$ and $f(R;k,P_{in}|m = 0)$ depending on the embedded bit.

Errors occur in two cases: the first case is when the message bit $m = 1$ is embedded into the signal and because of the noise, the number of points that fall into the first region is more than that of the second region. Therefore, at the receiver, the bit "0" is detected.

In this case, the probabilities that a point lays in the first region $f_R(R;k,P_{in}|m = 0)$ or in the second one

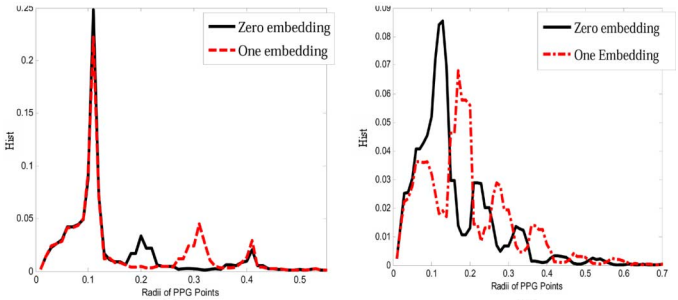


Fig. 6. Histogram of PPG points radius for "0" or "1" message bit embedding. a) Hard quantization and b) soft quantization.

$f_R(R;k,P_{in}\,|m=1)$ when the bit "1" is embedded can be written as

$$\begin{aligned} P_{F|1} &= E\{P_{fst-reg}(R)\} \\ &= \int_0^{\infty} P_{fst-reg}(R) f_R(R;k,P_{in}|m=1)dR \end{aligned} \tag{14}$$

$$\begin{aligned} P_{S|1} &= E\{P_{snd-reg}(R)\} \\ &= \int_0^{\infty} P_{snd-reg}(R) f_R(R;k,P_{in}|m=1)dR \end{aligned} \tag{15}$$

$$P_{others|1} = 1 - P_{F|1} - P_{S|1} \tag{16}$$

where $P_{others|1}$ is the probability that a point lays outside the determined first and second regions for the embeded message "1". Now, we define the parameters i_f, i_s, and i_{others} that represent the number of points located in the first, second and other regions, respectively. When bit "1" is embedded, since $i_f > i_s$, we have

$$\begin{aligned} P_{e|m=1} = & \sum_{i_{others}=0}^{N} \binom{N}{i_{others}, i_f, i_s} \\ & \times \left(P_{others|1}\right)^{i_{others}} \left(P_{F|1}\right)^{i_f} \left(P_{S|1}\right)^{i_s} \\ \text{subject to}: & \begin{cases} i_{others} + i_f + i_s = N \\ i_f > i_s. \end{cases} \end{aligned} \tag{17}$$

Similarly, the error probability for "0" embedding case can be written as

$$\begin{aligned} P_{e|m=0} = & \sum_{i_{others}=0}^{N} \binom{N}{i_{others}, i_f, i_s} \\ & \times \left(P_{others|0}\right)^{i_{others}} \left(P_{F|0}\right)^{i_f} \left(P_{S|0}\right)^{i_s} \\ \text{subject to}: & \begin{cases} i_{others} + i_f + i_s = N \\ i_f < i_s \end{cases} \end{aligned} \tag{18}$$

where $P_{F|0}$ and $P_{S|0}$ are similarly calculated using $f_R(R;k,P_{in}\,|m=0)$ instead of $f_R(R;k,P_{in}\,|m=1)$. Thus, the total error probability can be derived as follows:

$$P_e = \frac{1}{2}(P_{e|m=0} + P_{e|m=1}). \tag{19}$$

B. Error Probability of Soft Quantization Method

In the soft quantization method, if q_i and q_{i+1} are two adjacent quantization levels, the number of points within $[q_i, (q_i + q_{i+1})/2]$, N_0, and $[(q_i + q_{i+1})/2, q_{i+1}]$, N_1, for all valid i's, are important at the decoder. In other words, if $N_0 > N_1$, the embedded message bit is assumed to be "0", whereas, if $N_1 > N_0$, it is assumed to be "1". Similar to the hard quantization method, the N_0 and N_1 can be derived as follows, if P_{Zero_Ring} is the probability that a typical point with radius R in the PPG domain is located in a zero ring and P_{Unit_Ring} is its counterpart when the bit "0" is embedded, then

