Real-time Covert Communications Channel for Audio Signals

Ashraf Seleym[1], Dina Darwish[2]

[1]Ph.D. student in IT Department, Faculty of Computers and Information, Cairo University, Cairo, Egypt

[2]Assistant professor, Multimedia and Internet Department, International Academy for Engineering and Media Science, 6th October city, Egypt

Abstract

Covert communications channel is considered as a type of secure communications that creates capability to transfer information between entities while hiding the contents of the channel. Multimedia data hiding techniques can be used to establish a covert channel for secret communications within a media carrier. In this paper, a high-rate covert communications channel is developed to exploit an audio stream as a carrier signal using multiple embedding in the Quantization Index Modulation framework. The proposed approach uses multi quantization vectors to increase data transmission rate. The embedding algorithms consider the embedding process as a communications problem, that it uses structured scheme of Multiple Trellis-Coded Quantization jointed with Multiple Trellis-Coded Modulation. Using convolution codes based trellis coding returns a real-time communications, because it can be continuously encoded and decoded. The proposed approach exhibits a high channel capacity due to the increase in data embedding rate without severely increasing in embedding distortion.

Keywords: *Hidden communications, Real time communications, quantization index modulation, Embedding Process, Audio Signals communications.*

1. Introduction

COVERT channels are one recent method to provide reliable security and can be implemented using data hiding. Steganography is the science of hiding data into different covers such that the data embedded is imperceptible; the aim is to store and deliver secret messages in digital objects without suspicion [1]. The rapid development of the broadband communication networks and multimedia data available in a digital format opened many challenges and opportunities for securing data transmission [2]. When taking audio as a carrier for information hiding, it is called *Audio Steganography*[3]. There is an explosive growth of using audio data on the Internet today. Audio has become a very significant medium due to voice over IP (VOIP) popularity. Digital audio players and broadband Internet connections have made it possible for consumers from all over the world to create, exchange, and distribute large digital multimedia files. In addition sharing digital electronic files using mobile phones has grown extremely fast over the last decade. The Internet as a whole does not use secure links [4], thus information in transit may be vulnerable to interception, so it is important to reduce a chance of the information being detected during the transmission [5], [6].

Covert communications channel over audio signals can transfer hidden data embedded into the digital representation of an audio stream [7]. This data should not be hearable to human ear. In order to design a relative optimal data hiding system for audio, it is important to take the Human Auditory System (HAS) into account [8]. Embedding additional information into audio sequences is classified as especially challenging, because the HAS operates over a wide dynamic range. The human ears can perceive frequencies that are in the range of 20 Hz to 20 KHz. The frequencies that are located below and above this hearing bandwidth are called infra-sounds and ultra-sounds, respectively [9]. Embedded data that is not detectable by the human hearing system has to be neither located in the audible region nor have high level power in the audible region. Sensitivity to additive random noise is also acute [10]. The perturbations in a sound file can be detected by HAS as low as one part in ten million (80 dB below ambient level) [11]. However, there are some "holes" available, while the HAS has a large dynamic range, it often has a fairly small differential range. As an instance, loud sounds tend to mask out quieter sounds. Additionally, the HAS is unable to perceive absolute phase of a sound, only relative phase [11]. Moderately shifting in the absolute phase of an audio signal interprets as natural, perceptually not annoying ones hearing. Finally, there are some environmental distortions so common as to be ignored by the listener in most cases.

In general, two properties of the HAS dominantly exploited in data hiding algorithms are the frequency masking and the temporal masking. Frequency masking is a frequency domain phenomenon where low levels signal can be made inaudible by a simultaneously appearing stronger signal. A masking threshold can be found and is the level below which the audio signal is not audible. Thus, frequency domain is a good region to check for the possible areas that have imperceptibility [12]. In temporal masking, two phenomena of the HAS in the time domain play an important role in human auditory perception. Those are pre-masking and post-masking in time. The temporal masking effect appears before and after a masking signal has been switched on and off, respectively. The duration of the pre-masking is significantly less than one-tenth that of the post-masking, which is in the interval of 50 to 200 milliseconds.

Both pre- and post-masking have been exploited in the audio compression and several audio data hiding algorithms [13]. In this paper, a real time high-rate covert communications channel is developed to exploit an audio stream as a carrier signal using multiple embedding in the Quantization Index Modulation framework.

The rest of this paper is organized as follows: We introduce a general overview on steganography in Section II.. Next, in Section III, a real time covert communications channel for audio signals is proposed . It is an extension development to the Quantization Index Modulation (QIM) using multidimensional jointly source coding, and modulation. Experimental results for establishment of covert channels in audio signals are introduced in Section IV. Finally, conclusions remarks are included in Section V.

2. Steganography Overview

The general framework of a data hiding system is shown in Fig. 1, a secret message, m, need to be embedded into the coefficients of a host signal, c (named, cover-work), to get the composite signal, s (named, stego-cover). The received signal, y corrupted by channel noise, from which the receiver estimates the message, m' (a noisy copy of m), that was hidden. A natural requirement of a data hiding system is that, there should not be any perceptual distortion during the embedding process. This is modeled by a constraint on the amount of change that is made to the host signal [14- 16]. In audio steganography, the distortion constraint is a complex function motivated by the HAS. For simplicity of analysis, an average mean squared distortion D is quite commonly used as in the following formula:

$$D(c,s)=\frac{1}{w}\sum_{h=1}^{w}\left(c_h-s_h\right)^2 \qquad (1)$$

where, $h = \{1, 2 \ldots w\}$, and w is the total number of altered cover-work elements.

The data hider is allowed to induce the distortion of at most D_1, i.e., $D(c,s)\le D_1$, and the maximum distortion due to channel noise should not be greater than D_2, where, $D(c,y)\le D_2$. The constraint at the encoder is similar to power constraint in the classical communication setting. Likewise, the channel distortion constraint is equivalent to the noise power.

Multimedia steganographic systems can be classified according to the cover modifications strategy that applied in the embedding process. This classification fall in one of two main types [15, 16].Type-I, Known-Host-Statistics: this type performs statistical embedding using the host signal characterization. Data embedding systems that use this type are simple to understand and easy to implement. However, it can be shown that under certain conditions, they are not secure at all. Type-II, Known-Host-State: includes the data hiding methods that uses host signal state or Side Information (SI) at the encoder and performs deterministic decoding (deterministic means, the correct next step depends only on the current state) without resorting to host signal statistical characterization. As the original cover-work is known to the embedder, then the data hiding encoder can represent the secret message with a codeword vector dependant on the cover-work. The result is, the stego-cover is highly correlated with the cover-work. This enhances the perceptual fidelity. The proposed technique presented in this paper depends mainly on type II.

By considering the knowledge of the host signal state at the encoder, the data hiding problem can be modeled as communications with SI about the channel state at the encoder [16]. The general framework of this type is illustrated in Fig. 2, which can be considered as an update of Fig. 1.

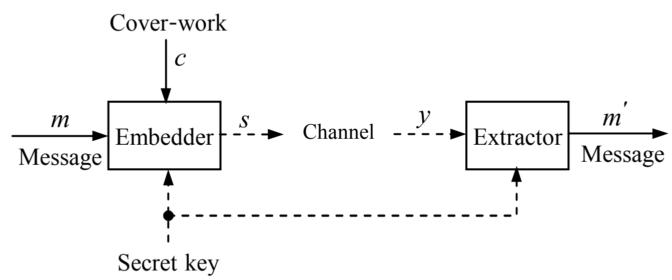


Fig. 1 General framework of a data-hiding system.

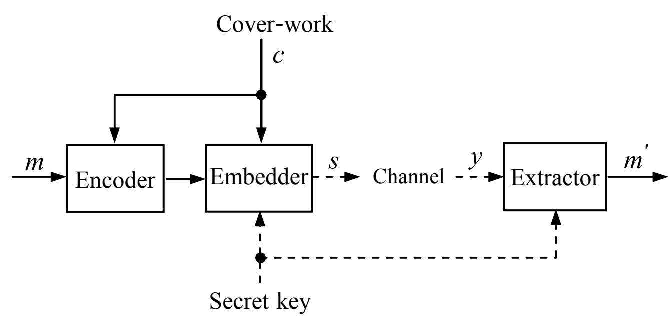
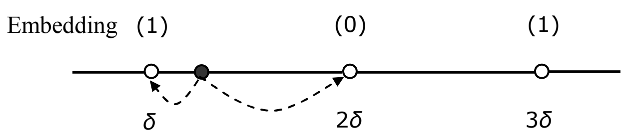


Fig. 2 Data hiding as communications with side information.

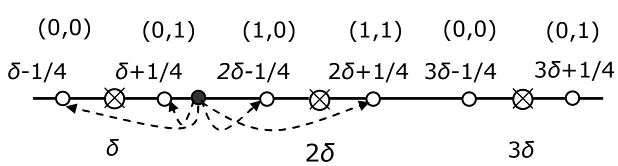
Millar *et al.* [17] modeled the data hiding with blind detection as a communication with SI at the embedding side. They developed the design of algorithms for informed coding and informed embedding. In informed coding, the secret message is represented by a codeword vector that is dependent on the cover-work. In informed embedding, the secret message added pattern is tailored according to the cover-work, attempting to attain an optimal trade-off between estimates of perceptual fidelity and robustness.

Chen and Wornell [18] proposed a high capacity embedding scheme called quantization index modulation, based on quantizing the host samples. The data is hidden by choice of a quantizer (based on the message to be hidden) at the encoder. The decoder just determines which of the possible quantizers were used to extract the hidden message. For the single embedding, let the QIM embedding logic be converting an element to the nearest even/odd multiple of the quantization interval, δ, to embed 0/1, respectively as shown in Fig. 3.a.

Sarkar and Manjunath [19] proposed the use of double embedding in the QIM framework, where a single coefficient is modified twice using two quantizers. The two quantizers can be obtained by decreasing/increasing every even/odd multiple of the quantization interval points by $\delta/4$ to have four quantization codewords. The embedding is done by modifying the cover-work coefficient once to one of the four nearest codewords to embed two bits as shown in Fig. 3.b.



