International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:2, No:3, 2008

A Tool for Audio Quality Evaluation Under Hostile Environment

Akhil Kumar Arya, Jagdeep Singh Lather,Lillie Dewan

Abstract**—**In this paper is to evaluate audio and speech quality with the help of Digital Audio Watermarking Technique under the different types of attacks (signal impairments) like Gaussian Noise, Compression Error and Jittering Effect. Further attacks are considered as Hostile Environment. Audio and Speech Quality Evaluation is an important research topic. The traditional way for speech quality evaluation is using subjective tests. They are reliable, but very expensive, time consuming, and cannot be used in certain applications such as online monitoring. Objective models, based on human perception, were developed to predict the results of subjective tests. The existing objective methods require either the original speech or complicated computation model, which makes some applications of quality evaluation impossible.

Keywords**—**Digital Watermarking, DCT, Speech Quality, Attacks.

I. Introduction

WITH the rapid growth of network distributions of digital media contents, there is an urgent need for copyright protection against piracy. The embedding of digital watermarks into multimedia content has been proposed to tackle this kind of problem. Audio watermarks are special information signals embedded into digital audio. These signals are extracted by detection mechanisms. Robustness and imperceptibility are important requirements of watermarking.
The evaluation of audio and speech quality is a very important field in current multimedia era. According to specific practice of long standing, the only way to measure the quality of an audio signal was through the use of subjective quality evaluation [1].In this test ten or more person were involved to listen a live or recorded conversation and assign a rating to it. Participants listened to the audio sequences and were asked to report using five-point scale: (5: imperceptible, 4: perceptible but not annoying, 3: slightly annoying, 2: annoying, 1: very annoying). The arithmetic mean of the collection of these was taken for quality of audio which is called the mean opinion score (*MOS*). This has been the most reliable method of speech quality assessment but it is highly unsuitable for online monitoring applications and is also very expensive and time consuming.

Due to these reasons, objective methods have been developed in recent years, classified into two categories: signal-based methods and parameters- based methods [5]. The signal-based methods use the reference and degraded signals as the input to the measurement; on the contrary, the parameters-based methods predict the speech quality through a computational model instead of using real measurement.

Objective methods can also be classified as intrusive and nonintrusive ones. Intrusive method takes both the original and the degraded speeches as the input. Non-intrusive methods only require the degraded speech. It is more challenging to design a nonintrusive method because no original speech information could be used during the quality evaluation. Recently, several nonintrusive speech quality evaluation methods have been proposed.

In the following, we will briefly introduce subjective, signal based objective and parameters-based objective speech quality evaluation methods.

1) *Signal Based Methods*: Signal-based methods use the reference and distorted signals as input. The two signals are compared based on some perceptual model and the predictions of subjective test results are generated. In order to achieve an estimate of perceived quality, a measurement should employ as much understanding of human perception and human judgments as possible. The common idea behind perceptual quality measurement is to mimic the situation of a subjective test [1].

2)*Parameter Based Methods*: Besides perceptual measurement, some other parameters based methods, such as Gaussian mixture models, artificial neural networks and E-models, have also been developed for audio and speech quality assessment.

In Gaussian mixture models (GMMs), a large pool of feature measurements is extracted and created from the distortion surface between the original speech signal and the degraded speech signal. Good features are then chosen [4]. The joint density of these selected features is modeled with the subjective MOS as a Gaussian mixture. Finally, using this model, the least squares estimate of the subjective MOS value is derived. This model outperforms the PESQ in root mean square errors but the improvement in correlation between the subjective MOS and predicted MOS is small.

In, artificial neural networks (ANNs) have been employed to assess audio quality in packet networks with the concern of several distortion parameters on transmitted audio, such as, arrival jitter, end-to-end delay, sampling rate and the number

Akhil Kumar Arya was M.Tech Student with the Electrical Engineering Department, National Institute of Technology Kurukshetra Haryana India (e-mail:akhilarya@gmail.com)
Jagdeep Singh Lather, is with the Electrical Engineering Department, National Institute of Technology Kurukshetra Haryana India (e-mail:jslather@nitkkr.ac.in)
Dr.Lillie Dewan is with the Electrical Engineering Department, National Institute of Technology Kurukshetra Haryana India (e-mail: l_dewan@nitkkr.ac.in)

of bits per sample of codec algorithm, echo, crosstalk effect, etc. To build a neural network, the most effective quality-affecting parameters of the network are chosen. Subjective experiments are then conducted for establishing the relation between parameter values and MOS scores. Once a stable neural network configuration is obtained, the trained ANN will take the given parameter values and correspondingly compute the subjective MOS quality score.

Digital watermarking technology has been around for more than ten years, which has been used in copyright protection, content authentication, copy control, broadcast monitoring, etc. In this dissertation, we propose a new application of digital watermarking: speech quality evaluation. The basis of this method is that the carefully embedded watermark in a speech will suffer the same distortion as the speech does. The proposed method needs neither the original speech, nor a training database. Furthermore, without the complicated signal processing on both the original and the degraded speech signals, the implementation of the proposed quality evaluation is very fast. In addition to speech quality evaluation, this objective method can also evaluate the quality of audio signals. This work includes audio and speech quality evaluation method using digital watermarking. Our algorithm evaluates the speech quality without the need of reference speech or any computational model. The watermark is embedded in the discrete cosine domain or temporal domain of a speech signal by using quantization technique. This algorithm can evaluate perceptual quality of audio and speech that is distorted by Gaussian Noise, Compression Error and Jittering Effect. Implementation and analysis of Audio Watermarking scheme using Modified Patchwork Algorithm (MPA). MPA is statistical technique for audio watermarking [17]. It inserts watermarks in frequency domain.

II IT'S A COMMUNICATION PROBLEM

Digital watermarking can be viewed as a communication problem. Information to be sent to the receiver is encoded into a signal called the watermark, which is then embedded into the media signal, referred to as *the cover signal*, to form the watermarked data [2]. This watermarked data is sent to the receiver through a channel, denoted as *the watermark channel*, where it might be further processed or even replaced by some other data. This process is also denoted as *the attack*. In the context of robust watermarking, the goal of an attacker is to impair or even remove the embedded watermark information without impairing the cover signal. Conversely, the aim of the defender is to design the transmitter in such a way that the watermark is still there, as long as the attack results in received signals of sufficient quality. This so-called robust watermarking was first proposed for multimedia copyright protection and then for many other possible applications.

Reliable communication was proven by Shannon to be theoretically possible providing the information rate does not exceed a threshold known as channel capacity [32]. With an idealized assumption regarding the form of the noise n corrupting a watermark, information theory can be used to derive the rules to decide about the strength of the watermark required and location of the watermark.

