AUDIO SECURITY SYSTEM BASED ON ECHO HIDING AND AUIO INPAINTING

[1]Aswathy Venugopal, [2]M.r. Jayadas C.K
[1]MTECH Student, [2]Professor
[1]Department of Electronics and Communication Engineering ,
[1]College of Engineering Cherthala, Alappuzha, India

Abstract : This paper propose an efficient audio security system based on Echo Hiding Audio Watermarking and Audio inpainting technique. For watermarking, echoes with two different delays are used. One is used to encode watermark bit “1” and the other one is used to encode watermark bit “0”. Echoes embedded in one channel are used to encode watermark bit “1” and echoes embedded in the other channel are used to encode watermark bit “0”. The embedded echoes or watermark are extracted by detecting time delay of each embedded echo in Cepstrum domain. Robustness and performance of this audio watermarking scheme under different types of attacks are investigated. The watermarked signal is decoded and the output is Audio inpainted to improve the quality of the audio signal. In this framework, the distorted data are treated as missing and their location is assumed to be known. The signal is decomposed into overlapping time-domain frames and the restoration problem is then formulated as an inverse problem per audio frame. Sparse representation modeling is employed per frame, and each inverse problem is solved using the Orthogonal Matching Pursuit algorithm. This approach is shown to outperform commercially available methods for audio declipping in terms of Peak Signal-to-Noise Ratio.

***IndexTerms* - Audio inpainting , Audio watermarking, Echo Hiding, Cepstrum, Orthogonal Matching Pursuit Algorithm,HAS**

I. INTRODUCTION

An audio signal is a representation of sound, typically as an electrical voltage for analog signals and a binary number for digital signals. Audio signals have frequencies in the audio frequency range of roughly 20 to 20,000 Hz. The technique of hiding information is known as watermarking. Watermarking describes methods and technologies that hide information, for example a number or text, in digital media, such as images, video or audio. The embedding takes place by manipulating the content of the digital data, which means the information is not embedded in the frame around the data. The hiding process has to be such that the modifications of the media are imperceptible. Audio watermarking is a well-known technique of hiding data through audio signals. It is also known as audio steganography and has received a wide consideration in the last few years. Perceptual properties of human auditory system (HAS) help to hide multiple sequences of audio through a transferred signal.

1.1 GENERAL BACKGROUND

Digital processing of audio on personal computer is becoming more and more common. Even today's standard PC's are capable of processing CD quality audio data in real time. The purpose of this research is to develop an effective audio security system. The most basic definition of any security system is found in its name. It is literally a means or method by which something is secured through a system of interworking components and devices.

Due to the rapid development in the field of internet music piracy has become a common issue. Most of the music piracy is because of rapid growth and easiness of current technologies for copying, sharing, manipulating and distributing musical data. Multimedia is involved in almost every aspect of human life, and many of us rely on various websites for information on events taking place all over the world. Concerns regarding how to authenticate multimedia data have led to research in forgery detection, but studies in audio are still limited compared to those in the image and video. Audio carrier includes the content of music, phone conversations, and other recording forms. Although some sources give us authentic information, others contain forged content. Audio forgery techniques could be used to conduct piracy over the Internet, falsify court evidence, or modify security device recordings of events taking place in different parts of the world. Real-world examples can best illustrate the importance of audio authentication.

Hiding data in audio signals presents a variety of challenges due in part to the wider dynamic and differential range of the human auditory HAS as compared to the other senses. The HAS perceives over a range of power greater than one billion to one and a range of frequencies greater than one thousand to one Sensitivity to additive random noise is also acute

1.2 OBJECTIVE

Most of the music piracy is because of rapid growth and easiness of current technologies for copying, sharing, manipulating and distributing musical data. As one promising solution, audio watermarking has been proposed for post-delivery protection of audio data. Digital watermarking works by embedding a hidden, inaudible watermark stream into the host audio signal.

The objective of this project is to develop a high end audio security system. The input signal is watermarked based on echo hiding technique. At the receiver end the audio inpainting technique is applied to improve the quality of the signal.

The echo hiding is widely used for audio watermarking scheme . The embedded echo kernels can be represented as several discrete impulses in the time domain. For simplicity, an echo kernel with only one echo is written in equation 1.1

$$h(n) = \delta[n] + \alpha\delta[n-d] \qquad (1.1)$$

Where δ(n) is Kronecker delta function, α represents amplitude, d is the delay offset. .Due to different audio watermarking techniques and process there is chance that the signal get distorted. Once the audio is properly watermarked the quality of the output signal should be ensured. It is done using audio inpainting technique. Audio inpainting as a general problem encountered in many applications: one observes a partial set of reliable audio data while the remaining unreliable data is either totally missing or highly degraded; the unreliable data is considered missing and it is estimated from the reliable data portion. Orthogonal matching pursuit algorithm is used in the audio inpainting technique.

1.3 SCOPE

Digital audio watermarking is more challenging as compared to watermarking other media. The proposed audio security system is based on Audio watermarking based on Echo Hiding and Audio inpainting. The Human Auditory System (HAS) is less sensitive to echo's, this property is used in this work. The Internet is an open network, being increasingly used for delivery of digital multimedia contents. In the digital format, content is expressed as streams of ones and zeros that can be transported flawlessly. The contents can be copied perfectly infinite times. A user can also manipulate these files. However, good business senses necessitates two transaction mechanisms content protection and secure transport over the Internet. The proposed audio security system can be used in the field of war to transmit information without getting interrupted by intruder. It has wide scope in the field of multimedia to avoid music piracy.