(a) Single embedding.



(b) Double embedding.

Fig. 3 Traditional QIM Techniques.

The motivation behind substituting single embedding with double embedding in the QIM framework for a certain steganographic scheme is to increase its embedding rate. QIM can be categorized as a communication with SI based data hiding model, because, information about the cover-work (its quantized and indexed coefficients) is known to the embedder.

3. PROPOSED REAL TIME COVERT COMMUNICATIONS CHANNEL FOR AUDIO SIGNALS

In this paper, a particular application of audio data hiding systems consists of using the audio signal as a hidden communications channel for binary information transmission. Exploiting the frequency masking property in the frequency domain of audio signals, and that the HAS is unable to perceive absolute phase but only relative phase, the proposed data hiding technique uses the phases of the low frequency components to be modulated by the required secret data. The Discrete Fourier Transform (DFT) is used to convert the audio cover-work from time domain to frequency domain where the hidden data is embedded in selected regions. After the data is embedded, the Inverse Discrete Fourier Transform (IDFT) is performed to revert back to the time domain.

Establishment of a covert channel in the audio host signal's phase is done by applying a phase modulation process. The proposed algorithm embeds secret data into the phase by performing independent multi-band M-ary PSK modulation techniques. This algorithm not only uses the frequency masking properties of the HAS as it works in the low frequency regions of the signal spectrum, but exploits the fact that, the HAS is insensitive to an absolute phase shift in a stationary audio signal. This leads to inaudible phase modifications. The PSK has been proved as an effective approach to embed data in audio signals [20], [21]. The idea of exploiting angel modulation was previously presented for the case of using single embedding in the QIM framework in [22], [23].

3.1 Frequency Spectrum Analysis

The Discrete Fourier Transform (DFT) takes a discrete signal in the time domain and transforms that signal into its discrete frequency domain representation. A sequence of N–points discrete-time signal, $x(t)$ can be transformed into another sequence of N-complex numbers according to the DFT using the following Euler based formula [24]:

$$X(f) = \sum_{t=0}^{N-1} x(t)\left[\cos 2\pi(ft/N) - j\sin 2\pi(ft/N)\right] \tag{2}$$

where, $f = 0, 1, \ldots N\text{-}1$

The Inverse Discrete Fourier Transform (IDFT) is given as:

$$x(t) = \frac{1}{N}\sum_{f=0}^{N-1} X(f)\left[\cos 2\pi(ft/N) + j\sin 2\pi(ft/N)\right] \tag{3}$$

where, $n = 0, 1, \ldots N\text{-}1$

The sinusoidal magnitude and phase from the complex modulus and argument of $X(f)$ can be obtained respectively as follows:

$$A_f = |X(f)| = \sqrt{\mathrm{Re}[X(f)]^2 + \mathrm{Im}[X(f)]^2} \tag{4}$$

$$\varphi_f = \arg[X(f)] \tag{5}$$

One of the basic properties of the Spectrum of DFT is its even symmetry around the center point. It is noted that, the coefficients appearing in DFT structures have two mirror image sides of complex conjugates magnitudes. The coefficients start with Direct Current (DC) component followed by components with ascending frequencies from low frequency elements represented in the right side to high frequency elements represented in the left side. The proposed embedding algorithm instates the altering region to be in the right side of the DFT spectrum to embed the secret data in the low frequency coefficients.

The Fast Fourier Transform (FFT) is a faster version of the DFT as it utilizes some clever algorithms to do the same thing as the DTF, but in much less time. The execution time of an FFT algorithm depends on the transform length. It is fastest when the transform length is a power of two [25].

3.2 Audio Signal Host Regions Selection

Among the frequencies of the DFT elements, only bins with high power and low frequency are selected by the proposed algorithm for embedding the secret data bits. The reason of hiding the data in these regions is that the low frequency is below the perceptual threshold, making it imperceptible for the human auditory system. The performed phase modifications are done in the high-energy peaks to satisfy two important features. The first is to ensure the audio signal perceptual transparency by introducing only small changes in the envelope. The low distortion in the envelope is due to selecting frequency components with high magnitudes. That because, the relative distortion in high amplitude (the altered relative to the unaltered) due to phase modification of the selected components is lowers than that of low amplitude. The second is the robustness against modification attacks due to embedding in the significant high power bins of the audio signal.

For selecting high-energy peaks in the proposed technique, a threshold is chosen above which all such peaks are considered as binning points. This threshold is taken as the all value or a fraction of the Root Mean Square (RMS) of the magnitudes in the frequency domain representation depending on the required degree of robustness and fidelity. One of the important winnings of this idea is the good balance between robustness and fidelity properties in the embedding algorithm.

3.3 Coded Quantization / Modulation

Consider a space ζ of C vectors, where C represents the coefficients corresponding to the original cover-work. To embed a bit sequence b of length l, we should be able to define a constellation with a minimum of 2^l points in ζ. The problem now is the choice of a signaling set or a signal constellation, such that any point in ζ can be relocated to a point in the constellation corresponding to the arbitrary bit sequence to be hidden without perceptual distortion. The new points to which the cover-work is moved is then the stego-cover. If the space ζ signal points are well separated in the constellation, then the hidden bits can survive under a severe level of noise distortions.

Motivation for the use of multidimensional signals for digital transmission dates back to the work of Shannon [26]. In his analysis of the limit performance achievable in digital communication over a given channel, he recognized that the performance of a signal constellation used to transmit information over the AWGN channel can be improved by increasing the dimensionality of the signal set used for transmission. In particular the dimension number grows to infinity; the performance tends to an upper limit that defines the capacity of the channel. Heuristically, as the number of dimensions increases we have more space to accommodate the signals, and hence the distance between signal points increases. In turn, a greater distance between signal points means a smaller error probability (at least for high enough signal-to-noise ratio) in detection of the received signal.

The motivation of this paper is increasing the signal space dimensionality by substituting the single or the double embedding with multiple embedding in the QIM framework base on using multi-quantization vectors. However, in spite of using multi-quantization vectors to increase the embedded data rate, there is a trade-off between the stego-cover perceptual fidelity and the error rate performance of the hidden message detection. Mean Square Error (MSE) between the cover-work and the stego-cover increases in case of low correlation between them due to the embedding. That because, the divergence between the quantization vectors increases due to the quantization multiplicity. To minimize the MSE we should minimize this divergence, but this can lead to a more convergence between the modulated quantized coefficients. This can return an increasing in the Bit Error Rate (BER) while detecting the hidden message due to increasing the correlation between the modulated coefficients.

As a solution to this problem we can increase the stego-cover coefficients power. But increasing the power leads to a lower perceptual quality again, that because the steganography based covert channel is a power-limited communication system. The power limit is due to the requirement of imperceptible altering effect.

Another effective and practical solution is applying the digital communication theory concepts of coding, modulation, and quantization. There are many common areas between modulation and quantization theory, both areas mostly depend on signal space concepts. In that respect, a structured scheme with jointed Trellis-Coded Quantization and Trellis-Coded Modulation (TCQ/TCM) [27] is proposed to generate the altered space. This eliminates the prohibitive task of high dimensional quantization design without increasing in the embedding distortion or the detection error rate. TCQ and TCM are jointed, as they use the same encoder, and hence the TCQ quantization rate equals to the symbol rate of TCM.

TCQ is exploited to generate a set of multidimensional quantizers for phases of selected audio coefficients. This can be executed by utilization of a structured multidimensional codebook with an expanded set of quantization vectors based on the notation of set partitioning. As in QIM, the data is hidden by the choice of a quantizer from the quantized cover-work space; hence it employs the concept of modulation in the communications theory. As a development to this idea, we used TCM as an efficient scheme to modulate the quantized phases of the audio signal coefficients. The secret message in a binary form which needs to be transmitted through the covert channel is coded to a codeword vector. Then, TCM is used for mapping the quantized host audio signal space to another set according to the secret message data bits, where, each codeword selects a quantization vector to quantize a selected cover-work audio signal coefficients vector.

In this paper Multi Trellis Coded Modulation (MTCM) is used to construct the multi-constellations scheme by using a multiple trellis encoders set. Every single trellis encoder represents a single constellation. Using MTCM returns a higher data rate compared with the single TCM that because the data rate is then equals the summation of the all data rates of every encoder. In addition, to improve the performance of a TCM scheme based on a given signal constellation, the number of states should be increased. However, as this number exceeds a certain value, the coding gain increases more slowly. Thus, if the signal constellation is not changed by using a higher state, the change of the constellation can be done by using the MTCM in order to increase the coding gain of a TCM scheme. The proposed technique is considered a real-time scheme that it depends mainly on trellis coding that performed based on a set of convolutional encoders. Convolutional codes are used for real-time applications, as in video or audio streams delivering, because they can be continuously encoded and decoded.

In the decoding process, error correction is essential for extracting the embedded message from the altered audio signal in low BER [28]. A distance metric is evaluated for all quantization vectors and the one with the minimum distance from the received signal identifies the embedded information. The Viterbi algorithm [29] is used in the detector to extract the hidden message bits by selecting the path with the highest correlation to the input work.

3.4 Embedding Algorithm

The proposed data hiding modulator block diagram is illustrated in Fig. 4. Assume that, the secret message data rate is, m bit/T, where T is a time interval. To embed this message in the cover-work we need a quantization vectors constellation with size of 2^m. Using trellis encoder with rate of $m/(m+1)$ doubles this constellation size to be 2^{m+1}. In the proposed technique, a set of k trellis encoders are used, and then the embedded data rate becomes mk bit/T. The resultant coding rate is $mk/(m+1)k$, hence there is k of signal constellations (one for each group of every encoder output symbols) in each transmitter interval. In case of using jointly matched modulation /quantization, the quantization rate R should equal to $(m+1)$. The quantization process is carried out under the concept of mapping by set partitioning to the asymmetric multidimensional codebook constellation controlled by the convolutional encoders outputs to change the cover-work vector to a number of near similar vectors. The modulation process is carried out to map the quantized cover-work coefficients in the stego-cover space according to the incoming message data bits over the trellis paths.