Let us write,

$$x_i + n_i = y_i \quad ; \quad 1 \le i \le N \tag{1}$$

where x_i is one element of a watermark vector of length N, n_i is an element of a noise vector and y_i is an element of a watermark distorted by image processing noise. Assuming the noise is additive, white, Gaussian:

$$p(y_i \mid x_i) = p(n_i) = (1/\sigma\sqrt{2\pi})\ \exp\left[-(y_i - x_i)^2 / 2\sigma\right] \tag{2}$$

Assuming that the n_i are uncorrelated and that

$$p(y_1, y_2 \cdots\cdots y_N \mid x_1, x_2 \cdots\cdots x_N) = \coprod_i^N p(y_i \mid x_i) \tag{3}$$

Channel capacity may be defined as

$$C = \max_{p(x)} I(X_i Y) \tag{4}$$

where the watermark probability density function p (x) is chosen to maximize the average mutual information I(X;Y).According to Proakis the capacity is maximized with respect to distribution P(x) if

$$p(x_i) = (1\ /(\gamma\sqrt{2}))\exp[-x_i^2/(2\gamma^2)] \tag{5}$$

which is a zero mean Gaussian density with variance γ^2. In this case,

$$I_{max} = \frac{1}{2} N \log_2(1 + \gamma^2/\sigma^2) \tag{6}$$

I extreme conditions, in which case $\sigma^2 >> \gamma^2$ which implies

$$\ln(1 + \gamma^2/\sigma^2) \cong \gamma^2/\sigma^2 \tag{7}$$

Substituting eqn. (2.7) into eqn. (2.6) we obtain the following condition for reliable communication:

$$\gamma^2/\sigma^2 > (2\ln 2)J/N \tag{8}$$

where the N is the number of sites used to hide watermark information bit and J is the information rate. It should be noted that the noise power can be considerably greater then the signal power and, in theory at least, the message may still be transmitted reliably.

The strategy for communicating the watermark is now clear. Because a watermark should be imperceptible the signal to noise ratio (SNR), is severely limited. Reliable communication can only be assured by increasing bandwidth B to compensate poor SNR. Hence in the case of watermarking the maximum number N of suitable transform domain coefficients should be exploited for hiding information in the signal. Watermarking may be considered as being an application of spread spectrum communications. The Shannon limit may be approached by applying error control codes. Robust error correction techniques can be employed if necessary. To answer the second part of the question "Where to embed ?" We again take recourse to information

theory concepts. Let us assume that the signal may be considered as a collection of parallel uncorrelated Gaussian channels which satisfy (1) above with the constraint that the total watermark energy is limited:

$$\sum_i^N \gamma_i^2 <= E \quad (9)$$

Using (2) and assuming the noise variances are not necessarily the same in each channel, Gallger[17] shows that the capacity is,

$$C = 1/2 \sum_i^N \log_2 (1 + \gamma_i^2 / \sigma_i^2) \quad (10)$$

where σ_i^2 is the variance of the noise corrupting the watermark and γ_i^2 is the average power of the watermark signal in the i^{th} channel. This is a mere general form of (6). Capacity is achieved when

$$\sigma_i^2 + \gamma_i^2 = T_h$$
$$\text{if} \quad \sigma_i^2 < T_h \quad (11)$$

where the threshold T_h is chosen to maximize the sum on the left-hand side of (9) and thus maximize the energy of the watermark. This result shows clearly that the watermark should be placed in those areas where local noise variance σ_i^2 is smaller than threshold T_h and not at all in those areas where local noise variance exceeds the threshold.

It should be noted that the simple analysis presented here assumes that the noise corruption suffered by the watermarks a result of common forms of signal processing is Gaussian. This is not an accurate assumption to make in many cases. However, the Gaussian assumption is not a bad choice given that the aim is to derive the rules and heuristics that apply in general to a number of fundamentally different signal processing scenarios.

III SPREAD SPECTRUMCOMMUNICATIONS

It is clear that the watermark should not be placed in perceptually insignificant regions of the signal (image or audio) or its spectrum since many common signals and geometric processes attack these components[30]. For example, a watermark placed in the high frequency spectrum of a signal can be easily eliminated with little degradation to the signal by any process that directly or indirectly performs low pass filtering. The problem then becomes how to insert a watermark into the most perceptually significant regions of a spectrum without such alternations becoming noticeable. Clearly, any spectral coefficient may be altered, provided such modification is small. However, very small changes are very susceptible to noise.

To solve this problem, the frequency domain of the image or sound is viewed as a communication channel, and correspondingly, the watermark is viewed as a signal that is transmitted through it. Attacks and unintentional signal distortions are thus treated as noise that the immersed signal must immune to.

Thus, the watermarking can be considered as an application of spread spectrum communications. In spread spectrum communication, one transmits a narrow band signal over a much larger bandwidth such that the signal energy present in any single frequency is imperceptible. Similarly, the watermark is spread over very many frequency bins so that the energy in any one bin is very small and certainly undetectable. Nevertheless, because the watermark verification process knows the location and context of the watermark, it is possible to concentrate these many weak signals with a high signal to noise ratio (SNR). However, to considerably destroy such a watermark would require noise of high amplitude to be added to all frequency bins. Spreading of the watermark throughout the spectrum of a signal ensures a large measure of security against unintentional or intentional attack. First the spatial location of the watermark is not obvious. Furthermore, frequency regions should be selected in a fashion that ensures severe degradation of the original data following any attack on the watermark.

IV INTRUSION ON WATERMARKS

A watermarked signal is likely to be subjected to certain manipulations, some intentional such as compression and transmission noise and some intentional such as cropping, filtering, etc [1],[2].

A. *Lossy Compression*: Many compression schemes like JPEG and MPEG can potentially degrade the data's quality through irretrievable loss of data.

B. *Geometric Distortions*: Geometric distortions are specific to images videos and include such operations as rotation, translation, scaling and cropping.

C. *Common Signal Processing Operations*: They include the followings.

a. D/A conversion
b. A/D conversion
c. Resampling
d. Requantization
e. Dithering distortion
f. Recompression
g. Linear filtering such as high pass and low pass filtering
h. Non-linear filtering such as median filtering
i. Color reduction
j. exchange of pixels
k. Addition of a constant offset to the pixel values
l. Addition of Gaussian and Non Gaussian noise
m. Jittering effect
n. Compression error

Other intentional attacks:

a. Printing and Rescanning.
b. Watermarking of watermarked host signal (re-watermarking).
c. Forgery: A number of authorized recipients of the image should not be able to collude to form a copy of watermarked image with the valid embedded watermark of a person not in the group with an intention.

V AUDIO OR SPEECH QUALITY EVALUATION METHOD

Figure1 illustrates the proposed speech quality evaluation method using digital watermarking. As shown in Figure1, the proposed method consists of three parts: 1) watermark embedding; 2) watermark extraction; and 3) quality

evaluation. The watermark embedding and extraction are quantization based. To provide best performance, we make the quantization scale adaptive to a speech signal, which will be discussed in detail. The optimized quantization scale is used for both watermarking embedding and extraction.

For evaluating the speech quality after Compression Error, Gaussian Noise Addition and Jittering Effect we embed the watermark in the discrete cosine transform (DCT) coefficients of the speech [15],[18],[27]. The quality evaluation algorithm is the same to all the distortions (Compression Error, Gaussian Noise Addition, and Jittering Effect). As the watermark will undergo the same distortion as the speech does, we can evaluate the quality of the speech having undergone distortions by evaluating the percentage of correctly extracted watermark bits (PCEW). Furthermore, from Fig. 1, it can be seen that the proposed method does not need the original speech signal for quality evaluation.

