Quantization Based Audio Watermarking in a New Transform Domain

Mohammad A. Akhaee, Amir Nikooienejad, and Farokh Marvasti
Advanced Communications Research Institute (ACRI),
Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran
Emails: akhaee@ee.sharif.edu, nikooienejad@ee.sharif.edu, and marvsti@sharif.edu
http://acri.shrif.edu

Abstract

In this paper, a novel blind watermarking technique based on quantization is proposed. Quantization is performed in a special domain which converts one dimensional signal to a 2-D one named Point to Point Graph (PPG). Basis of the method is on the separation of this domain into two portions; while, only one portion is quantized. Furthermore, in the dewatermarking procedure, by using the unquantized portion and zero norm, the embedded data can be extracted. The performance of the proposed method is analytically investigated and verified by simulation with artificial Gaussian signals. Experimental results over several audio signals shows the great robustness of the technique in comparison with the algorithms presented so far.

***Index Terms*—** Audio watermarking, Quantization, Point to Point Graph Transform.

I. Introduction

Audio watermarking, has various applications such as authentication, fingerprinting, copyright protection, and broadcasting monitoring [1]-[2]. It is necessary that the watermark signal is readily extracted from the watermarked signal in order to illustrate the copyright owner. Three important and basic issues that should be considered in watermarking systems are: 1) perceptual transparency, 2) robustness against attacks, and 3) data rate of the watermark. There are trade-offs between the capacity of the watermarked data and the robustness against watermark attacks, while keeping the perceptual quality of the watermarked audio signal at acceptable levels. It is not possible to attain a high robustness against various attacks and high watermark data rate simultaneously. The trade-offs among the embedding rate, distortion, and robustness have been examined from an information-theoretic perspective [3], [4]. In [3], a model of the watermarking game is introduced and upper and lower bounds on the data embedding capacity are determined. In [4], Cohen and Lapidoth investigated information rates for Gaussian host signals and the squared-error distortion measure.

In fact, [1] is the first paper which relates the watermarking techniques and communications over a channel with side information at the decoder. After this work, data embedding processes are involved in practical schemes [5]. The aim of these schemes is to embed some information in the host signal in such a way that causes the least interference; while at the same time, has a high robustness against different attacks. In this framework, the Quantization Index Modulation (QIM) method is an important blind watermarking class performing close to the system capacity [6], [7]. 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 a different message. For example, the selection of scalar quantizers results in a method named Dither Modulation (DM) with bit repetition or a Distortion Compensated version of DM (DC-DM) [6], [8].The main problem of this approach is the development of the quantizer codebooks which work for data embedding. This problem is solved by using the lattice-based quantizers which are efficient in data embedding and decoding [9], [10]. Recently, it has been shown that the channel capacity may be achievable using lattice codebooks [11].

Here, we propose a novel technique based on the quantization in a new transform domain coined PPG. This transformation is a mapping which increases the dimension of the signal and utilizes the correlation among the samples as well as their magnitudes. In the proposed technique ,half-zone Quantization method, the data embedding is performed by dividing the PPG domain into two portions and quantizing the points in just one part of them. In fact, these method is based on forming different patterns in the radii of points according to the embedded bit. At the decoder, in spite of the QIM method, a histogram based dewatermarking procedure has been used for all proposed techniques. Simulation results and analytical discussion demonstrate that these kinds of code-words and decoders yield better performance in comparison with some improved version of QIM methods reported so far [12].

II. Preliminary

A. PPG Transform

This transformation is based on putting N samples of the signal together in a specific order to convert a 1-dimensional signal to an N-dimensional counterpart. For example, for the case of N=2, this transformation is in such a way that sample 1 and sample $k+1$, sample 2 and sample $k+2$, ..., sample k and sample $2k$ are formed in pairs. This process is performed on the next $2k$ samples too, where k is the index of PPG transformation. Using this method, a 1-D signal can be

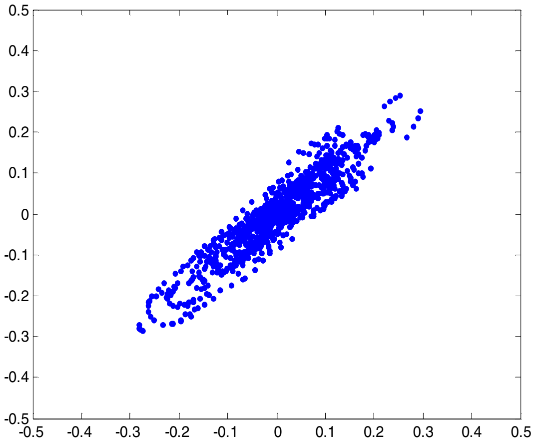

Fig. 1. The PPG points of a 1000 length audio for k=4.