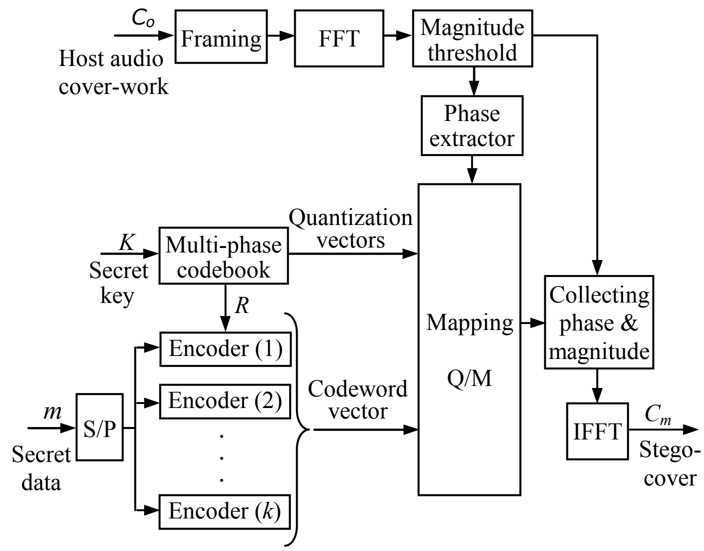


Fig. 4 MTCQ / MTCM using M-PSK Modulator.

To embed a secret message in the audio signal using the proposed technique, apply the procedures in the embedding algorithm (Fig. 5) as follows:

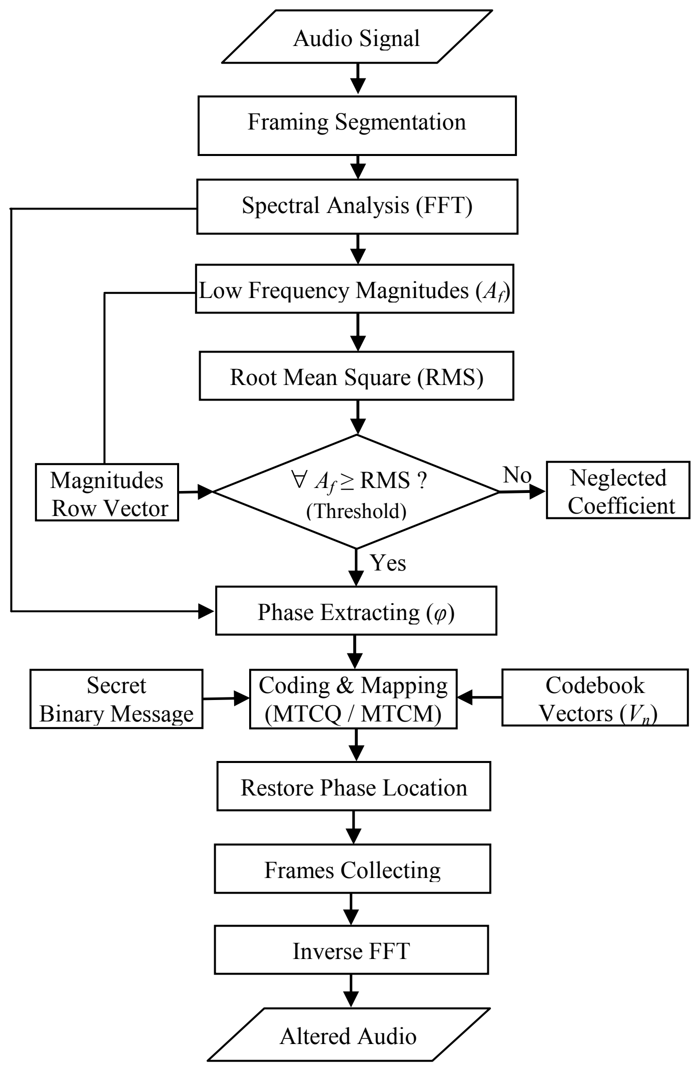


Fig. 5 Flow Diagram of the Embedding Algorithm in Audio Signals.

1) The original host audio signal stream is first segmented into frames. The segmentation process is done by a predetermined constant frame size value. This frame size satisfies the following prerequisites, which will provide important parameters such as the number of frames and the sample rate:
   i. Avoid introducing perceptible distortion: after embedding the secret data, the altered signal should be relatively similar to the original signal. Therefore, the embedding should not add any audible distortion into the host audio file.
   ii. The power of two criterions: in order to evaluate the Fourier transform faster and more efficiently, the number of samples should be a power of two.
   iii. Avoid non-overlapping frames during segmentation.
2) Calculate the magnitude and phase spectrum of each frame using DFT formula in (2) to have

$$X(f)=\sum_{l=0}^{L-1} x(l)\, e^{-j2\pi fl/L} \tag{6}$$

where, l = 0, 1,…L-1, and L is the frame size length.
3) Remove the Direct Current (DC) component F(0), which has a frequency of (0 Hz) to avoid embedding in an insignificant area. It is important to mention that any modification on the signal at this specific frequency is completely inaudible to the human auditory system which starts from 20 Hz.
4) Extract the binning magnitudes up to power considerations using the RMS threshold applied for magnitudes that in the low frequency side of the frames DFT response.
5) Extract the phase $\varphi_k=\arg[X(f)]$ of the only selected magnitudes for each frame.
6) Generate a matrix of symmetric multidimensional codebook using multi-quantization row vectors V_n given by the following formula:

$$V_n= N\delta(m-1)+n\delta-180 \tag{7}$$

where, n = {0, 1 . . . N-1} is the vector index, m = {1, 2 . . . $\lfloor 360/N\delta \rfloor$ } is the vector length, $N=2^R$ is the total number of vectors, R is the quantization rate, and δ is the vectors translation factor in degrees.
7) Generate the codeword vector of the secret message, m, using a set of, k identical trellis encoders with same polynomial function satisfies the matching between the embedding rate and the quantization rate. If the coding rate is (R-1)/R, then the embedding rate is k(R-1). The message bits are converted from serial to parallel (S/P), according to the scheme embedding rate, to be applied to the encoders.
8) Quantize the binning phase's vector using the predefined codebook multi-phase vectors.
9) Map the quantized vectors according to the codeword vector of the secret message data bits to form a unique quantized phase vector represents the embedded data bits stream.
10) Restore the elements of the quantized phase vector to their original location in the frequency domain components of each frame.
11) Modify the conjugate symmetric elements of the DFT.
12) Apply the inverse fast Fourier transform (IFFT) on each frame and collect them again to reconstruct the audio signal.

3.5 Extracting Algorithm

This process is fairly similar to the insertion process in some of procedures; however, the objective is now to extract the secret data from the audio file. To accomplish this, the audio signal is portioned uniformly into frames. Therefore, each frame is processed in the frequency domain and the secret data is finally extracted.

Figure 6 illustrates the demodulator of the proposed covert communications channel approach. The received signal contains the secret data embedded in a noisy copy of stego-cover $C_m^{'}$. Information that must be sent to the decoder as side information includes: the encoder polynomial function and its initial state (IS), the quantization rate I, and the vectors translation factor (VTF).

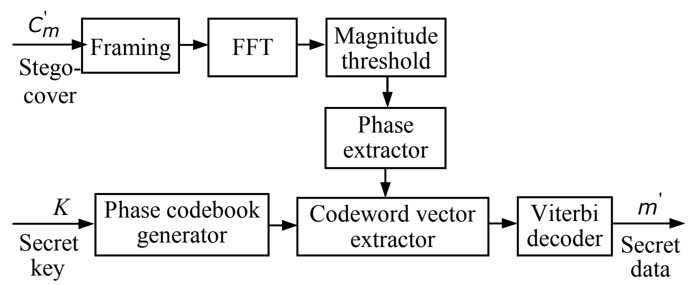


Fig. 6 MTCM using M-PSK Demodulator.

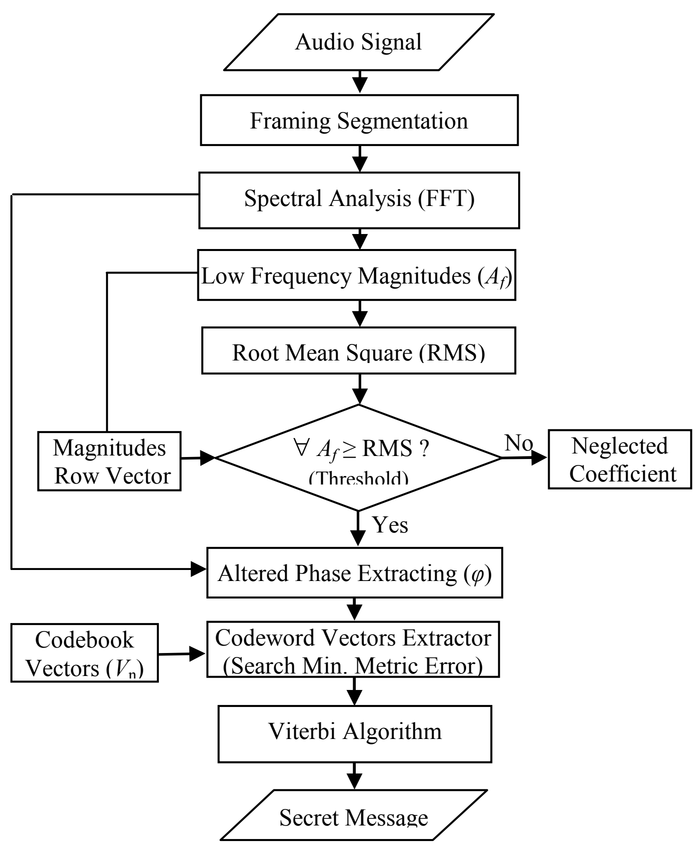


Fig. 7 Flow Diagram of the Extracting Algorithm from Audio Signals.