Audio or Speech Quality Evaluation:

After watermark extraction, the PCEW is calculated by comparing the extracted watermark with the original one using the following relation.

$$PCEW = \frac{1}{N}\sum_{j=1}^{N} w(j) \oplus w*(j)$$

where is $w(j)$ the original watermark, $w*(j)$ is the extracted watermark, N is the length of the watermark, and $\oplus$ is the exclusive- OR operator. The PCEW value lies between 0 and 1.

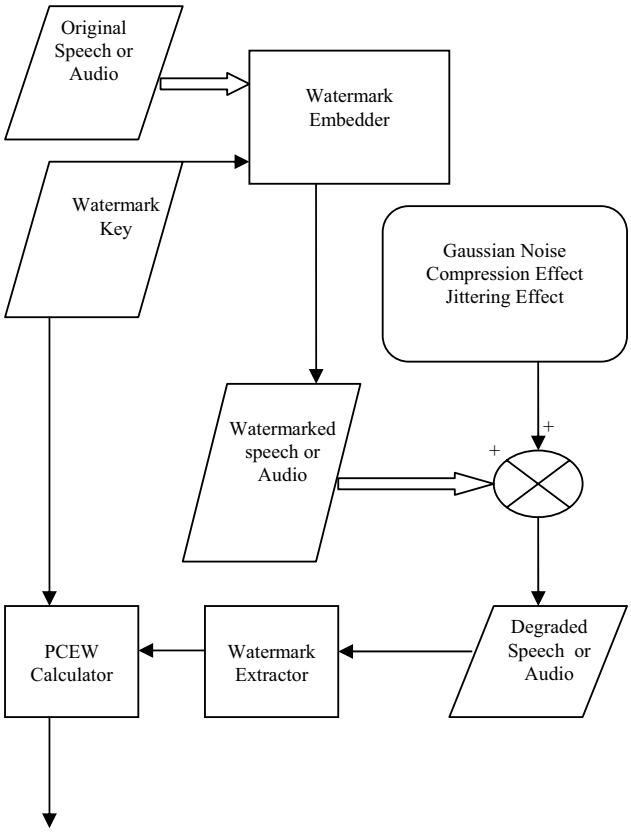

Fig. 1 Speech quality evaluation using digital watermarking.

VI IMPLEMENTATION STRATEGIES

In this section, about the strategies which are used in the implementation of the audio or speech quality evaluation are discussed in detailed. The watermark embedding and extraction are Modified Patchwork Algorithm based. It is statistical technique for audio watermarking. It inserts watermarks in frequency domain. For evaluating the speech quality after Compression Error, Gaussian Noise Addition and Jittering Effect we embed the watermark in the discrete cosine transform (DCT) coefficients of the speech. As the watermark will undergo the same distortion as the speech does, we can evaluate the quality of the speech having undergone distortions by evaluating the percentage of correctly extracted watermark bits (PCEW). Proposed method does not need the original speech signal for quality evaluation.

A. Encoder:

Let us denote a signal by I, a signature by S = {s1, s2 … } the watermarked signal by I'. E is an encoder function, it takes an signal I and a signature S, and it generates a new signal which is called watermarked signal I', i.e.

$$E(I, S) = I' \qquad (12)$$

It should be noted that the signature S may be dependent on signal I. In such cases, the encoding process described by (12) still holds.

Embedding Steps are summarized as follows:

a. Firstly, random numbers are generated from random number generator, which is initiated by secret key. The secret key is seed of random number generator. Then make an index set $I = \{I_1 \ldots\ldots\ldots\ldots I_{2n}\}$ from these random numbers. These random numbers are selected from $[K_1, K_2]$, $1 \le K_1 < K_2 \le N$. Where K_1 and K_2 are the range from which random numbers are selected. For example K_1 =500 and K_2=1000 then random numbers are selected from the range varies from (500-1000). It means all random numbers have values between 500 and 1000. The choice of K_1 and K_2 are crucial step in embedding the watermark because these values control the robustness and inaudibility of watermark. Where N is size of block in which DCT is applied and one bit code is embedded. Two index sets, I^0 and I^1, are needed to denote watermark bits 0 and 1, respectively. For example, in order to embed 0 the index set I^0 is used and to embed 1 index set I^1 is used. The whole audio signal is divided into number of blocks. Two index sets are needed in order to embed one bit in one block i.e. either 0 or 1. In similar way four index sets such as I^{00}, I^{01}, I^{10} and I^{11} are needed in order to embed 00 or 01or 10 or 11 respectively in one block i.e. two bits in one block. Therefore a Distinct multiple index sets are used to designate multiple bits of code information in just one block. In the thesis, only one bit is embedded in each block.

b. Take the Discrete Cosine Transform (DCT) of whole audio signal. After that divide the DCT coefficients of audio signal into different number of blocks. In each block one bit is embedded i.e. either 0 or 1. Let F= $\{F_1 \ldots\ldots\ldots\ldots F_N\}$ be the DCT coefficients whose subscript denote frequency range from lowest to highest frequencies and N is size of block in which one bit is embedded. Define $A = \{a_1 \ldots\ldots\ldots\ldots a_n\}$

as the subset of F corresponds to the first n elements of the index set I^0 or I^1 according to the embedded code with similar definition for B for the last n elements, that is , $a_i = F_{I_i}$, and $b_i = F_{I_{n+1}}$, for i = 1,2…….n. i.e.

take the first element of index set I , let it be 980 then take the 980th element of F, let it be .043 then first element of A be .043 then take the second element of index set I , repeat the above procedure till all the elements of A are selected. Suppose 50 random numbers are generated, it means index set I contains 50 elements i.e. 2n=50 and n = 25. It means both A and B contains 25 elements each. Take 26th element of index set I, let it be 775 then take the 775th element of F, let it be 0.098 then the first element of B be 0.098.After that take 27th element of index set I in order to get the second element of B. Repeat the above procedure till all the elements of B are selected. A and B are sets used for embedding. If random numbers are selected from index se I^0 then 0 is embedded otherwise 1 is embedded.

c. Calculate the sample means $\bar{a} = n^{-1} \sum_{i=1}^{n} a_i$

and $\bar{b} = n^{-1} \sum_{i=1}^{n} b_i$ the Pooled Sample Standard

error (S):

$$S = \sqrt{\frac{\sum_{i=1}^{n} (a_i - \bar{a})^2 + \sum_{i=1}^{n} (b_i - \bar{b})^2}{n(n-1)}}$$

d. The Embedding function presented below introduces a location shift change

$$a_i^* = a_i + sign\ (\bar{a} - \bar{b}) \sqrt{C}\ \frac{S}{2}$$

$$b_i^* = b_i + sign\ (\bar{a} - \bar{b}) \sqrt{C}\ \frac{S}{2}$$

where C is a constant and "sign" is a sign function. Choose the value of C always greater than threshold. This function make large value set larger and small value set smaller so that distance between two sample means is always bigger than $d = \sqrt{C}\ S$ Where d is distance between two sample means.

e. Finally, replace the selected elements a_i and b_i by a_i^* and b_i^*, respectively, i.e. the 1st element of A which is a_1 is replaced by a_i^* this process goes on till all the elements of A is replaced. Similar, is the case for B, then place the replaced elements of A and B in F at the same position from where it was selected. For example, let the first element of A be 0.0045 i.e. a_1 = 0.0043 and a_1^* = 0.0097 then a_1 = 0.0097. The first element of A was taken from the 980th element of F; therefore replace the 980th element of F by 0.0097. This process goes on till all the elements of A and B are replaced which further replace the corresponding elements of F. After that, apply the inverse DCT.