$$\begin{aligned} P_{Zero\_Ring}(R) &= \sum_{k=1}^{M-1} \int_{q_k}^{q_k+\gamma} f_A(A,R)dA \\ &= \sum_{k=1}^{M-1} \int_{q_k}^{q_k+\gamma} \frac{A}{{\sigma_n}^2} e^{-(A^2+R^2)/2{\sigma_n}^2} \\ &\quad \times I_0\left(\frac{AR}{{\sigma_n}^2}\right) dA \end{aligned} \tag{20}$$

$$\begin{aligned} P_{Unity\_Ring}(R) &= \sum_{k=2}^{M} \int_{q_k-\gamma}^{q_k} f_A(A,R)dA \\ &= \sum_{k=2}^{M} \int_{q_k-\gamma}^{q_k} \frac{A}{{\sigma_n}^2} e^{-(A^2+R^2)/2{\sigma_n}^2} \\ &\quad \times I_0\left(\frac{AR}{{\sigma_n}^2}\right) dA \end{aligned} \tag{21}$$

where $\gamma = (q_1 - q_2)/2$. Thus, the average probability that a point lies in either zero or unity rings given the embedded message "1" can be calculated as follows:

$$\begin{aligned} P_{Z|1} &= E\{P_{Zero\_Ring}\} \\ &= \int_0^{\infty} P_{Zero\_Ring} f_R(R,k,p_{in}\,|m=1)dR \end{aligned} \tag{22}$$

$$\begin{aligned} P_{U|1} &= E\{P_{Unity\_Ring}\} \\ &= \int_0^{\infty} P_{Unity\_Ring} f_R(R,k,P_{in}\,|m=1)dR \end{aligned} \tag{23}$$

$$P_{others|1} = 1 - P_{Z|1} - P_{U|1} \tag{24}$$

where $f_R(R;k,P_{in}|m=1)$ is the distribution of PPG point radius after embedding "1" and $P_{others|1}$ represents the probability that the points are located in either the first quantization circle (no region change) or outside the last quantization circle due to noise. Now, similar to the hard quantization method, we define the variables i_Z, i_U, and i_{others} as the number of points that are in zero rings, unity rings, and the other unchanged regions, respectively. Error occurs when after embedding "1", "0" is detected at the receiver, thus

$$\begin{aligned} P_{e|m=1} = & \sum_{i_{others}=0}^{N} \binom{N}{i_{others}, i_Z, i_U} \\ & \times \left(P_{others|1}\right)^{i_{others}} \left(P_{Z|1}\right)^{i_Z} \left(P_{U|1}\right)^{i_U} \\ \text{subject to}: & \begin{cases} i_{others} + i_Z + i_U = N \\ i_Z > i_U. \end{cases} \end{aligned} \tag{25}$$

In fact, error happens when the number of points in zero rings i_Z, is greater than i_U when message "1" is inserted. Similarly, the probability of error when the bit "0" is embedded can be written as

$$P_{e|m=0} = \sum_{i_{others}=0}^{N} \binom{N}{i_{others}, i_Z, i_U} \times (P_{others|0})^{i_{others}} (P_{Z|0})^{i_Z} (P_{U|0})^{i_U}$$
$$\text{subject to}: \begin{cases} i_{others} + i_Z + i_U = N \\ i_U > i_Z \end{cases} \quad (26)$$

where $P_{Z|0}$ and $P_{O|0}$ are similarly calculated using $f_R(R; k, P_{in}|m=0)$ instead of $f_R(R; k, P_{in}|m=1)$. Thus, the total error probability can be derived using (19).