To extract the hidden data (assuming blind detection), apply the procedures in the extracting algorithm (Fig. 7) as follows:

1) Apply the procedures typically as in the step (1) to step (6) in the transmitter side. That include segmentation into frames, performing the frames spectrum analysis using the DFT, removing the DC component, extracting the altered phases that in the low frequency band and high-energy components of each frame, and generating the multidimensional codebook.
2) Extract the codeword vector using a search closest function. A distance metric error is evaluated between the extracted altered phases and the all codebook multi-phase vectors; the one with the minimum error distance from the received signal identifies the embedded codeword vector elements according to the assigning fashion in the set partitioning that was used in the coding step in the transmitter side.
3) Apply the Viterbi algorithm to identify the trellis path whose code vector yields the highest correlation and then extract the secret message data bits along that path.

3.6 Altered Audio Perceptual Quality Evaluation

Perceptual quality refers to the imperceptibility of embedded secret data within the host audio signal to be undetectable to a listener. This ensures that the quality of the host signal is not perceivably distorted, and does not indicate the presence or location of the covert channel. Measuring the quality of the embedded audio cover-work sequence after the embedding process with the secret data is an important to ensure the system fidelity. Audio quality evaluation can be performed using both objective and subjective measures .

Objective measures compare the original and the altered waveform and calculate a measure of the distortion between the two signals. Subjective measures involve listening tests with different quality ratings and test conditions. They include audio quality tests, where judgment is passed on the intelligibility and quality of the reconstructed audio signal. This type of measuring tests is generally performed on both standardized coders to give a benchmark and on the experimental audio coders.

Objective measures are simpler to repeatedly implement and evaluate, allowing the audio quality to be continually assessed during the embedding process [30]. However, subjective measures are always important to assess the human perception of the quality of a speech coder. In this work, Objective measures are taken into consideration to evaluate the perceptual quality of the altered audio cover-work signal. The most frequently used objective audio quality measures are the Segmental Signal-to-Noise Ratio (SEGSNR) and the Cepstral Distance (CD) measures.

- *Segmental Signal-to-Noise Ratio:*

Segmental Signal-to-Noise Ratio (SEGSNR) is a waveform distortion measure that is defined as the distortion between the original and altered audio signals. The SEGSNR is defined in time domain by determination the average of the SNR values over short segments. It is calculated over a quasi-stationary interval of an audio frame, thus, the distortion from a high-energy portion of audio will not overwhelm the distortion evaluation from a low-energy portion of audio. The audio signal is divided into segments of 10-20 ms (about 300 samples for 16 kHz signal); this gives proper weighting to low energy audio segments and therefore gives values more closely related to the subjective audio quality. Acceptable SEGSNR value is in the range of (0 to 40 dB), because outside this interval it becomes uncorrelated with subjective quality judgments. SEGSNR can be calculated using the following formula [31]:

$$\text{SEGSNR}= \frac{10}{M}\sum_{m=1}^{M} \log_{10} \frac{\sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N}\left(x_i - y_i\right)^2} \qquad (8)$$

where $m=\{1,2\ldots M\}$, M is the number of segments of length typically 15 to 20 ms, x_i is the original audio signal, y_i is the altered audio signal, $i=\{1,2\ldots N\}$, and N is the number of elements within each frame.

- *Cepstral Distance:*

Cepstral Distance (CD) is a frequency spectrum distortion measure between the original and the altered audio signals defined by using the cepstrum transformation. Cepstrum is calculated by determining the natural logarithm of magnitude of the Fourier transform, then obtaining the inverse Fourier transform of the resulting sequence. The cepstrum transform always returns complex coefficients, but for simplicity the real part is exploited and it is simply called, the cepstrum. The correlation distortion between the altered and the original audio signals can be measured by calculating the cepstral distance using the cepstrum coefficients of the two signals [31] as follows:

$$\text{CD}=\sqrt{\left(cx_1 - cy_1\right)^2 + 2\cdot\sum_{i=2}^{\infty}\left(cx_i - cy_i\right)^2} \qquad (9)$$

where cx_i is the cepstrum coefficients of the original audio signal, cy_i is the cepstrum coefficients of the altered audio signal.

4. Experimental Results

The proposed technique is tested to communicate a high data rate secretly over a covert communications channel established in various host audio files. A secret message is transferred through a using of the multiple embedding technique based on applying the multi-constellations scheme (MTCQ/MTCM) using multiple trellis encoders set.

4.1 Host Audio Media

Five standard testing audio signals are selected to carry the secret data in the proposed covert channel. Various factors have been taken into consideration in selecting theses signals including the number of signal samples and sampling frequency. In addition, these audio signals are representing different categories; for example: speech, naturals, objects, music … etc.

The technical properties of the test audio signals which had been used are listed in Table I. The signals are from Matalb V.7.9.0 (R2009b), and the audio CD issued by European Broadcasting Union (EBU) for sound quality evaluation purposes (Sound Quality Assessment Material, SQAM). This CD could be downloaded from EBU SQAM page [32].

TABLE 1: COVER-WORK TEST AUDIO STREAM SIGNALS

Audio Signal Name	Category	No. of Samples	Sampling Frequency
handel.wav	Vocal	73113	8192
speech.wav	Speech	110033	22050
thunder.wav	Naturals	36868	11025
heli.wav	Objects	101150	22050
pop.wav	Songs	430242	48000

4.2 Codebook Construction

For instance, assume using an 8-PSK scheme, the quantization rate is, $R=\log_2(8)=3$ bits/sample, hence the coding rate of each of trellis encoders is assumed to be, $(R\text{-}1)/R=2/3$. A symmetric quantization codebook of $N=2^R=8$ vectors constellation are used for every trellis encoder. These vectors represents the phase band $[-\pi \geq \varphi \leq \pi]$ of the host samples for the audio sequence cover-work represented in the frequency domain.

The multi-phase codebook vectors V_n can be calculated using the formula (7), where n is the vector index (n = 0, 1 . . . N-1) as follows:

$$
\begin{aligned}
V_0 &= 8m\delta - 8\delta - 180 \\
V_1 &= 8m\delta - 7\delta - 180 \\
V_2 &= 8m\delta - 6\delta - 180 \\
V_3 &= 8m\delta - 5\delta - 180 \\
V_4 &= 8m\delta - 4\delta - 180 \\
V_5 &= 8m\delta - 3\delta - 180 \\
V_6 &= 8m\delta - 2\delta - 180 \\
V_7 &= 8m\delta - \delta - 180
\end{aligned} \tag{10}
$$

Codebook of these eight uniform scalar quantization vectors is used for the quantization process of the phase band $[-\pi \;\; \pi]$. The vectors translation factor between every adjacent quantization vectors is assumed to be $\pi\delta/180$. In case of symmetric constellation, these vectors are equally separated by a distance equals to the VTF, it is assumed that they are in a contour of a circuit with a normalized separation distances of unity as shown in Fig. 8. The constellation vectors are set partitioned as shown in Fig. 9 based on the Ungerboeck rules with increasing the minimum relative distances between the subsets. The distances in each of the partition level can be calculated as: d_0=2sin ($\pi\delta/360$), d_1=2sin ($\pi\delta/180$), and d_2 = 2sin ($\pi\delta/90$), where, $d_0 < d_1 < d_2$.

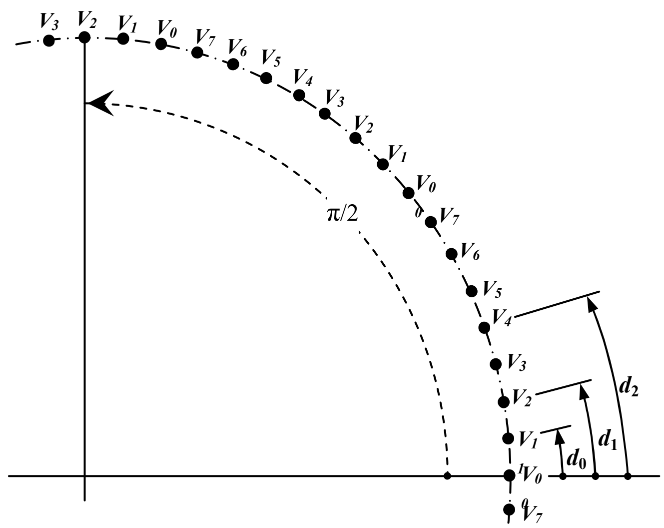


Fig. 8 A Part of Codebook Vectors Constellation.

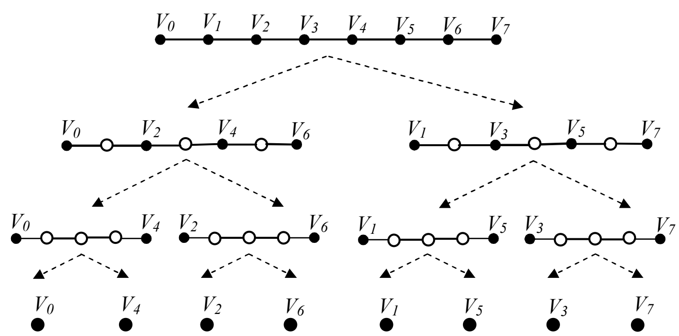


Fig. 9 A constellation set partitioning based on Ungerboeck [39].

4.3 8-PSK MTCQ/MTCM Constellations Set Partitioning

The embedding technique using 8-PSK MTCQ/MTCM in three constellations of codebook vectors (k=3) for the embedding rate 6 bit/T is discussed in details. Coding gains is calculated for two cases: the full parallel transition between the sequent 2-state trellis encoders and the no parallel transition between the sequent 64-state trellis encoders. In case of using a 3-tuple signal sets, set partitioning is carried to assign quantization vectors to the trellis branches using Dariush *et al*. method [33] as follows:

Starting by defining the 3-tuples set as:

$$
\mathrm{E}_0 = \begin{pmatrix} 0 & 0 & 0 \\ 4 & 4 & 4 \end{pmatrix} \tag{11}
$$

The Minimum Euclidean Distance (MED) between the 3-tuples elements is:

$$
d_{\min}(\mathrm{E}_0) = \sqrt{3 \cdot d^2(0,4)} = \sqrt{3\left[2\sin(5\pi/90)\right]^2} = 0.602
$$

Next, the following sets can be formed:

$$E_1 = E_0 + [0 \quad 0 \quad 4]$$
$$E_2 = E_0 + [0 \quad 4 \quad 0] \tag{12}$$
$$E_3 = E_0 + [4 \quad 0 \quad 0]$$

Each of these sets has the same MED equal to that of E_0.