.

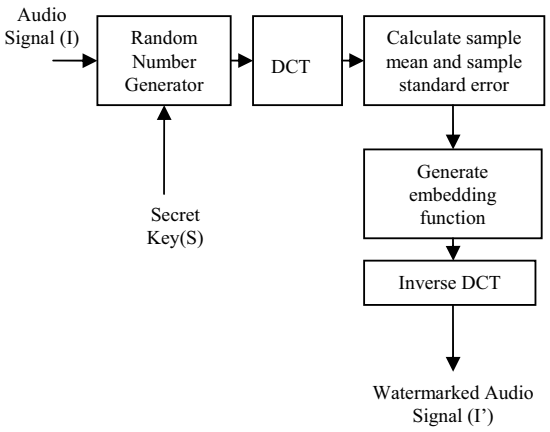

Fig. 2. Block Diagram of Encoder

B. Decoder:

A decoder function takes an signal J (J can be a watermarked or un- watermarked signal, and possibly corrupted) whose ownership is to be determined and recovers a signature S' from the signal.

Decoding Steps are summarized as follows:

Since the embedding function introduces relative changes of two sets in location, a natural test statistic which is used to decide whether or not the watermark is embedded should concern the distance between the means of A and B.

b. First, generate the same random numbers as generated during embedding by using the same secret key of random number generator as used during embedding. Then made the same index sets I^0 or I^1 as used during encoding process.

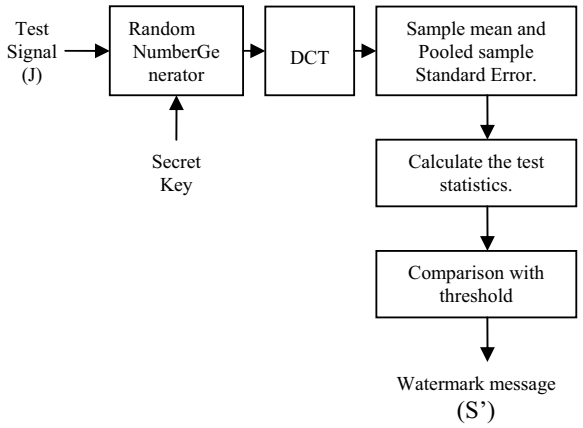

Fig. 3. Block diagram of decoder

c. Take the Discrete Cosine Transform (DCT) of watermarked audio signal. Divide the DCT coefficients into same number of blocks as made during embedding process. Then obtain the subsets A_1 and B_1 from F= $\{F_1 \ldots\ldots\ldots F_N\}$ and compute the sample means and the pooled sample standard errors. For example, obtain the

subsets $A_0 = \{a_{01}.....a_{0n}\}$ and $B_0 = \{b_{01}.....b_{0n}\}$ from index set I^0, $A_1 = \{a_{11}.....a_{1n}\}$ and $B_1 = \{b_{11}.....b_{1n}\}$ from the index set I^1, all from F = {F_1……..F_N} and compute the sample means $\overline{a_0}, \overline{b_0}, \overline{a_1}$ and $\overline{b_1}$ and pooled sample standard errors S_0 and S_1 similar to encoding process.

d. Calculate the test statistics

$$T_0^2 = \frac{(\overline{a}_0 - \overline{b}_0)^2}{S_0^2} \quad \text{and}$$

$$T_1^2 = \frac{(\overline{a}_1 - \overline{b}_1)^2}{S_1^2}$$

e. Define T^2 as the larger value obtained from two statistics.

f. In order to decide whether watermark was embedded or not, threshold M is compared with test statistics T^2. Watermark is embedded only and only if test statistics T^2 is greater than threshold i.e. T^2>M. Now in order to decide whether 0 was embedded or 1, test statistics T_0^2 and T_1^2 is compared. If $T_0^2 > T_1^2$ and T^2>M then bit 0 was embedded otherwise bit 1 was embedded. Choose the value of M in such a way that whenever watermark is embedded then the value of test statistics T^2 should always be greater than threshold M and vice versa.

VII PARAMETERS CHOSEN TO IMPLEMENT

Following parameters are used to evaluate the performance of the Audio Watermarking algorithm. These are discussed as below:

A. Random Numbers (n):

Random numbers are generated from random number generator which is initiated by secret key. This Secret key is the seed of random number generator. Random numbers are used to select the samples of audio signal in which we have to embed the watermark message. In other words random numbers are used to select the frequency of DCT.

B. Watermark Message:

It is message embedded into the host audio signal. Signal to noise ratio of watermarked signal depends upon the length of watermark message. Larger the length of watermark message smaller is the signal to noise ratio. Five different lengths of watermark message which are 5, 10, 15, 20 and 25 bits are taken. Then total number of bits embedded into the audio signal is given by:

B = (length of Watermark Message) * ((Random numbers/2)).

C. K:

It is the range from which random numbers are selected. Three different values of K have been taken for testing the MPA. Three different values are: 500-1000, 100-1000 and 350-700. The choice of K is crucial step in embedding the watermark because these values control the robustness and inaudibility of watermark.

.

VIII EXPERIMENTAL RESULTS AND EVALUATIONS

The Algorithm has been tested on standard audio signal which is a recorded sound. The input audio signal is passed to the encoder module. The encoder generates a watermarked audio signal at its output. The watermarked audio signal is then passed through the decoder module in order to get an embedded watermark message. Performance of the Algorithm is evaluated by obtaining Signal to Noise Ratio (SNR) for various parameters.

The audio signal 'Test1.wav' and 'Test2.wav' has been used for evaluating the performance of the Algorithm. The audio signals are 16 bits stereo signal with sampling frequency of 44.1 KHz. The duration of first signal is 19 sec. Another audio signal which has been used for testing the performance is 'Test2.wav' signal which is recorded mp3 song of duration 26 sec with sampling frequency of 44.1 kHz. It is also a 16 bit stereo signal. For testing, three different values of K and have been taken. These values are given as:
K1= 500-100, K2= 100- 1000, K3= 350-700.

For watermark messages, alphabets are chosen and each alphabet is given 5 bits code. For example if 'AKHIL' is to be embedded in host audio signal 'Test1.wav' then it means 25 bits are to be embedded in host audio signal since watermark message 'AKHIL' contains 10 alphabets i.e. A,K,H,I,L and for each alphabet 5 bits are needed. In this case length of watermark message is 25 bits.