C. Difficulties in Analytical Evaluations

In order to evaluate the above equations numerically, the integration of the Rice PDF (12)–(15), and (20)–(21) should be calculated. However, the growth of the two parts of (6) is at opposite directions, i.e., for small values of the Bessel function, the envelope becomes large and vice versa. Thus, whenever the argument of the Bessel function is large, $I_0(\cdot)$ cannot be computed easily. Therefore, in this case, we use the following approximation:

$$I_\alpha(x) \cong \frac{1}{\sqrt{2\pi x}} e^x \text{ if } x \gg \left|\alpha^2 - \frac{1}{4}\right|.$$

By using this approximation, (6) simplifies to

$$\frac{A}{{\sigma_n}^2} e^{-(A^2+R^2)/2{\sigma_n}^2} I_0\left(\frac{AR}{{\sigma_n}^2}\right) \cong \frac{A}{{\sigma_n}^2} e^{-(A^2+R^2)/2{\sigma_n}^2} \frac{\sigma_n}{\sqrt{2\pi AR}} e^{AR/{\sigma_n}^2} = \frac{1}{\sigma_n}\sqrt{\frac{A}{2\pi R}} e^{-(A-R)^2/2{\sigma_n}^2}. \quad (27)$$

Thereafter, by using $f_R(R, k, p_{in}|m=\{0,1\})$, P_{Unity_Ring} and P_{Zero_Ring}, P_e can be calculated numerically. In addition, for calculation of (17)–(18) and (25)–(26), we need to evaluate $n!$ which is computationally expensive for large values of n. However, using Sterling approximation, we get

$$n! \approx \left(\frac{n}{\pi}\right)^n \cdot \sqrt{2\pi n} \text{ for } n \gg 1.$$

V. Simulation Results

For simulations, one minute audio signals of various types such as classic, jazz, rock, and vocal with male and female singers are used; the signals are sampled at 44 100 Hz and quantized with 16 bits. The host signal is divided into frames of length 400 for both hard and soft quantization methods. The results are averaged over 50 simulations. In the hard quantization method, g_1, q_1, q_2, and g_2 are taken as 0.05, 0.15, 0.23, and 0.33, respectively, for acceptable distortion. The number of quantization levels in the soft quantization method is fixed to 15. Here, the synchronization is assumed to be perfect.

Fig. 7(a) demonstrates the bit error rate (BER) of the hard quantization after AWGN attack. As shown in this figure, the method is completely robust for SNRs greater than 25 dB and it smoothly increases when the SNR decreases. From Fig. 7(b) it is shown that the soft quantization outperforms the hard quantization method when SNR $\geq$ 15 dB; otherwise, its BER is worse than the other method. For both cases, there is a good similarity between the theoretical and experimental curves which confirms the analytical derivations and approximations in Section IV.

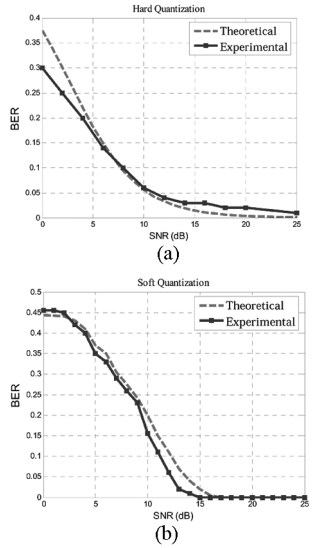


Fig. 7. BER versus SNR for theoretical and experimental simulations for (a) hard quantization and (b) soft quantization.