The next higher partition level F_0, is formed from the union of the sets in (11) and (12) as:

$$F_0 = E_0 \cup E_1 \cup E_2 \cup E_3 = \begin{pmatrix} 0 & 0 & 0 \\ 4 & 4 & 4 \\ 0 & 0 & 4 \\ 4 & 4 & 0 \\ 0 & 4 & 0 \\ 4 & 0 & 4 \\ 4 & 0 & 0 \\ 0 & 4 & 4 \end{pmatrix} \tag{13}$$

The MED between the 3-tuples elements of this set is:

$$d_{\min}(F_0) = d(0,4) = 2\sin(5\pi/90) = 0.347$$

Next, the following sets can be formed:

$$F_1 = F_0 + [0 \quad 2 \quad 2]$$
$$F_2 = F_0 + [2 \quad 0 \quad 2] \tag{14}$$
$$F_3 = F_0 + [2 \quad 2 \quad 0]$$

Each of these sets has the same MED equal to that of F_0.

The next higher partition level A_0, is formed from the union of the sets in (13) and (14) as:

$$A_0 = F_0 \cup F_1 \cup F_2 \cup F_3$$

$$= \begin{pmatrix} 0 & 0 & 0 & & 0 & 2 & 2 & & 2 & 0 & 2 & & 2 & 2 & 0 \\ 4 & 4 & 4 & & 4 & 6 & 6 & & 6 & 4 & 6 & & 6 & 6 & 4 \\ 0 & 0 & 4 & & 0 & 2 & 6 & & 2 & 0 & 6 & & 2 & 2 & 4 \\ 4 & 4 & 0 & & 4 & 6 & 2 & & 6 & 4 & 2 & & 6 & 6 & 0 \\ 0 & 4 & 0 & & 0 & 6 & 2 & & 2 & 4 & 2 & & 2 & 6 & 0 \\ 4 & 0 & 4 & & 4 & 2 & 6 & & 6 & 0 & 6 & & 6 & 2 & 4 \\ 4 & 0 & 0 & & 4 & 2 & 2 & & 6 & 0 & 2 & & 6 & 2 & 0 \\ 0 & 4 & 4 & & 0 & 6 & 6 & & 2 & 4 & 6 & & 2 & 6 & 4 \end{pmatrix} \tag{15}$$

The MED between the 3-tuples elements of the A_0 set is:

$$d_{\min}(A_0) = \sqrt{2 \cdot d^2(0,2)} = \sqrt{2[2\sin(5\pi/180)]^2} = 0.247$$

Next, the following set can be formed:

$$A_1 = A_0 + [0 \quad 0 \quad 2]$$

$$= \begin{pmatrix} 0 & 0 & 2 & & 0 & 2 & 4 & & 2 & 0 & 4 & & 2 & 2 & 2 \\ 4 & 4 & 6 & & 4 & 6 & 0 & & 6 & 4 & 0 & & 6 & 6 & 6 \\ 0 & 0 & 6 & & 0 & 2 & 0 & & 2 & 0 & 0 & & 2 & 2 & 6 \\ 4 & 4 & 2 & & 4 & 6 & 4 & & 6 & 4 & 4 & & 6 & 6 & 2 \\ 0 & 4 & 2 & & 0 & 6 & 4 & & 2 & 4 & 4 & & 2 & 6 & 2 \\ 4 & 0 & 6 & & 4 & 2 & 0 & & 6 & 0 & 0 & & 6 & 2 & 6 \\ 4 & 0 & 2 & & 4 & 2 & 4 & & 6 & 0 & 4 & & 6 & 2 & 2 \\ 0 & 4 & 6 & & 0 & 6 & 0 & & 2 & 4 & 0 & & 2 & 6 & 6 \end{pmatrix} \tag{16}$$

The MED between the 3-tuples elements of the A_1 equal to that of A_0.

The next higher partition level A, is formed from the union of the sets in (15) and (16) as:

$$A = A_0 \cup A_1 \tag{17}$$

The MED between the 3-tuples elements is:

$$d_{\min}(A) = d(0,2) = 2\sin(5\pi/180) = 0.174$$

To generate the sets of odd-numbered vectors, the following sets can be formed:

$$B_0 = A_0 + [1 \quad 1 \quad 1]$$

$$= \begin{pmatrix} 1 & 1 & 1 & & 1 & 3 & 3 & & 3 & 1 & 3 & & 3 & 3 & 1 \\ 5 & 5 & 5 & & 5 & 7 & 7 & & 7 & 5 & 7 & & 7 & 7 & 5 \\ 1 & 1 & 5 & & 1 & 3 & 7 & & 3 & 1 & 7 & & 3 & 3 & 5 \\ 5 & 5 & 1 & & 5 & 7 & 3 & & 7 & 5 & 3 & & 7 & 7 & 1 \\ 1 & 5 & 1 & & 1 & 7 & 3 & & 3 & 5 & 3 & & 3 & 7 & 1 \\ 5 & 1 & 5 & & 5 & 3 & 7 & & 7 & 1 & 7 & & 7 & 3 & 5 \\ 5 & 1 & 1 & & 5 & 3 & 3 & & 7 & 1 & 3 & & 7 & 3 & 1 \\ 1 & 5 & 5 & & 1 & 7 & 7 & & 3 & 5 & 7 & & 3 & 7 & 5 \end{pmatrix} \tag{18}$$

The MED between the 3-tuples elements of the B_0 set is:

$$d_{\min}(B_0) = \sqrt{2 \cdot d^2(1,3)} = \sqrt{2[2\sin(5\pi/180)]^2} = 0.247$$

$$B_1 = A_1 + [1 \quad 1 \quad 1]$$

$$= \begin{pmatrix} 1 & 1 & 3 & & 1 & 3 & 5 & & 3 & 1 & 5 & & 3 & 3 & 3 \\ 5 & 5 & 7 & & 5 & 7 & 1 & & 7 & 5 & 1 & & 7 & 7 & 7 \\ 1 & 1 & 7 & & 1 & 3 & 1 & & 3 & 1 & 1 & & 3 & 3 & 7 \\ 5 & 5 & 3 & & 5 & 7 & 5 & & 7 & 5 & 5 & & 7 & 7 & 3 \\ 1 & 5 & 3 & & 1 & 7 & 5 & & 3 & 5 & 5 & & 3 & 7 & 3 \\ 5 & 1 & 7 & & 5 & 3 & 1 & & 7 & 1 & 1 & & 7 & 3 & 7 \\ 5 & 1 & 3 & & 5 & 3 & 5 & & 7 & 1 & 5 & & 7 & 3 & 3 \\ 1 & 5 & 7 & & 1 & 7 & 1 & & 3 & 5 & 1 & & 3 & 7 & 7 \end{pmatrix} \tag{19}$$

The MED between the 3-tuples elements of the B_1 equal to that of B_0.

The next higher partition level B, is formed from the union of the sets in (18) and (19) as:

$$B = B_0 \cup B_1 \tag{20}$$

The MED between the 3-tuples elements is:

$$d_{\min}(B) = d(1,3) = 2\sin(5\pi/180) = 0.174$$

- Evaluation of Coding Performance Gain

Case (1): 2-State Encoder

This represents the case of full parallel transition between the sequent trellis states. For an error event in case of all ‘‘0’’ sequence transmitted using the 2-state trellis encoders as in Fig. 10, the set partitions A_0 and A_1 in (15) and (16) are used for the first state S0, and the sets B_0 and B_1in (18) and (19) are used for the second state S1. The minimum Euclidean distance can be calculated as:

$$\begin{aligned}
\text{MED} &= \sqrt{d^2_{\min}(A_0, A_1) + d^2_{\min}(A_0, B_0)} \\
&= \sqrt{d^2(000{,}002) + d^2(000{,}111)} \\
&= \sqrt{d^2(0{,}2) + 3 \cdot d^2(0{,}1)} \\
&= \sqrt{[2\sin(5\pi/180)]^2 + 3[2\sin(5\pi/360)]^2} = 0.231
\end{aligned}$$

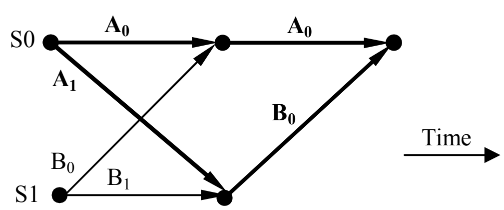


Fig. 10 Mapping of Codebook Constellations Vectors (2-state, k=3).