A. Implementation Results for Test1.wav audio signal:

For different values of, K, Random Numbers and different lengths of watermark message, the values of SNR calculated for 'Test1.wav' audio signal are shown in Fig4 to Fig.10:

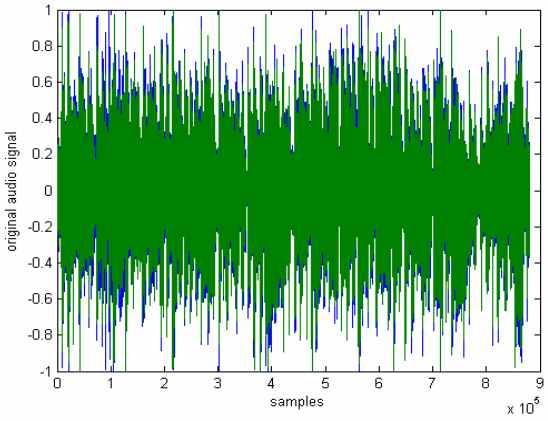

Fig. 4. Original Audio Signal

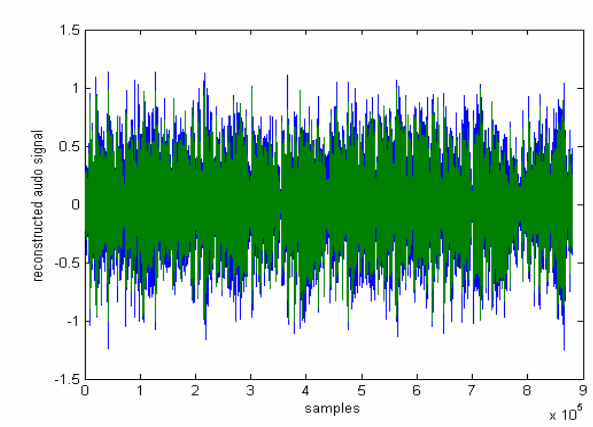

Fig. 5. Reconstructed Watermarked Audio Signal

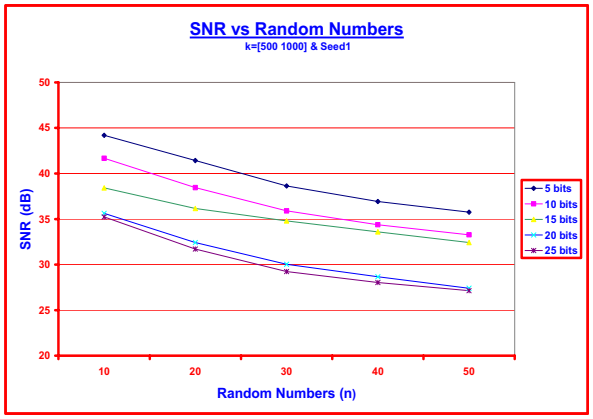

Fig. 6. (a) Graphical Representations of SNR for different values of Random Numbers for k= [500 1000]

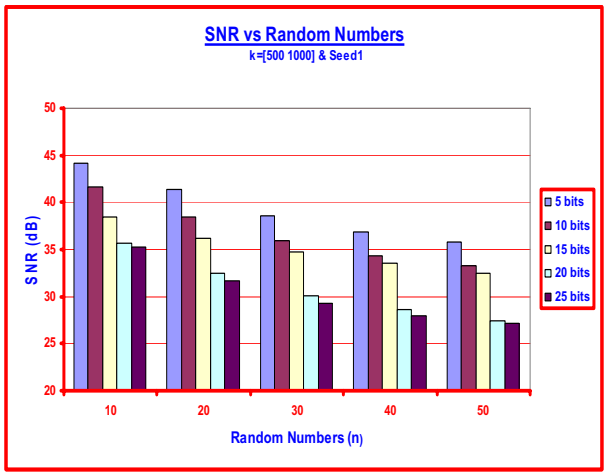

Fig. 6. (b) Graphical Representations of SNR for different values of Random Numbers for k= [500 1000]

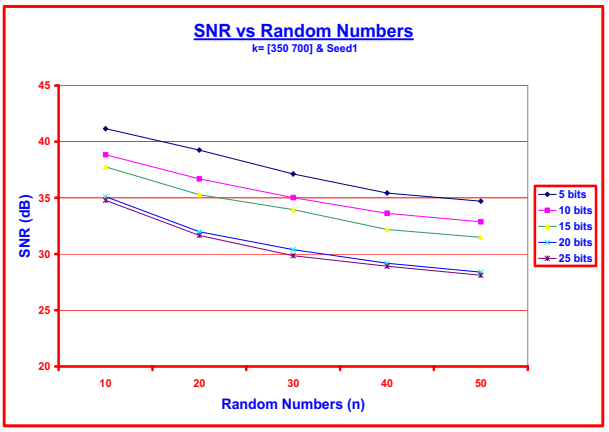

Fig. 7. (a) Graphical Representations of SNR vs. Random Numbers k= [350 700]

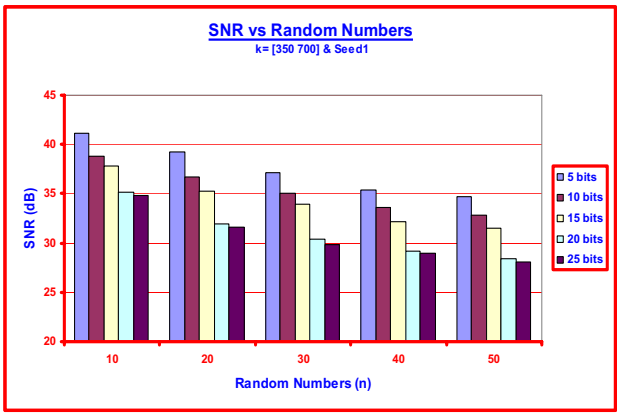

Fig.7. (b) Graphical Representations of SNR vs. Random Numbers for k= [350 700]

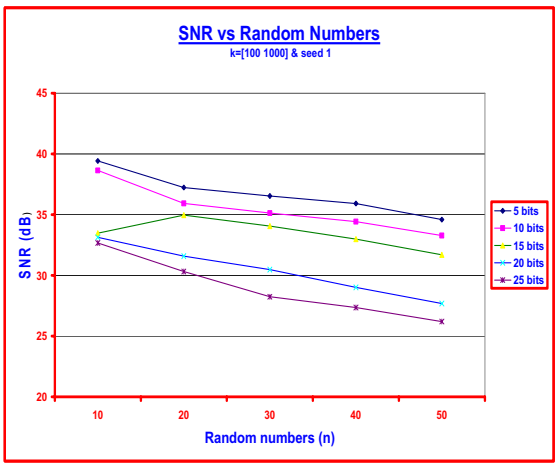

Fig. 8. (a) Graphical Representations of SNR vs. Random Numbers for k= [100 1000]

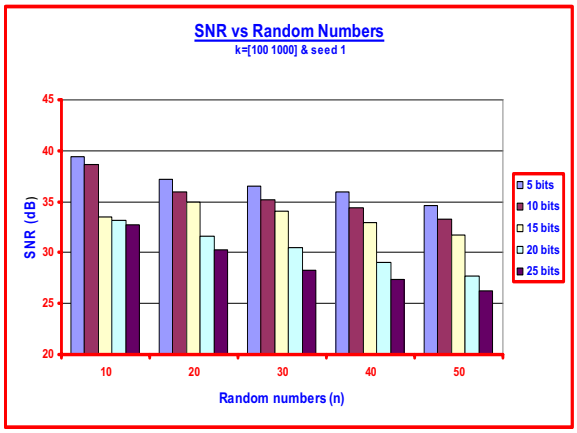

Fig. 8. (b) Graphical representations of SNR vs. Random Numbers for k= [100 1000]