TABLE I
Robustness of the Proposed Methods Against Some Common Attacks

	Echo		Requantization	LP		DC
	0.5,100msec	0.2,100msec	16-8	6 kHz	3kHz	attack
Hard quant.	11%	8%	0	0	10%	0
Soft quant	9%	6%	0	0	3%	0

TABLE II
BER(%) of the Proposed Methods for Various Sampling Rates (Hz)

	44/11/44	44/6/44	44/4/44
Hard Quant.	0	0	2.6%
Soft Quant.	0	0	0

Robustness of the algorithms against various common attacks such as echo, requantization, LP and DC is investigated in Table I. As demonstrated in this table, both techniques do not have high robustness against echo attack. For example, for echo (50%, 100 ms), the BER of both methods are more than 9%, whereas, the soft quantization method works slightly better. Instead, both approaches are completely robust against requantization and dc attacks. As the watermarked message is embedded in low-frequency components (lower than 6 kHz), the proposed methods are robust against LP attack as well. As shown, in most attacks, soft quantization method outperforms the hard quantization one. This can be explained by the fact that, for soft quantization, the code-words are rings instead of circles and thus each sample bears less distortion than that of the hard quantization method. In this case, by considering inaudibility, more quantization levels can be used for data hiding, which results in more robustness against most attacks.

TABLE III
ROBUSTNESS OF THE TWO METHODS AGAINST VARIOUS ATTACKS USING DIFFERENT PPG INDEX PARAMETER k

	Noise (SNR=25)			LP(3 kHz)			Resampling(50%)			Echo(100msec,20%)		
PPG Index k	2	4	10	2	4	10	2	4	10	2	4	10
Hard Quant.	0	0	0	15%	10%	4%	0	0	0	7%	8%	12%
Soft Quant.	0	0	1%	15%	3%	1%	0	0	0	5%	6%	10%

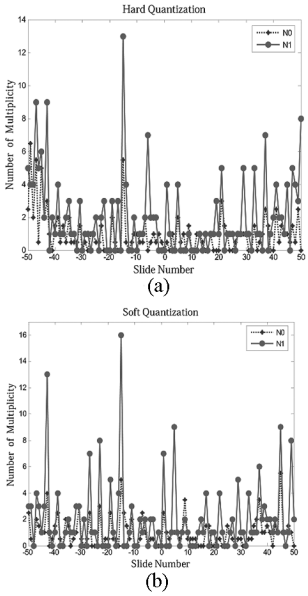


Fig. 8. Number of multiplicity for different sliding shifts in (a) hard quantization and (b) soft quantization.

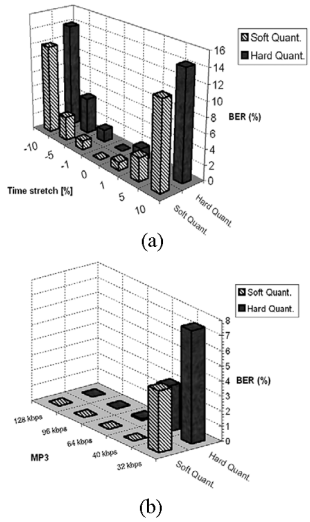


Fig. 9. BER versus (a) the time stretching and (b) MP3 attacks with various rates.

The performance of the synchronization method for finding the location of the frame is demonstrated in Fig. 8. The samples of the frames are shifted by 17 and then the synchronization process is applied. In Fig. 8(a) the values of ρ_0 and ρ_1 are depicted for the hard quantization method. The same results are shown in Fig. 8(b) for soft quantization. As we can see, for both cases the maximum value of ρ is obtained at the 17th shifted sample which confirms the accuracy of the method.

The stability of the algorithms against time scaling and coding attacks is investigated in Fig. 9. Fig. 9(a) depicts the BER of suggested techniques against the time stretching, and Fig. 9(b) demonstrates the robustness of the methods against MP3 attack with various rates. Robustness of the proposed method against resampling is depicted in Table II. This table shows the proposed methods are highly robust against this type of time scaling attack when the scaling factor is known.

In Table III, the sensitivity of the methods against PPG mapping index k, is investigated. By increasing k, the BER increases slightly against noise attack. This can be explained by (8) and (10). By increasing k, the distribution function in (11) tends slightly to right which in turn results in higher BER according to (14)–(15) and (22)–(23). However, since the index k is very small in (8), it does not have significant effect on BER.