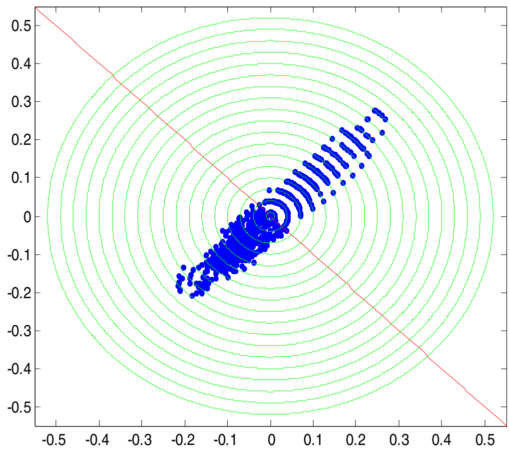

Fig. 2. The ppg points of length 1000 audio after embedding message '0'

converted to a 2-D one. It is worthwhile to mention that the index of PPG transformation (k) can be chosen arbitrarily depending on the desired characteristics. For example, for the case of small k, because of a strong correlation between samples, two components of each PPG points are close to each other. Thus, the PPG points, form an approximate ellipse with an inclination line $y = x$. The direction along the line $y = x$ is caused by its inherent low pass property and this is why each sample cannot have significant variation from its adjacent samples. On the other hand, choosing a greater k causes more dispersion in the PPG points due to the less 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 an approximate circle instead of an ellipse. Fig. 1 demonstrates the PPG points of a 1000 length audio with the index of PPG transformation of 4.

For convenience, we show this transformation with $PPG(n, k)$, in which n is the dimension of the domain and k is the PPG transformation index. Furthermore, in this paper we only concentrate n the 2-D case of $n = 2$, where we denote $PPG(k) = PPG(2, k)$ In the following, the basic concepts of the data embedding in this domain based on the quantization are explained.

B. Quantization in PPG Transform Domain

Quantization process is the basis of QIM watermarking approaches. These methods are blind and cause a controllable and measurable distortion due to the determined distance between levels. First, in order to have symmetry about the origin, the DC component of the signal is eliminated. Then, the quantization is performed on the radii of points. Quantization levels should be co-center circles with the same distance between each other. Sometimes that's better not to have same distance between levels in order to make the procedure more transparent, i.e. in real audio signals. In order to quantize the points, each point is stretched in the direction of the line which connects that point to the origin either upward or downward depending on the embedded bit or the method of quantization.

Quantization in the PPG domain produces correlated noise on the components of each point. In other words, it creates specific patterns in the PPG domain which can be utilized in the dewatermarking procedure. Furthermore, by adjusting PPG transformation index, the time domain effect which is caused by quantization can be controlled, In other words, the greater the index k, the more transparency is in the time domain.

III. PROPOSED METHOD

In this section, we introduce the proposed method which uses the quantization in the PPG transform domain. We call this technique as *Half-Zone Quantization* method.

A. Data Embedding

In order to embed watermark data, just some PPG points are quantized while the others are kept with no change. For instance, for the case of bit '0' embedding, one way is to quantize only points above the line $y = -x$ whose radii are greater than threshold. To embed the bit '1', points outside the threshold level and below the line $y = -x$ are quantized. The quantization in this method is such that each point is mapped to the nearest level in the direction of its radius. The data embedding algorithm can be summarized as follows:

Step 1: The PPG transformation with index k is applied to each frame.

Step 2: To embed bit '0', the appropriate points upper than the line $y = -x$, and to embed message '1', the points below that line are quantized in the direction of their lines to origin. Fig. 2, shows the PPG points of a 1000 length audio, in which the message '0' is embedded. The quantized points with respect to the line $y = -x$ can be easily seen.