Case (2): 64-State Encoder

This represents the case of no parallel transition between the sequent trellis states, which returns the maximum value of MED. For an error event in case of all ''0'' sequence transmitted using the 64-state trellis encoders as in Fig. 11. The machine output states are assignments and the set partitions are given as follows:

$$\begin{bmatrix} S0 \\ S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \\ S7 \\ S8 \\ S9 \\ S10 \\ S11 \\ \vdots \\ S62 \\ S63 \end{bmatrix} = \begin{bmatrix} a_0\ a_1\ a_2 \ldots\ldots\ldots\ldots\ldots\ldots a_{63} \\ b_0\ b_1\ b_2 \ldots\ldots\ldots\ldots\ldots\ldots b_{63} \\ a_1\ a_2\ a_3 \ldots\ldots\ldots\ldots\ldots a_{63}\ a_0 \\ b_1\ b_2\ b_3 \ldots\ldots\ldots\ldots\ldots b_{63}\ b_0 \\ a_2\ a_3\ a_4 \ldots\ldots\ldots\ldots a_{63}\ a_0\ a_1 \\ b_2\ b_3\ b_4 \ldots\ldots\ldots\ldots b_{63}\ b_0\ b_1 \\ a_3\ a_4\ a_5 \ldots\ldots\ldots a_{63}\ a_0\ a_1\ a_2 \\ b_3\ b_4\ b_5 \ldots\ldots\ldots b_{63}\ b_0\ b_1\ b_2 \\ a_4\ a_5\ a_6 \ldots\ldots a_{63}\ a_0\ a_1\ a_2\ a_3 \\ b_4\ b_5\ b_6 \ldots\ldots b_{63}\ b_0\ b_1\ b_2\ b_3 \\ a_5\ a_6\ a_7 \ldots a_{63}\ a_0\ a_1\ a_2 \ldots\ldots a_4 \\ b_5\ b_6\ b_7 \ldots b_{63}\ b_0\ b_1\ b_2 \ldots\ldots b_4 \\ \vdots \\ a_{31}\ a_{32}\ a_{33} \ldots a_{63}\ a_0\ a_1\ a_2 \ldots a_{30} \\ b_{31}\ b_{32}\ b_{33} \ldots b_{63}\ b_0\ b_1\ b_2 \ldots b_{30} \end{bmatrix} \tag{21}$$

where, $a_0 = (0\ 0\ 0)$, $a_1 = (4\ 4\ 4)$, $a_2 = (0\ 0\ 4)$, $a_3 = (4\ 4\ 0)$,
$a_4 = (0\ 4\ 0)$, $a_5 = (4\ 0\ 4)$, $a_6 = (4\ 0\ 0)$, $a_7 = (0\ 4\ 4)$,
$a_8 = (0\ 2\ 2)$, $a_9 = (4\ 6\ 6)$, $a_{10} = (0\ 2\ 6)$, $a_{11} = (4\ 6\ 2)$,
$a_{12} = (0\ 6\ 2)$, $a_{13} = (4\ 2\ 6)$, $a_{14} = (4\ 2\ 2)$, $a_{15} = (0\ 6\ 6)$,
$a_{16} = (2\ 0\ 2)$, $a_{17} = (6\ 4\ 6)$, $a_{18} = (2\ 0\ 6)$, $a_{19} = (6\ 4\ 2)$,
$a_{20} = (2\ 4\ 2)$, $a_{21} = (6\ 0\ 6)$, $a_{22} = (6\ 0\ 2)$, $a_{23} = (2\ 4\ 6)$,
$a_{24} = (2\ 2\ 0)$, $a_{25} = (6\ 6\ 4)$, $a_{26} = (2\ 2\ 4)$, $a_{27} = (6\ 6\ 0)$,
$a_{28} = (2\ 6\ 0)$, $a_{29} = (6\ 2\ 4)$, $a_{30} = (6\ 2\ 0)$, $a_{31} = (2\ 6\ 4)$,
$a_{32} = (0\ 0\ 2)$, $a_{33} = (4\ 4\ 6)$, $a_{34} = (0\ 0\ 6)$, $a_{35} = (4\ 4\ 2)$,
$a_{36} = (0\ 4\ 2)$, $a_{37} = (4\ 0\ 6)$, $a_{38} = (4\ 0\ 2)$, $a_{39} = (0\ 4\ 6)$,
$a_{40} = (0\ 2\ 4)$, $a_{41} = (4\ 6\ 0)$, $a_{42} = (0\ 2\ 0)$, $a_{43} = (4\ 6\ 4)$,
$a_{44} = (0\ 6\ 4)$, $a_{45} = (4\ 2\ 0)$, $a_{46} = (4\ 2\ 4)$, $a_{47} = (0\ 6\ 0)$,
$a_{48} = (2\ 0\ 4)$, $a_{49} = (6\ 4\ 0)$, $a_{50} = (2\ 0\ 0)$, $a_{51} = (6\ 4\ 4)$,
$a_{52} = (2\ 4\ 4)$, $a_{53} = (6\ 0\ 0)$, $a_{54} = (6\ 0\ 4)$, $a_{55} = (2\ 4\ 0)$,
$a_{56} = (2\ 2\ 2)$, $a_{57} = (6\ 6\ 6)$, $a_{58} = (2\ 2\ 6)$, $a_{59} = (6\ 6\ 2)$,
$a_{60} = (2\ 6\ 2)$, $a_{61} = (6\ 2\ 6)$, $a_{62} = (6\ 2\ 2)$, $a_{63} = (2\ 6\ 6)$,
$b_0 = (1\ 1\ 1)$, $b_1 = (5\ 5\ 5)$, $b_2 = (1\ 1\ 5)$, $b_3 = (5\ 5\ 1)$,
$b_4 = (1\ 5\ 1)$, $b_5 = (5\ 1\ 5)$, $b_6 = (5\ 1\ 1)$, $b_7 = (1\ 5\ 5)$,
$b_8 = (1\ 3\ 3)$, $b_9 = (5\ 7\ 7)$, $b_{10} = (1\ 3\ 7)$, $b_{11} = (5\ 7\ 3)$,
$b_{12} = (1\ 7\ 3)$, $b_{13} = (5\ 3\ 7)$, $b_{14} = (5\ 3\ 3)$, $b_{15} = (1\ 7\ 7)$,
$b_{16} = (3\ 1\ 3)$, $b_{17} = (7\ 5\ 7)$, $b_{18} = (3\ 1\ 7)$, $b_{19} = (7\ 5\ 3)$,
$b_{20} = (3\ 5\ 3)$, $b_{21} = (7\ 1\ 7)$, $b_{22} = (7\ 1\ 3)$, $b_{23} = (3\ 5\ 7)$,
$b_{24} = (3\ 3\ 1)$, $b_{25} = (7\ 7\ 5)$, $b_{26} = (3\ 3\ 5)$, $b_{27} = (7\ 7\ 1)$,
$b_{28} = (3\ 7\ 1)$, $b_{29} = (7\ 3\ 5)$, $b_{30} = (7\ 3\ 1)$, $b_{31} = (3\ 7\ 5)$,
$b_{32} = (1\ 1\ 3)$, $b_{33} = (5\ 5\ 7)$, $b_{34} = (1\ 1\ 7)$, $b_{35} = (5\ 5\ 3)$,
$b_{36} = (1\ 5\ 3)$, $b_{37} = (5\ 1\ 7)$, $b_{38} = (5\ 1\ 3)$, $b_{39} = (1\ 5\ 7)$,
$b_{40} = (1\ 3\ 5)$, $b_{41} = (5\ 7\ 1)$, $b_{42} = (1\ 3\ 1)$, $b_{43} = (5\ 7\ 5)$,
$b_{44} = (1\ 7\ 5)$, $b_{45} = (5\ 3\ 1)$, $b_{46} = (5\ 3\ 5)$, $b_{47} = (1\ 7\ 1)$,
$b_{48} = (3\ 1\ 5)$, $b_{49} = (7\ 5\ 1)$, $b_{50} = (3\ 1\ 1)$, $b_{51} = (7\ 5\ 5)$,
$b_{52} = (3\ 5\ 5)$, $b_{53} = (7\ 1\ 1)$, $b_{54} = (7\ 1\ 5)$, $b_{55} = (3\ 5\ 1)$,
$b_{56} = (3\ 3\ 3)$, $b_{57} = (7\ 7\ 7)$, $b_{58} = (3\ 3\ 7)$, $b_{59} = (7\ 7\ 3)$,
$b_{60} = (3\ 7\ 3)$, $b_{61} = (7\ 3\ 7)$, $b_{62} = (7\ 3\ 3)$, $b_{63} = (3\ 7\ 7)$.

The minimum Euclidean distance is:

$$\begin{aligned}
\text{MED} &= \sqrt{d^2_{\min}(a_0, a_{32}) + d^2_{\min}(a_0, a_{16})} \\
&= \sqrt{d^2(000{,}002) + d^2(000{,}202)} \\
&= \sqrt{d^2(0{,}2) + 2 \cdot d^2(0{,}2)} \\
&= \sqrt{[2\sin(5\pi/180)]^2 + 2[2\sin(5\pi/180)]^2} = 0.302
\end{aligned}$$

where, $a_0 = (0\ 0\ 0)$, $a_{32} = (0\ 0\ 2)$ and $a_{16} = (2\ 0\ 2)$.

This provides a performance improvement gain over the cases of 2-state equals to 2.33 dB.

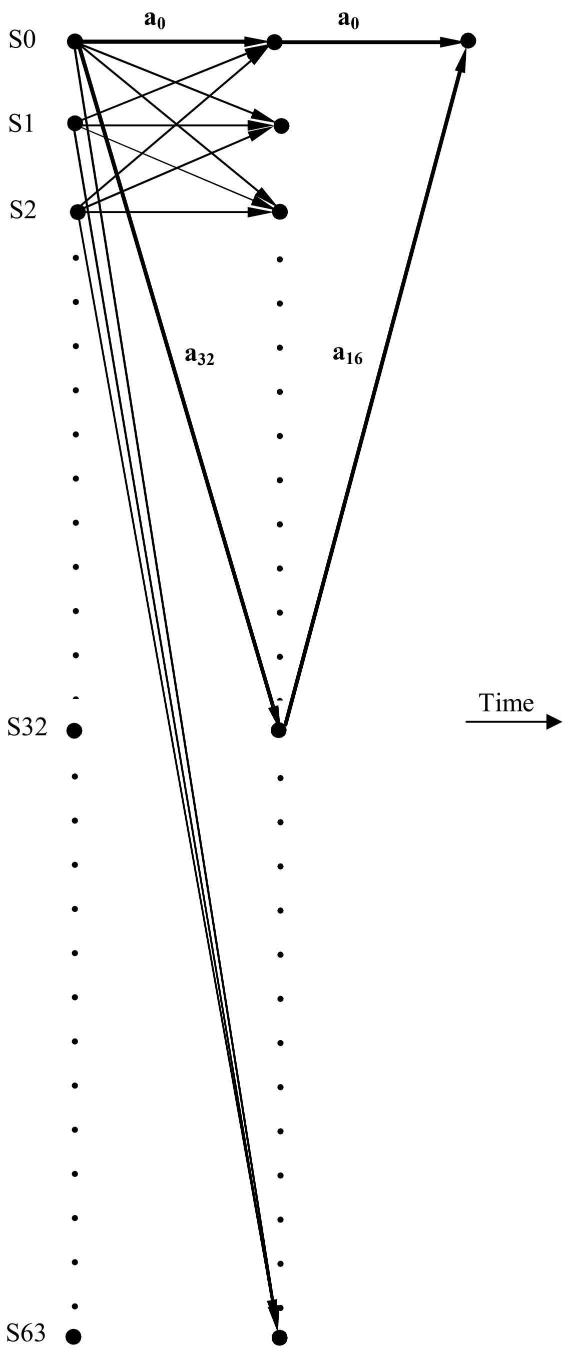


Fig. 11 Mapping of Codebook Constellations Vectors (64-state, k= 3).