For LP attack, increasing k, we get better performance for all methods. Thus, by choosing k, there is a trade-off between robustness against various attacks such as echo, noise and LP. In addition, as mentioned in Section IV, the greater k, the more transparency of the embedded message is in the time domain shape.

TABLE IV
BER PERFORMANCE COMPARISON FOR DIFFERENT KINDS OF ATTACKS WITH THE TECHNIQUE IN [19]

Methods	Payload (bps)	AWGN, 15 dB	Resampling , 44/4/44	MP3, 64kbps
[19]	172	7.5%	12.21%	4.34%
Hard Quant.	176	8.8%	13.1%	10.2%
Soft Quant.	176	3.9%	4%	0.06%

Finally, Table IV summarizes the watermark detection results against various attacks. The comparative criterion is BER in detection procedure. In this scheme, the bit rate contains both watermarked data and synchronization code. Due to the diversity of data embedding approaches, different audio samples, visibility of the watermarked data, and payload, the comparison is not straightforward. We set the bit rate of the proposed method near the payload presented in [19] and also we calculate the average BER over both light and march music. Thus, we decrease the frame length from 400 to 250. The reason for this comparison is that for both proposed methods and the reported technique in [19], data embedding is performed by quantization; while at the decoder, we use hard decoding instead of quantization for data extraction.

Table V demonstrates the BER of the proposed technique with one of the recently reported techniques introduced by Chen and Wu [23]. We set the frame length of our methods to 400 which is far less than the frame size of their algorithm (4000). The distortion of our data embedding is less than this method too. Nevertheless, the proposed method has outstanding

TABLE V
BER COMPARISON OF THE PROPOSED METHODS WITH THE TECHNIQUE OF [23] FOR SOME COMMON ATTACKS

Methods	Payload bps	AWGN 20dB	Filtering 8kHz	MP3 64kbps	Resampling 44/22/44	Requantization 16-8	Echo 40%, 100 ms
[23]	11	22.8%	4.2%	6.5%	6.4%	11.9%	1.6%
Hard Quant.	110	2.9%	0%	0.2%	0%	0%	10.2%
Soft Quant.	110	0%	0%	0%	0%	0%	8.1%

robustness in comparison with their algorithm. The only exception is the echo attack where our method has higher BER than their method. The higher resistance against attacks with less imposed distortion makes the efficiency of the suggested scheme attractive.

VI. CONCLUSION

By using PPG transform, two novel techniques for digital watermarking have been proposed. In both methods, data embedding is performed by shaping the configuration of the PPG points. Among them, the soft quantization method has better performance than the hard quantization method except for the highly noisy environments where hard quantization works better. Besides, the complexity cost of the hard quantization method is slightly less than that of the soft quantization for both embedding and detection. Since the watermark signal is embedded in low-frequency components of the audio signal, both algorithms are highly robust against lowpass, resampling and MP3 attacks; but both methods are not robust against echo attacks. Sensitivity of the proposed algorithms against the PPG mapping index shows that by increasing k, there is more transparency visually. On the other hand, by increasing k, the robustness of the algorithm against AWGN and echo attack is slightly decreased. Hence, there is a trade-off for choosing k. Our future work may be performed on inserting synchronization codes as well as the watermark data to make the algorithm robust against de-synchronization and echo attacks.

REFERENCES

[1] I. Cox, M. Miller, and A. McKellips, "Watermarking as communication with side information," *Proc. IEEE*, vol. 87, no. 7, pp. 1127–1141, Jul. 1999.

[2] P. Moulin and A. Ivanovic, "The zero-rate spread-spectrum watermarking game," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1098–1117, Apr. 2003.

[3] M. Swanson, B. Zhu, A. Tewlik, and L. Boney, "Robust audio watermarking using perceptual masking," *Signal Process.*, vol. 66, no. 3, pp. 337–355, May 1998.

[4] W. Trapper, M. Wu, Z. J. Wang, and K. J. R. Liu, "Anti-collusion finger-printing for multimedia," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1069–1087, Apr. 2003.