B. Data Extraction

In the dewatermarking process, it is sufficient to find the section which is quantized. To achieve this aim, the number of points that are close to each quantization level is counted. Simulation results show that using the zero norm has better

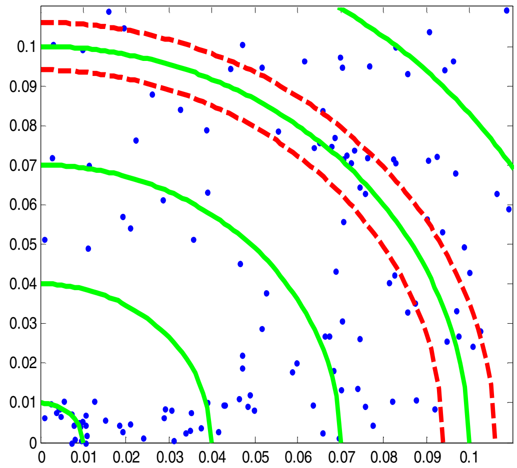

Fig. 3. Levels of quantization and margins to extract the embedded bit

performance than the first and the second norm. Consequently, we use the zero norm in the dewatermarking procedure by counting the points near each quantization level. This closeness is measured by a specific margin around the quantization levels. Therefore, by comparing the number of points in the quantization margins of two counterparts, the embedded bit can be extracted. It is obvious that the portion which is not quantized has fewer points in these margins than the other one. The extraction algorithm can be summarized as follows:

Step 1: The PPG transformation with index k is applied to each frame.

Step 2: The histogram of each portion is derived around quantization margins.

Step 3: By comparing the number of PPG points in the margins of two parts, the embedded bit is extracted.

Here, the value of the margin is such that the algorithm reaches to its optimum performance that is, the least Bit Error Rate (BER). In fact, this margin is hand-optimized via several simulations.

Fig. 3 shows a part of the PPG points in the receiver where we can see the quantization levels and the margins. The margin in this figure is $\frac{1}{5}$ of the distance between two levels of quantization.

IV. Performance Analysis

Consider a zero mean Gaussian signal with $\sigma_x{}^2$ variance contaminated by AWGN with zero mean and variance $\sigma_n{}^2$. Then, a typical PPG point at the receiver is $(x+n_1, y+n_2)$. Since both signals have zero mean and are Gaussian, the distribution of its envelope $r=\sqrt{(x+n_1)^2+(y+n_2)^2}$ is Rayleigh, i.e.,

$$f_R(r)=\frac{r}{\sigma^2}e^{\frac{-r^2}{2\sigma^2}} \qquad r>0 \qquad (1)$$

Where $\sigma^2=\sigma_x{}^2+\sigma_n{}^2$

In half-zone quantization method, at the receiver, the number of PPG points near the quantization levels is counted and the decision is based on the comparison of the obtained numbers of each side (upper and lower of the line y=-x). For evaluation, we need two probabilities. In the region where the quantization process is not performed, the probability of laying a typical PPG point in a noisy environment near the quantization levels, in our threshold margin, can be calculated as:

$$P_{r\,n\_q}=\sum_{i=1}^{M}\int_{q_i-thr}^{q_i+thr}\frac{r}{\sigma^2}e^{\frac{-r^2}{2\sigma^2}}\,dr \qquad (2)$$

Where q_i is the quantization levels and $\sigma^2=\sigma_x{}^2+\sigma_n{}^2$.

In the quantized region, we have $\sqrt{(x^2+y^2)}$. Thus, the envelope of $r=\sqrt{(x+n_1)^2+(y+n_2)^2}$ has a Rician distribution. In fact, the quantization causes a specific mean in the Gaussian host signal. Therefore, the distribution of quantized noisy PPG points can be express as:

$$f_R(r|q_i)=\frac{r}{\sigma_n{}^2}e^{\frac{-(r^2+q_i{}^2)}{2\sigma_n{}^2}}I_0\left(\frac{rq_i}{\sigma_n}\right) \quad r>0 \qquad (3)$$

In fact, in the quantized region, the PPG points which already have Rayleigh distribution are concentrated in some specific points and makes some Dirac function at the quantization levels. Equation (3) shows that after Gaussian noise attack, a Rician distribution is convolve with these Dirac function. Thus, although noise spread the PPG points around quantization levels, still there is higher concentration around these levels than no quantized region. Fig. 4, shows the distribution of noisy PPG points in both region simultaneously.