4.4 Design Parameters based Channel Performance

Experimentally, the proposed technique is used to transfer binary data message over a covert channel for secret communications established over audio signals. The performance of the covert channel is analyzed for: the cover-work perceptual quality due to embedding, and the secret message detection error performance due to channel noise.

1) Cover-work Perceptual Quality

The very important precaution is working within the imperceptible fidelity region for the human auditory system. The most effective covert channel design parameters in the proposed technique are: the required quantization rate (R), the size length of segmented frames (L), and the vectors translation factor (δ). In case of using a fixed quantization rate ($R = 3$; i.e. 3 bit/symbol), the testing audio cover-works are experimentally analyzed to identify the optimal design parameters (L, and δ) to keep the embedding system within a predefined perceptual fidelity measure quantities (CD and SEGSNR) between the original and altered audio signals.

Two objects should be taken into consideration during the channel design. First: increasing the δ as much as possible, this returns a less intersymbol interference between modulated samples at the detection side, consequently a robust communication system against channel noise. Contrary, increasing the δ returns a distortion in magnitudes of the altered samples due the divergence between the altered and unaltered points in the signal constellation. Second: increasing the frame length returns a low distortion between the altered and unaltered signals due to the less number of modified signal samples, but it is required to minimize the frame length as much as possible to have an increasing in system capacity.

As a main required feature of covert channels is the acceptable fidelity, the stego-work perceptual quality is taken as a master guide to identify the design parameters that keep the system to be in the fidelity region all the time. Experimentally, the acceptable value of CD is found to be (CD < 0.9), and the SEGSNR is found to be (SEGSNR > 20 dB) to have a good balance between robustness, capacity, and fidelity in designing a relative optimal data hiding system using the proposed technique.

The proposed technique is tested using the host audio streams listed in Table 1 and a program written in Matlab V.7.9.0 (R2009b). These files are used as covert channels using the proposed 8-PSK MTCQ/MTCM embedding in frequency domain. The following tests are carried to identify the optimal design parameters that return acceptable covert channel fidelity.

- *Test: using (handel.wav) Signal*

The audio signal (handel.wav) is analysed to identify the design parameters of the covert communications channel. Tables 2 and 3 illustrate the values of CD and SEGSNR respectively for multi frame sizes at variant values of vectors translation factor. Figures 12 and 13 illustrate the previous relations plotted in curves. It is obvious that values $\delta = 4$, and L = 64 can return SEGSNR >20, and CD < 0.9 which are keeping the system to be in the fidelity region. Figure 14 illustrates the levels of the original and altered signals in (a), and (b) respectively, it is noted that there is no considerable perceptual distortion between the two signals as shown in plotting them at the same axis in (c).

TABLE 2: CD OF ALTERED "handel" AUDIO SIGNAL

Frame Length (L)	Cepstral Distance (CD)			
	VTF: δ = 4	δ = 6	δ = 8	δ = 10
32	0.3007	0.4017	0.4917	0.577
64	0.2404	0.3294	0.4015	0.4812
128	0.1932	0.258	0.3274	0.3821
256	0.1354	0.2032	0.2439	0.2828
512	0.1033	0.1447	0.1764	0.2146

TABLE 3: SEGSNR OF ALTERED "handel" AUDIO SIGNAL

Frame Length (L)	Cepstral Distance (CD)			
	VTF: δ = 4	δ = 6	δ = 8	δ = 10
32	19.011	15.527	12.927	10.93
64	20.926	17.229	14.677	12.566
128	21.931	18.001	15.546	13.531
256	23.449	19.779	17.866	15.648
512	25.509	21.203	18.792	17.856

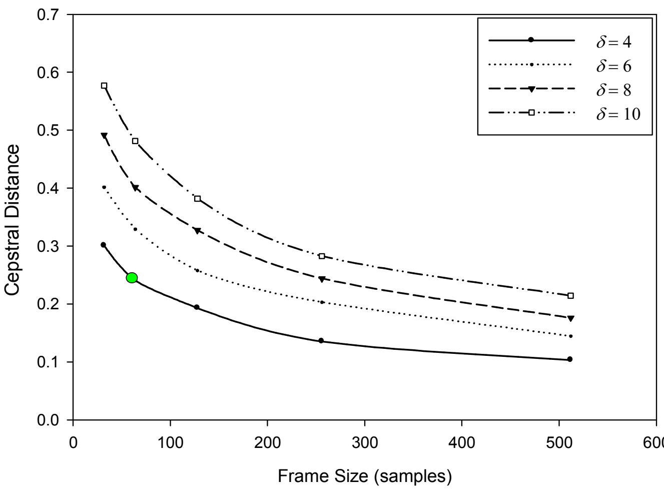


Fig. 12 CD vs. Frame Size of Altered "handel" Stream at Variant VTFs.

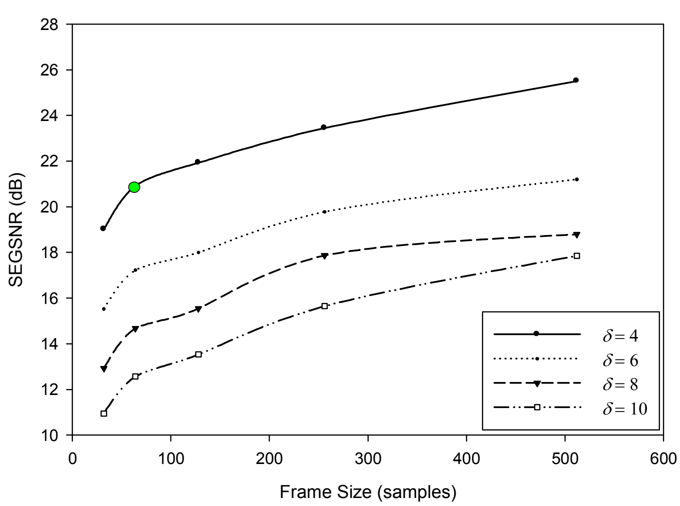
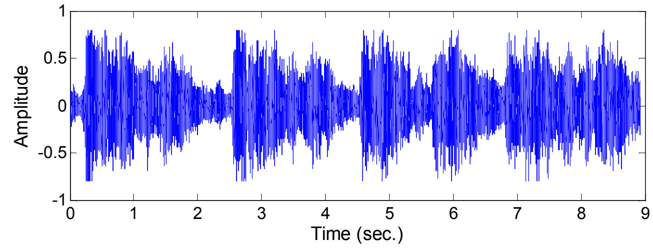
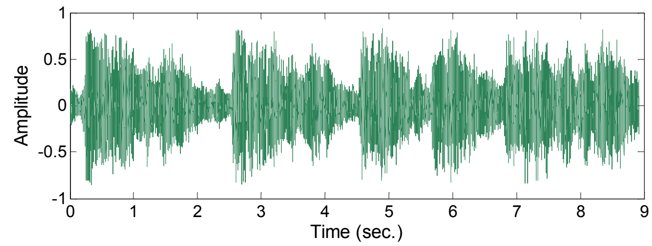


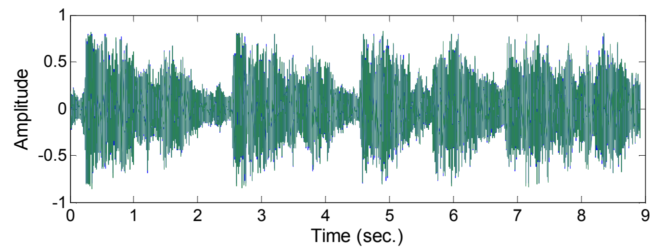
Fig. 13 SEGSNR vs. Frame Size of Altered "handel" Stream at Variant VTFs.



(a) Original Signal.



(b) Altered Signal.



(c) Original and Altered Signals.

Fig. 14 Original and Altered "handel" Audio Signal Levels.

Table 4, illustrates the experimental results obtained in testing the proposed technique for the rest of testing audio signals. Design parameters (L and VTF) required for keeping perceptual quality quantities (CD and SEGSNR) within the imperceptible regions are tabulated.

TABLE 4: OPTIMAL CD & SEGSNR OF ALTERED TESTED AUDIO SIGNALS

Audio File Name	Frame Length (L)	VTF (δ)	CD	SEGSNR
handel	64	4	0.2404	20.926
speech	256	4	0.6437	23.0215
thunder	256	6	0.8589	20.4515
heli	32	4	0.2479	22.3784
pop	32	6	0.9408	21.8462

2) Secret Message Error Performance

The error performance of secret message sent over the proposed covert communications channel established in cover-work audio stream signals is analyzed experimentally. Table 5 illustrates the experimental results for the BER in the detection of secret message data bits transferred in the frequency domain of a covert channel audio over the AWGN. The results are compared with the traditional two types of embedding (single and double) in the QIM framework. Figure 15 illustrates the previous relations plotted in curves. Compared with the related techniques error performance, the multiple embedding in the proposed scheme achieves a better performance in BER and data transmission rate over the single and double embedding schemes. This performance is due to maximizing the minimum distance between the modulated quantized coefficients. At BER $\approx 10^{-5}$, the gain performance improving is about 10 dB and 15 dB in case of no parallel transitions between the sequent trellis states in using coding multiplicity, k=3 over the cases of single embedding and double embedding respectively.