[5] M. M. Yeung, "Digital watermarking," *Commun. ACM*, vol. 41, no. 7, pp. 31–33, Jul. 1998.

[6] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security: theory and practice," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3976–3987, Oct. 2005.

[7] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 563–593, Mar. 2003.

[8] S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1639–1667, Jun. 2002.

[9] M. Barni and F. Bartolini, *Watermarking Systems Engineering: Enabling Digital Assets and Other Applications*. New York: Marcel Dekker, 2004.

[10] O. T.-C. Chen and W.-C. Wu, "Highly robust, secure, and perceptual-quality echo hiding scheme," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 3, pp. 629–638, Mar. 2008.

[11] H. M. A. Malik, R. Ansari, and A. A. Khokhar, "Robust data hiding in audio using allpass filters," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1296–1304, May 2007.

[12] F. Balado, K. M. Guenole, C. M. Silvestre, and N. J. Hurley, "Joint iterative decoding and estimation for side-informed data hiding," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 4006–4019, Oct. 2005.

[13] B. Chen and G. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[14] B. Chen and G. Wornell, "Achievable performance of digital watermarking schemes," in *Proc. IEEE Int. Conf. Multimedia Computing Systems*, Florence, Italy, Jun. 1999, vol. 1.

[15] J. J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, "Scalar costa scheme for information embedding," *IEEE Trans. Signal Process*, vol. 51, no. 4, pp. 1003–1019, Apr. 2003.

[16] Q. Zhang and N. Boston, "Quantization index modulation using E8 lattice," in *Proc. 41th Annual Allerton Conf. Communication, Control and Computing*, Allerton, IL, 2003.

[17] J. H. Conway and N. J. Sloane, *Sphere Packing, Lattices, and Groups*. New York: Springer-Verlag, 1988.

[18] R. Zamir, S. Shamai, and U. Ertz, "Nested linear/lattice codes for structured multi-terminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.

[19] S. Wu, J. Huang, D. Huang, and Y. Q. Shi, "Efficiently self-synchronized audio watermarking for assured audio data transmission," *IEEE Trans. Broadcast.*, vol. 51, no. 1, pp. 69–76, Mar. 2005.

[20] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2004.

[21] H. O. Kim, B. K. Lee, and N. Y. Lee, Wavelet-Based Audio Watermarking Techniques: Robustness and Fast Synchronization. [Online]. Available: http://amath.kaist.ac.kr/research/paper/01-11.pdf.

[22] J. W. Huang, W. Yong, and Y. Q. Shi, "A blind audio watermarking algorithm with self-synchronization," in *Proc. IEEE Int. Symp. Circuits Systems*, Phoenix-Scottsdale, AZ, 2002, vol. 3, pp. 627–630.

[23] O. T.-C. Chen and W.-C. Wu, "Highly robust, secure, and perceptual-quality echo hiding scheme," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 3, pp. 629–638, Mar. 2008.

Mohammad A. Akhaee (S'07) received the B.Sc. degree in both electronic and communication from Amirkabir University of Technology, Tehran, Iran, and the M.S.c degree from Sharif University of Technology, Tehran, in 2005. He is currently pursuing the Ph.D. degree in the area of statistical signal processing at the Electrical Engineering Department of Sharif University of Technology.

His research interests include multimedia security, watermarking, speech, and image processing.

Mohammad J. Saberian, biography and photo not available at time of publication.

Soheil Feizi, biography and photo not available at time of publication.

Farokh Marvasti (S'72–M'74–SM'83) received the B.Sc., M.Sc. and Ph.D. degrees, all in electrical engineering, from Rensselaer Polytechnic Institute, Troy, NY, in 1970, 1971, and 1973, respectively.

He was an Associate Professor at the Illinois Institute of Technology, Chicago, during 1987–1991, and from 1992 to 2003 he was an instructor at King's College, London, U.K. He was granted several research grants. Currently, he is a Professor of Sharif University of Technology, Tehran, Iran. His general research interests lie in signal processing, multiple access, MIMO techniques, and information theory.