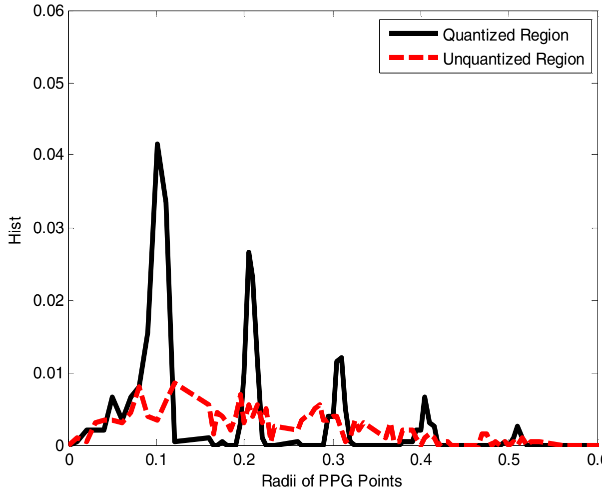

Fig. 4. Distribution of noisy PPG points in both quantized and unquantized regions

As a consequence, the probability of laying a typical noisy PPG point in the quantized region near the quantization levels can be written as:

$$P_{r\,q}=\sum_{i=1}^{M}\int_{q_i-thr}^{q_i+thr}\frac{r}{\sigma_n{}^2}e^{\frac{-(r^2+q_i{}^2)}{2\sigma_n{}^2}}I_0\left(\frac{rq_i}{\sigma_n}\right)\,dr \qquad (4)$$

Error occurs when the upper portion is quantized but at the receiver, the number of PPG points belongs to the desired regions (near the quantization levels) in the non-quantized region (i_{nq}) is more than that of the quantized region (i_q). Therefore

$$P_{e|0} = \sum_{i_q=0}^{\frac{N}{2}} \left(\left[\binom{\frac{N}{2}}{i_q} P_{r\,q}^{i_q} (1 - P_{r\,q})^{(\frac{N}{2} - i_q)} \right] \sum_{i_{nq}=i_q+1}^{\frac{N}{2}} \binom{\frac{N}{2}}{i_{nq}} P_{r\,nq}^{i_{nq}} (1 - P_{r\,nq})^{(\frac{N}{2} - i_{nq})} \right) \quad (5)$$

In fact the first summation is related to the quantized region, while the second one ($\sum_{i_{nq}=i_q+1}^{\frac{N}{2}}(.)$) is related to the unquantized region.

The probability of error in the case of sending '1' is the same as probability of error in the case of sending '0' by symmetry. Therefore the probability of error is:

$$P_e = \frac{1}{2} P_{e|0} + \frac{1}{2} P_{e|1} \quad (6)$$

Which is exactly equal to (5).

V. Experimental Results

A. Artificial Signal

In order to confirm the analytical calculations, first we applied the Half-zone quantization method to an artificial zero mean Gaussian signal with unity variance, which is attacked by a white Gaussian noise. Since evaluating (5) is time consuming, the frame length is set to 512. Fig. 5 demonstrates the Bit Error Rate (BER) of the proposed method versus SNR. As we can see, using 9 quantization levels, the proposed technique is highly robust against noise attack. Moreover, the result of the analytical calculation in (5) is well agree with the BER of the artificial signal which confirms the validity of our performance analysis.