TABLE 5: BER OF COVERT CHANNEL IN "handel" OVER AWGN COMPARED WITH TRADITIONAL EMBEDDING IN QIM

SNR (dB)	BER		
	k = 3, 64-state	Single	Double
5	0.4868	0.4842	0.5053
10	0.4864	0.5044	0.5004
15	0.4373	0.4956	0.4956
20	0.257	0.4623	0.5206
25	0.1088	0.4361	0.4658
30	0.0408	0.2382	0.3884
35	0.0053	0.1217	0.2434
40	0	0.0324	0.0976

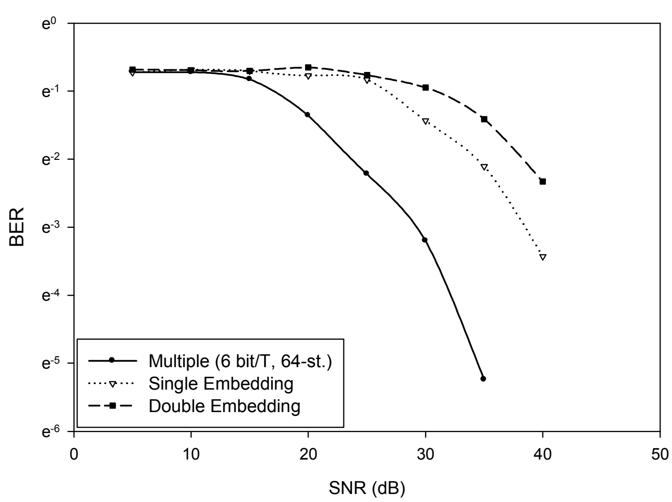


Fig. 15 Error Performance of the Proposed Covert Channel in Frequency Domain of "handel" Audio Stream over AWGN Compared with Single and Double Embedding.

5. CONCLUSION

A proposed technique for a structured covert channel for audio signals based on using M-PSK MTCQ/MTCM is developed to increase the data rate for hidden communications. This scheme is based on multi-dimensional quantization for the cover-work space using trellis coding techniques, and then it is modulated by the secret message data bits. This achieves good separation between the modulated quantized stego-cover elements, hence, low correlations with one another. The proposed technique has advantages over previous related techniques in terms of complexity, and decoding time, that it uses only one identical trellis encoder for quantization and modulation, besides using the well known simple decoding Viterbi algorithm. In addition, the error performance of the proposed covert channel in detection the hidden message over AWGN returns low BER compared with the related schemes that use single or double embedding.

REFERENCES

[1] J. M. Bahi, J. F. Couchot, and C. Guyeux, "Steganography: a Class of Algorithms having Secure Properties," *IEEE Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (*IIH-MSP*), Oct. 2011.

[2] R. M. Goudar, P. N. Patil, A. G. Meshram, S. M. Yewale, and A. V. Fegade, "Secure Data Transmission by using Steganography," *International Institute for Science, Technology and Education* (*IISTE*), Vol. 2, No.1, 2012.

[3] R. Amirtharajan, K. Nathella, and J. Harish, "Info Hide – A Cluster Cover Approach," *International Journal of Computer Applications,* Vol. 3, No.5, Jun. 2010.

[4] N. D. Jambhekar, "Steganography: To Preserve the Document Security," *Golden Resrach Thoughts*, Vol.1, Issue VI, Dec. 2011, pp.1-4.

[5] B. Alrouh1, "Towards Secure Web Services: Performance Analysis, Decision Making and Steganography Approaches," *Ph.D. dissertation, Brunel University, UK*, Oct. 2011.

[6] B. Alrouh1, A. Almohammad, and G. Ghinea, "Information Hiding in SOAP Messages: A Steganographic Method for Web Services," *International Journal for Information Security Research* (*IJISR*), Vol. 1, Issues 1/2, Mar. /Jun. 2011.

[7] H. B. Kekre, A. Athawale, S. Rao, and U. Athawale, "Information Hiding in Audio Signals," *International Journal of Computer Applications,* Vol. 7, No. 9, Oct. 2010.

[8] P. Dutta, D. Bhattacharyya, and T.Kim, "Data Hiding in Audio Signal: A Review," *International Journal of Database Theory and Application,* Vol. 2, No. 2, Jun. 2009.

[9] M. L. Mat Kiah, B. B. Zaidan, A. A. Zaidan, A. M. Ahmed, and S. H. Al-bakr, "A Review of Audio based Steganography and Digital Watermarking," *International Journal of the Physical Sciences*, Vol. 6(16), Aug. 2011, pp. 3837-3850.

[10] P. Dutta, D. Bhattacharyya, and T.Kim, "Data Hiding in Audio Signal: A Review," *International Journal of Database Theory and Application,* Vol. 2, No. 2, Jun. 2009.

[11] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for Data Hiding," *IBM Systems Journal*, Vol. 35, Nos. 3 and 4, 1996.

[12] R. Ravula, "Audio Watermarking using Transformation Techniques," *MSc. dissertation, Louisiana State University*, Oct. 2010.

[13] N. Cvejic, "Algorithms for Audio Watermarking and Steganography," *Ph.D. dissertation, University of Oulu*, 2004.

[14] Z. K. AL-Ani, A. A. Zaidan, B. B. Zaidan, and H. O. Alanazi, "Overview: Main Fundamentals for Steganography," *Journal of Computing*, Vol. 2, Issue 3, Mar. 2010.

[15] P. Gonzalez, F. Balado, F. Martin, and J. Hernandez, "Performance Analysis of Existing and New Methods for Data Hiding with Known-host Information in Additive Channels," *IEEE Transactions on Signal Processing*, Vol. 51, Issue 4, Mar. 2003.

[16] A.Selem, D. Darwish, and G. Darwish, " A High-Rate Hidden Communications Channel," International Journal of Digital Image Processing, Vol.3, No.3, March 2011, pp. 186-198

[17] M.L. Miller, G.J. Doerr, and I.J. Cox, "Applying Informed Coding and Embedding to Design Robust, High Capacity Watermark," *IEEE Transactions on Image Processing*, Vol. 13, 2004, pp. 792 – 807.

[18] B. Chen, and G. W. Wornell, "Quantization index modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding," *IEEE Transactions on Information Theory*, Vol. 47, May 2001, pp.1423-1443.

[19] A. Sarkar, and B. S. Manjunath, "Double Embedding in the Quantization Index Modulation Framework," *IEEE International Conference on Image Processing (ICIP)*, Nov. 2009, pp. 3653-3656.

[20] W. Barkouti, S. Nasri, and A. Cherif, "Audio Watermarking Based On the PSK Modulation," *International Journal on Computer Science and Engineering* (*IJCSE*), Vol. 3, No. 9, Sept. 2011.

[21] M. Arnold, P. G. Baum, and W. Voeßing, "A Phase Modulation Audio Watermarking Technique," *11th International Workshop on Information Hiding* (*IH 2009*), Darmstadt, Germany, Jun. 2009.

[22] E. Nezhadarya, Z. J. Wang, and R. K. Ward, "A new data hiding method using angle quantization index modulation in gradient domain," *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), May 2011, pp. 2440 – 2443.

[23] S. Yang, W. Tan, Y. Chen, and W. Ma, "Quantization-Based Digital Audio Watermarking in Discrete Fourier Transform Domain," *Journal of Multimedia*, Vol. 5, No. 2, Apr. 2010.

[24] M. Mandal, and A. Asif, "Continuous and Discrete Time Signals and Systems," Cambridge University Press, 2007.

[25] N. K. Govindaraju, B. Lloyd, Y. Dotsenko, B. Smith, and J. Manferdelli, "High performance discrete Fourier transforms on graphics processors," *International Conference for Computing, Networking, Storage and Analysis*, 2008, pp. 1–12.

[26] C. E. Shannon, "Communication in the Presence of Noise," *Proc. IRE*, Vol. 37, Jan. 1949, pp. 10-12.

[27] H. A. Aksu, and M. Salehi, "Joint Optimization of TCQ-TCM Systems," *IEEE Trans. on Comm.*, Vol. 44, May 1996, pp. 529–533.

[28] A. Chadha, S. Gangundi, R. Goel, H. Dave, and M. M. Roja, "Audio Watermarking with Error Correction," *International Journal of Advanced Computer Science and Applications* (*IJACSA*), Vol. 2, No. 9, 2011.

[29] G. D. Forney, "The Viterbi Algorithm," *Proceedings of the IEEE*, Vol. 61, Mar. 1973, pp. 268–278.

[30] L. Hanzo, F. C. A. Somerville, and J. Woodard, "*Voice and Audio Compression for Wireless Communications*," IEEE Press, Second Edition, 2007.

[31] Y. Hu, and P. C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 16, No. 1, Jan. 2008.

[32] http://tech.ebu.ch/publications/sqamcd

[33] D. Divsalar, and M. K. Simon, "Multiple Trellis Coded Modulation (MTCM)," *IEEE Trans. on Comm.*, Vol. 36, Apr. 1988, pp. 410-419.

Ashraf Seleym received the B.SC. degree in Air Defense Science from Air Defense College, Egypt, in 1993, and the B.Sc. degree in Electrical Communications Engineering with honors degree from Alexandria University, Egypt in 1994. He received the M.Sc. degree in Electronics and Communications Engineering from Mansoura University, Egypt, in 2003. Now, he is a Ph.D. student in IT, Faculty of Computers and Information, Cairo University, Egypt. His research interests include multidimensional coding, networks security, and hidden communications.

Dina Darwish received the B.Sc. in 2004 and the M.Sc. in 2006 with honors degree from Arab Academy for Science and Technology, Egypt. She received the Ph.D. degree from Cairo University, Egypt, 2009. Her main interests include communications systems, computer networks, internet technology, and multimedia systems. She is assistant professor of communications and computer networks, International Academy for Engineering and Media Science (IAEMS), Egypt, since September 2009.