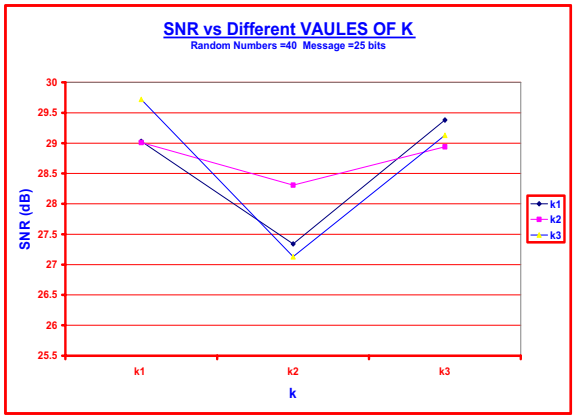

Fig. 9. Graphical representations SNR Random numbers =40 and length of watermark message = 25 bits

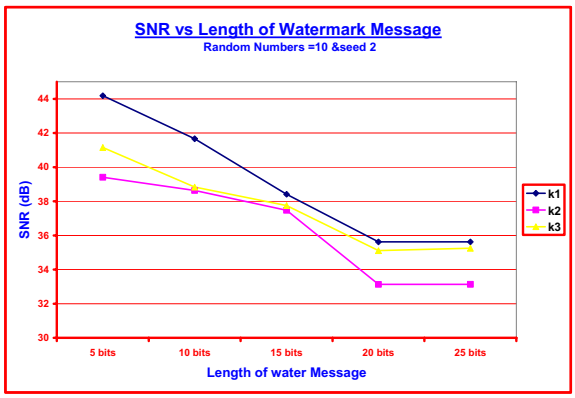

Fig. 10. SNR vs. Length of Watermark Message for Random Numbers =10

B. Implementation Results for Test2.wav audio signal:

For different values of, K, Random Numbers and different lengths of watermark message, the values of SNR calculated for 'Test2.wav' audio signal are shown in Fig.11 to Fig.14:

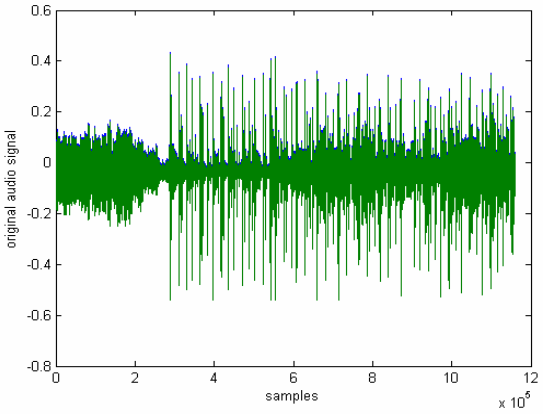

Fig. 11. Original Audio Signal

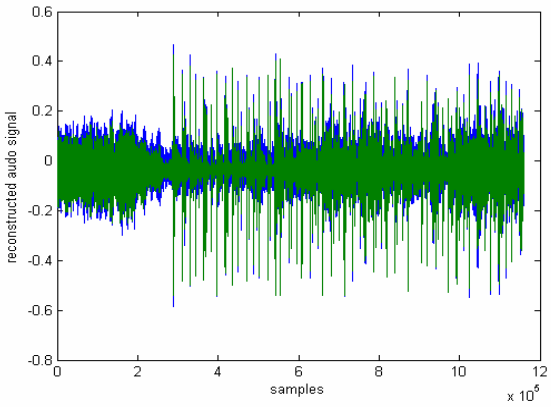

Fig. 12. Reconstructed Watermarked Audio Signal

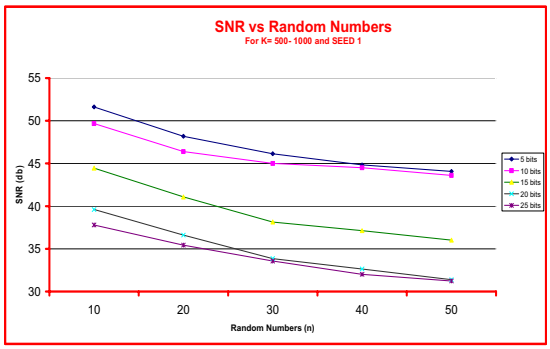

Fig. 13. (a): SNR vs. Random Numbers k= [500 1000]

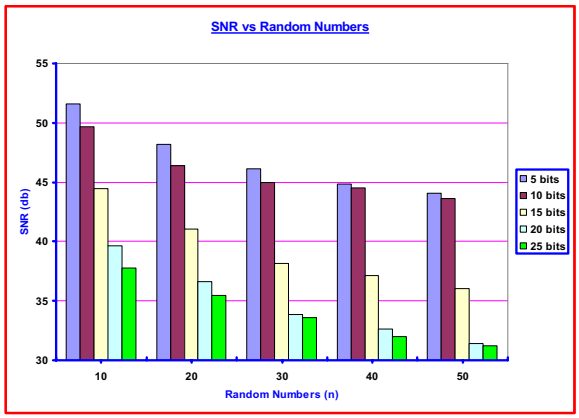

Fig. 13. (b): Graphical representation of SNR values for different Random Numbers

The above plot Fig.13(a) and (b)shows that the SNR values obtained for 'Test2.wav' audio signal is much better than that of 'fanna.wav' audio signal because of larger duration of 'Test2.wav' audio signal. The maximum value of SNR is 51.62 db

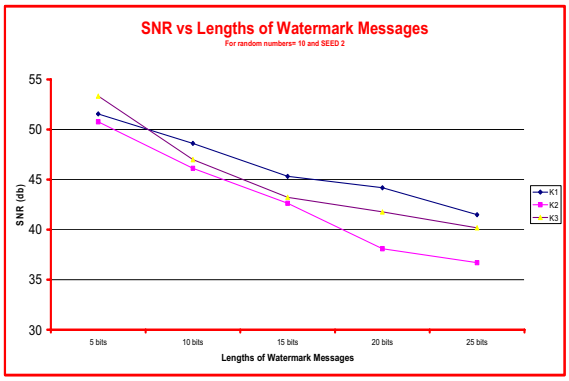

Fig. 14. Graphical representation of SNR values for different values of Length of Watermark Message

C. Hostile Environment Results for Test1.wav:

Here results of Test1.wav watermarked audio signal in the presence of hostile environment are shown. The hostile environment consists of Compression Error, Gaussian Noise Addition and Jittering Effect. Fig. 15 shows the Compression Error which is the result of difference of original signal and watermarked signal. Fig. 16 shows Gaussian Noise Addition on the watermarked Signal, the order of noise is 6.5×10^{-3}. Fig. 17 and Fig. 18 shows Jittering Effect on the watermarked Signal. Jittering Effect is negligible due to this in the original it is not possible to see the Jittering Effect normally. So to examine the Jittering Effect, response is magnified in sample range as from 0.20 to 0.21.The blue color signal is original watermarked signal and green color signal is after Jittering Effect.