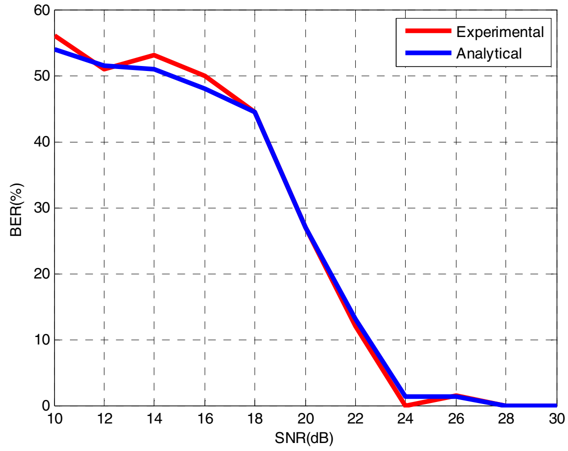

Fig. 5. BER versus SNR for proposed technique, using experimental result and analytical calculations

B. Audio Signal

We also applied the proposed method to several audio signals to test the proposed method for real signals. For this kind of simulation the signals are sampled at 44100 Hz and quantized with 16 bits. The host signal is divided into frames of length 2000 and one bit is embedded in each frame. In order to obtain more transparency as well as robustness, the bits are embedded into the frames with higher energy. The index of PPG transformation, is set to four in all frames.

For a real audio signal, the algorithm would be somehow different, since the transparency and robustness play an important role in this case. The enhanced version of the algorithm can be written as:

Step 1: The DC component of the signal is eliminated.

Step 2: The signal is lowpass filtered at 6 kHz.

Step 3: The PPG transformation with index k is applied to each frame.

Step 4: To embed the bit '0', the appropriate points upper than the line $y = -x$, and to embed the bit '1', the points below that line are quantized.

Step 5: The signal is reconstructed by taking inverse PPG transform

Step 6: The reconstructed signal is again low-pass filtered with the same cut off frequency.

Step 7: The high frequencies content of the original signal is added to the lowpass filtered signal and the watermarked audio signal is sent through channel.

We embed data in low-frequency component of the signal to make it more robust in front of attacks. Since quantization causes producing high frequency components, in step 6 we again apply the watermarked signal through lowpass filter to make it transparent. Finally, we add the high frequency components of the host signal to reconstruct the watermarked audio signal. Fig. 6 shows the block diagram of this enhanced algorithm.

It noteworthy that in the simulation of the real signals, since imperceptibility of the watermarked signal is the most important factor, we may not have equidistance quantization levels. In order to have the best transparency and robustness simultaneously, the location and the number of quantization levels are attained using an iterative technique. To this aim, we use Perceptual Evaluation of Audio Quality software [14] to measure the transparency of the method. This software uses some features of both reference and test audio signals and represents the quality in the form of Objective Difference Grade (ODG). ODG values are between 0 and -4 for PEAQ. Higher ODG values show more perceptual similarity of the reference and test signals. By using this software, the location and the number of quantization levels are hand-optimized in such a way that with the same ODG value of -0.3, the best robustness is achieved. By using this optimization approach,

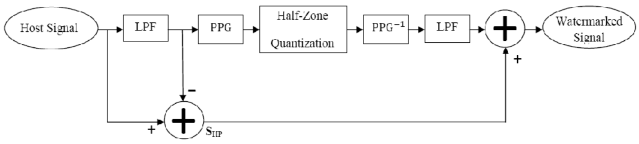

Fig. 6. Block Diagram of Half-Zone Method in the PPG domain

the proposed watermarking method is well matched to audio signals.

Another point that should be mentioned is that after low-pass filtering the density around quantization levels is slightly decrease. Therefore, to have acceptable robustness against common attacks, the frame length is chosen higher than artificial signals. In the following, the robustness of the algorithm against different common attacks such as echo, LP, Noise and MP3 is investigated.

1. Noise Attack

In this part, the AWGN attack is investigated. Fig. 7 shows the robustness of the proposed method against this attack. The results are evaluated over 20 simulations.

2. LP Attack

For this kind of attack, we low pass the watermarked signal with 3 cut off frequencies i.e. 5, 4 and 3 KHz, shown in Table 1. The results show that the method has high robustness against the low-pass attack. Particularly, for the case of 3KHz LP attack which destroys significantly the quality of the audio signal, the algorithm has acceptable performance.

TABLE I

ROBUSTNESS OF THE PROPOSED METHOD AGAINST LP ATTACK

Cut-off Frequency	5KHz	4KHz	3KHz
Average BER (%)	0.25	0.3	1.6

3. Echo Attack

The robustness against echo attack is demonstrated in Table. 2. The time delay for this attack is 100 milliseconds, with two different attenuation coefficients.

TABLE II

ROBUSTNESS OF THE PROPOSED METHOD AGAINST ECHO ATTACK

Attenuation Coefficient	20%	50%
Average BER (%)	25	52.4

As demonstrated in above table, the proposed technique does not have high robustness against echo. For example, for echo (20%, 100 msec), we have BER around 25%

4. MP3 Attack

The robustness of proposed method against MP3 attack with various rates is shown in Table. 3.

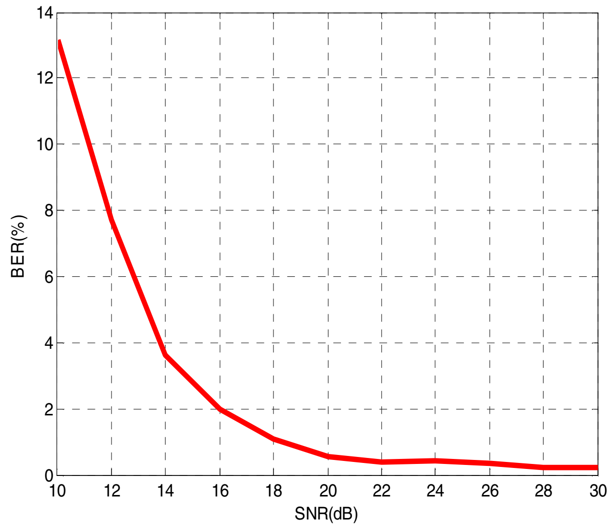

Fig. 7. BER versus SNR, applying proposed technique to an audio signal with noise attack

TABLE III

ROBUSTNESS OF THE PROPOSED METHOD AGAINST MP3 ATTACK

Bit Rate	128	96	64
Average BER (%)	0	0	3.2

It shows that the technique has a good robustness against MP3 attack and

C. Comparison with other Techniques

Finally, Table 4 summarizes the watermark detection results comparing to the schemes proposed in [12], against various attacks. The comparative criterion is BER in detection procedure. Although, due to the diversity of data embedding approaches, different audio samples, visibility of the watermarked data and data payload the comparison is not straightforward, we set the bit rate of the proposed method near the payload presented in [12] and also we calculate the average BER over both march and light music.

TABLE IV

ALGORITHM COMPARISON FOR DIFFERENT KINDS OF ATTACKS (BER)

Method	Payload (bps)	AWGN (15dB)	LPF (3KHz)	MP3 (64Kbps)
[12]	44	2.7%	12.21%	8.34%
Half-Zone Quant	44	3.1%	1.6%	3.2%

We can see that the proposed technique competes with [12] which is one of the best audio watermarking methods proposed recently.

VI. CONCLUSION

Using PPG transform, a blind technique for digital watermarking has been introduced. In the proposed method, data embedding is performed by shaping the configuration of the PPG points. Since the watermark signal is embedded in low frequency components of the audio signal, the suggested algorithms are highly robust against lowpass, echo and MP3 attacks. The performance of the technique is analytically investigated and verified via simulation. The half zone quantization can be applied in all 1-dimansion signals such as audio and speech signals. As a future work, we are intending to extend the proposed algorithm to 2-dimansion signals.

REFERENCES

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

[4] A. S. Cohen and A. Lapidoth, "The gaussian watermarking game," IEEE Trans. inf. Theory, vol. 48, no. 6, pp. 1639-1667, June 2002.

[5] 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.

[6] 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.

[7] B. Chen and G. Wornell, "Achievable performance of digital watermarking schemes," *Proc. IEEE Int. Conf. Multimedia Comput. Syst*., vol. 1, Florence, Italy, June 1999.

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

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

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

[11] 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, June 2002.

[12] 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, Mar. 2005.

[13] F. Perez-Gonzalez, F. Balado, and J. R. Hernandez, "Performance analysis of existing and new methods for data hiding with known-host information in additive channels," *IEEE Trans. Signal Process*., vol 51, no 4, pp. 960-980, Apr. 2003.

[14] http://www-mmsp.ece.mcgill.ca/Documents/Software/ index.html