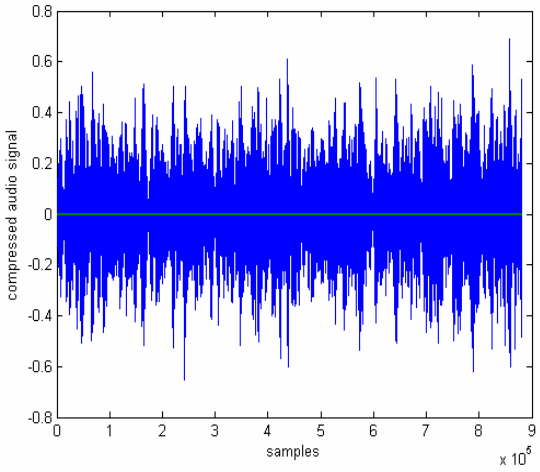

Fig. 15. Compressed Audio Signal

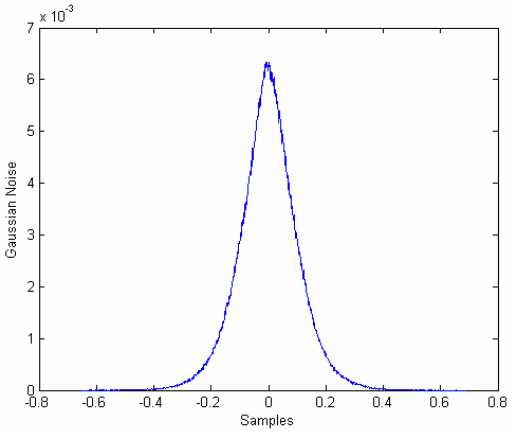

Fig. 16. Gaussian Noise Addition

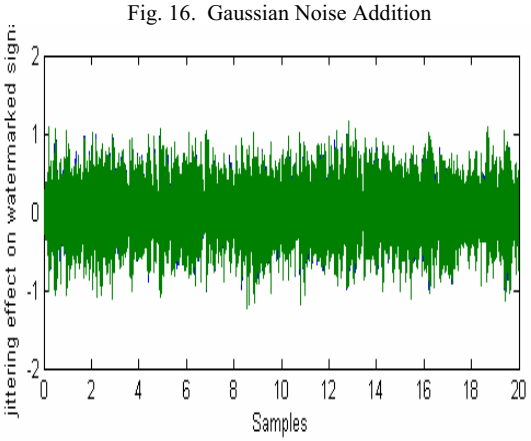

Fig. 17. Jittering Effect

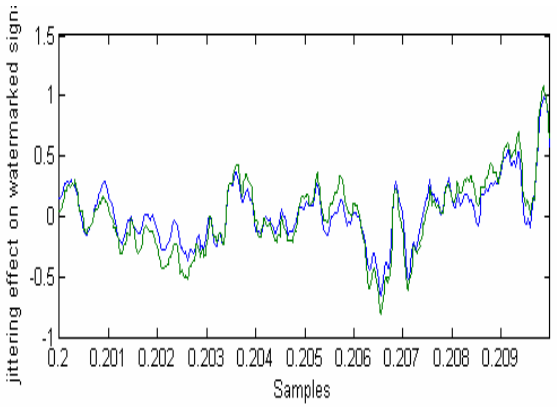

Fig. 18. Magnified Jittering Effect

D. Hostile Environment Results for Test2.wav:

Here results of Teat2.wav watermarked audio signal in the presence of hostile environment are shown. The hostile environment consists of Compression Error, Gaussian Noise addition and Jittering Effect. Fig. 19 shows the Compression Error which is the result of difference of original signal and watermarked signal. Fig. 20 shows Gaussian Noise Addition on the watermarked Signal, the order of noise is 5.2×10^{-3}. Fig. 21 and Fig. 22 shows Jittering Effect on the watermarked Signal. Jittering Effect is negligible due to this in the original it is not possible to see the Jittering Effect normally. So to examine the Jittering Effect, response is magnified in sample range as from 0.20 to 0.21.The blue color signal is original watermarked signal and green color signal is after Jittering Effect.

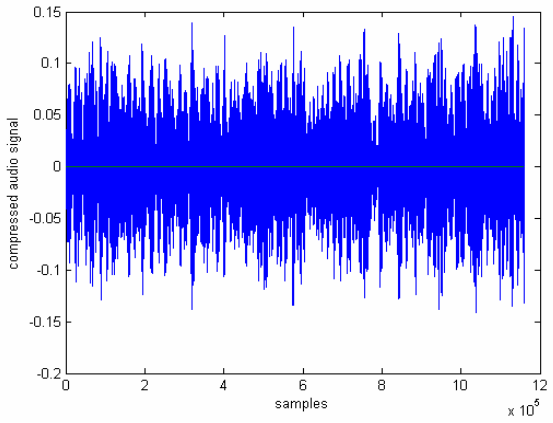

Fig. 19. Compressed Audio Signal

E. PCEW calculation Results:

According to the formula for PCEW calculation, the result for Test1.wav watermarked signal is 0.87 and for Test2.wav watermarked signal is 0.91.It shows that both signals lie nearly excellent level.

IX CONCLUSIONS

In this paper an objective method is proposed for audio or speech quality evaluation with the help of digital watermarking. This proposed method is based on the techniques of discrete cosine transform (DCT). Original signal is not required at detector, so it is a blind watermark detection technique. Our method is based on the fact that the embedded watermark and the speech will undergo the same attack (signal impairments). Therefore, the percentage of correctly extracted watermark bits is used to predict speech quality. As length of watermark message bits increases then signal to noise ratio of watermarked audio signal decreases. For 5 bits watermark message K=500-1000 is best. For 10 bits and 15 bits watermark message K=100-1000 is best. For 20 bits and 25 bits watermark message K=350-700 is best. From this method of Audio or Speech Quality Evaluation Quality Evaluation is much easier with Watermarking.

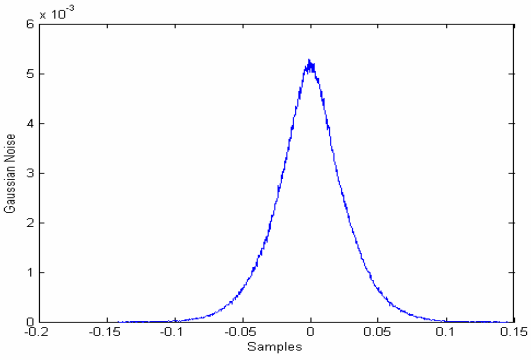

Fig. 20. Gaussian Noise Addition

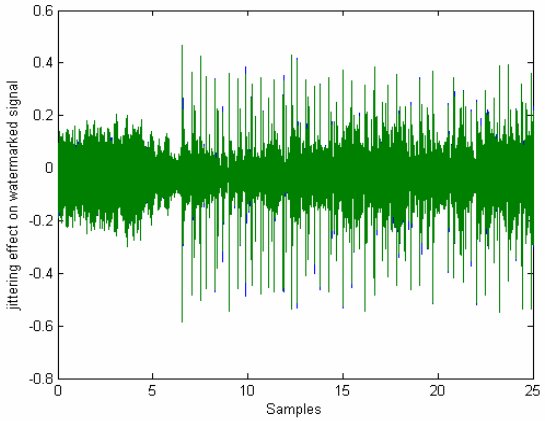

Fig. 21. Jittering Effect

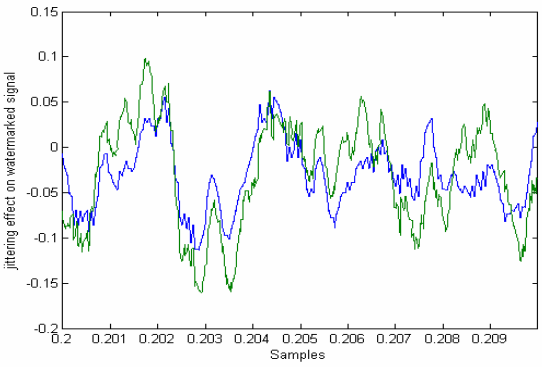

Fig. 22. Magnified Jittering Effect

REFERENCES

[1]. Libin Cai, Ronghui Tu, Jiying Zhao,, and Yongyi Mao, "Speech Quality Evaluation: A New Application of Digital Watermarking" IEEE Transaction on Instrumentation and Measurement, vol. 56, NO. 1, February 2007.

[2]. Abdellatif Zaidi, Rémy Boyer & Pierre Duhamel, "Audio Watermarking Under Desynchronization and Additive Noise Attacks". 1053-587X: p 570 – 584, © 2006 IEEE.

[3]. Sharkas, ElShafie, & Hamdy, "A Dual Digital-Image Watermarking Technique". Transactions on Engineering, Computing and Technology V5 April 2005 ISSN 1305-5313: p136 -139.

[4]. Akira Takahashi, Ryouichi Nishimura & Yôiti Suzuki, "Multiple Watermarks for Stereo Audio Signals Using Phase-Modulation Techniques". IEEE Transactions on signal processing, vol. 53, no. 2, February 2005: p 806 – 815

[5]. T. Falk, Q. Xu, and W. Chan, "Non-intrusive GMM-based speech quality measurement," in *Proc. 2005 IEEE ICASSP'05*, Mar. 2005, pp. 125–128.

[6]. D.-S. Kim and A. Tarraf, "Perceptual model for nonintrusive speech quality assessment," in *Proc. 2004 IEEE ICASSP'04*, May 2004, pp. 1060–1063.

[7]. T. Falk and W.-Y. Chan, "Objective speech quality assessment using gaussian mixture models," in *Proc. 22nd Biennial Symp. Commun.*, Kingston, ON, Canada, Jun. 2004, pp. 169–171.

[8]. L. Ding and R. Goubran, "Assessment of effects of packet loss on speech quality in VoIP," in *Proce. IEEE Int. Workshop HAVE2003*, Ottawa, ON, Canada, Sep. 2003, pp. 49–54.

[9]. V. Licks, F. Ourique, R. Jordan, and G. Heileman, "Performance of dirtypaper codes for additive white Gaussian noise," presented at the IEEE Workshop of Statistical Signal Processing (WSSP03), MO, 2003.

[10]. R. Bäuml, J. J. Eggers, and J. Huber, "A channel model for watermarks subject to desynchronization attacks," in *Proc. Int. ITG Conf. Source Channel Coding*, Berlin, Germany, Jan. 2002, pp. 28–30.

[11]. Kundur D, "Watermarking with diversity: Insights and implications". IEEE Multimedia 8(4): p 46–52, 2001.

[12]. J. W. Seok and J. W. Hong, "Audio watermarking for copyright protection of digital audio data," *Electron. Lett.*, vol. 37, no. 1, pp. 60–61, 2001.

[13]. P. Meerwald, A. Uhl, "A Survey of Wavelet-Domain Watermarking Algorithms" EI San Jose, CA, USA, 2001.

[14]. Yu H, Kundur D & Lin C, "Spies, thieves, and lies: The battle for multimedia in the digital era". IEEE Multimedia 8(3): p 8–12, 2001.

[15]. J.R. Hernandez, M.Amado, and F. Perez-Gonzalez, "DCT- Domain Watermarking Techniques for Still Images: Detector Performance Analysis And a New Structure", in IEEE Trans. Image Processing, vol. 9, pp 55-68, Jan. 2000.

[16]. F.A.P. Petitcolas, "Watermarking Schemes Evaluation" ", in IEEE Signal Processing Magazine, Vol 17, pp 58-64, September 2000.

[17]. G. Langelaar, I. Setyawan, R.L. Lagendijk, "Watermarking Digital Image and Video Data", in IEEE Signal Processing Magazine, Vol 17, pp 20-43, September 2000.

[18]. J.R. Hernandez, M.Amado, and F. Perez-Gonzalez, "DCT- Domain Watermarking Techniques for Still Images: Detector Performance Analysis And a New Structure", in IEEE Trans. Image Processing, vol. 9, pp 55-68, Jan. 2000.

[19]. M. Arnold, "Audio watermarking: Features, applications and algorithms," in *IEEE Int. Conf. Multimedia and Expo 2000*, vol. 2, 2000, pp. 1013–1016.

[20]. M. Kutter, F. Hartung, "Introduction to Watermarking Techniques" in Information Techniques for Steganography and Digital Watermarking, S.C. Katzenbeisser et al., Eds. Northwood, MA: Artech House, Dec. 1999, pp 97-119.

[21]. Information hiding techniques of steganography and digital watermarking Stefan Katzenbeisser. Fabien A. P. Petitcolas (Editors), Artech House Books, 1999.

[22]. I.J. Cox, M.L. Miller, J.M.G. Linnartz, T. Kalker, "A Review of Watermarking Principles and Practices" in Digital Signal Processing for Multimedia Systems, K.K. Parhi, T. Nishitani, eds., Marcel Dekker, Inc., 1999, pp 461-482.

[23]. F.A.P. Petitcolas, "Introduction to information hiding" in Information Techniques for Steganography and Digital Watermarking, S.C. Katzenbeisser et al., Eds. Northwood, MA: Artech House, Dec. 1999, pp 1-11.

[24]. J. Dugelay, S. Roche, "A Survey of Current Watermarking Techniques" in Information Techniques for Steganography and Digital Watermarking, S.C. Katzenbeisser et al., Eds. Northwood, MA: Artech House, Dec. 1999, pp 121-145.

[25]. Kii. H, Onishi. J,Ozawa.S, "The digital watermarking method by using both patchwork and DCT," Mulmedia Computing and Systems. IEEE International Conference on vol. 1, pp. 895-899, 1999.

[26]. I. Cox, M. Miller, and A. McKellips, "Watermarking as communication with side information," in *Proc. Int. Conf. Multimedia Computing Systems*, Jul. 1999, pp. 1127–1141.

[27]. M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "DCT-domain system for robust image watermarking," *Signal Process.*, vol. 66, pp. 357–372, 1998.

[28]. F. Hartung and B. Girod, "Watermarking of uncompressed and compressed video," *Signal Process.*, vol. 66, no. 3, pp. 283–301, 1998.

[29]. S Craver, et. al., "Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks and Implications", IEEE Jou.. Selected Areas in Communications, Vol. 16, No. 4, May 1998, pp 573-586.

[30]. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for images, audio and video," in *IEEE Int. Conf. Image Processing*, vol. 3, pp. 243–246, 1996.

[31]. Allen, J.B, "How do human process and recognize speech?" IEEE Trans. on Speech and Audio Processing, vol. 2, no. 4, pp.567-577, 1994.

[32]. R.C.Gonzalez and R.E.Woods, Digital Image Processing, Addison-Wesley Publishing company, Inc., 1993.