II LITERATURE SURVEY

Generally, many audio watermarking techniques have been developed. The well known methods of audio watermarking based on the limitations of perceptual properties of HAS are including simple least significant bits (LSB) scheme or low bit encoding, phase coding, spread spectrum, patchwork coding, echo coding, andnoise gate technique.

2.1 THEORETICAL INVESTIGATIONS

A pathway for watermarking especially for the famous patchwork algorithm was proposed. His method improves the performance of the original patchwork algorithm. Another method called as modified patchwork algorithm (MPA) enhanced the power of Arnold's algorithm and improved its performance in terms of robustness and inaudibility. A mathematical formulation has also been presented that aids to advance the robustness. Spread-spectrum technology has been utilized in audio watermarking . Another method based on the spreadspectrum technology is a multiple echo technique that replaces a large echo into the host audio signal with multiple echoes with different offsets. Next method is the positive and negative echo-hiding scheme . Each echo contains positive and negative echoes at adjacent locations. In the low-frequency band, the response of positive and negative echoes forms a smooth shape that is resulted by similar inversed shape of a negative echo with that of a positive echo. When positive and negative echoes are employed, the quality of the host audio is not obviously depreciated by embedding multiple echoes.

In [1] Mohamed F. Mansour and Ahmed H. Tewk proposes a new system for data embedding. The interval length between the audio signal's salient points are changed. Then the embedding is done in the quantization indices of the intervals. The wavelet extreme of the signal envelope are used as the salient points. The algorithm is robust against different signal processing attacks. But watermark detection depends on the selection of threshold and is not robust against pitch shifting . The salient point lengths are modified for embedding the data. The salient point interval lengths are quantized after extracting and refining them in the extraction process.

Wei Li, Xiangyang Xue, and Peizhong Lu proposes a novel content-dependent localized robust audio watermarking scheme , to combat synchronization attacks in [2]. First select the steady high-energy local regions that represent music edges by using different methods .The methods are either selects embedding regions on the original audio waveform directly, select the peaks on the audio envelope as reference to determine the embedding regions or selection based on music content analysis .After selecting the region then embed the watermark in these regions. Such regions will not be changed much for maintaining high auditory quality and in this way, the embedded watermark is robust against different distortions. The prime factor in watermark detection is accurately locating the embedding region .

In [3] Nima Khademi Kalantari .et.al proposes a multiplicative patchwork method.In this system two subsets of the host signal features are selected and the selection is based on the secret keys inorder to embed watermark data within the host signal. The watermark data is embedded by multiplying one of the subsets .The other subset is leaved unchanged. Embedding is performed in the selected frames of the host signal which satisfies a certain condition.The wavelet domain is used for the implementation of the method and approximation coefficients are used for data embedding. T he inaudibility of the watermark insertion is controlled by Perceptual Evaluation of Audio Quality (PEAQ) algorithm .The error probability is being derived. Energy of the subset is changed according to the embedding functions and due to multiplication. In order to extract the watermark data the energy ratio of two subsets should be compared with a threshold.

Another technique was introduced by Xiangui Kang, Rui Yan and Jiwu Huang in [4],they proposes a multi-bit spread-spectrum audio watermarking scheme based on a geometric invariant log coordinate mapping (LCM) feature which is very robust to audio geometric distortions. The embedding of watermark is done in the LCM feature. Actually embedding is done in the Fourier coefficients which are mapped to the feature via LCM. the embedding is actually performed in the DFT domain without interpolation, and this will completely eliminate the severe distortion formed due to the non-uniform interpolation mapping. One of the advantage of the watermarked audio is its high auditory quality in both objective and subjective quality assessments. A mixed correlation between the LCM feature and a key-generated PN tracking sequence is proposed to align the logcoordinate mapping, thus synchronizing the watermark efficiently with only one FFT and one IFFT

In [5] Ryouichi Nishimura, proposes a watermarking technique for audio signals, based on the spatial masking phenomenon and ambisonics. Ambisonics is a well known technique to encode and reproduce spatial information related to sound.A watermark data is created from the original sound scene by slightly rotating it. Ambisonic signals are synthesized from mono or stereo signals. Watermarks are embedded near the signal by adding a small copy of the host signal which provides reversible watermarking is possible if loudspeakers are arranged properly for playback. This will represent audio signals of first order ambisonics as mutually orthogonal fourchannel signals.And this make it possible to present a larger amount of data than in the original version.

Yong Xiang.et.al proposes a novel dual-channel timespread echo method for audio watermarking in [6], This method improves the robustness and perceptual quality of the system. At the embedding stage, the audio signal is divided into two subsignals and watermarks are added into the two subsignals simultaneously. The embedded subsignals are combined to form the watermarked signal. The watermarked signal is split up into two watermarked subsignals at the decoding stage.The similarity of the cepstra to the watermarked subsignals is used to extract the embedded watermarks. The performance of watermark extraction can be enhanced by using large peaks of pseudonoise sequences auto-correlation function.

In [7] Guang Hua.et.al, proposes a Convex optimization based finite-impulse-response(FIR) filter design.FIR is used to obtain the optimal echo filter coefficients. Proposed maximum power spectral margin (MPSM) and the absolute threshold of hearing (ATH) of human auditory system (HAS) is used to shape the desired power spectrum of the echo filter. This will ensure the optimal imperceptibility. The auto-correlation function of the echo filter coefficients will controls the robustness in terms of watermark detection. To design the echo kernel, convex optimization is used to obtain a set of filter coefficients which will replace the PN or MPN sequence.

A different method was introduced by Andrea Abrardo and Mauro Barni in [8], proposes a watermarking schema based on antipodal binary random binning, which allows the possibility of relying on simple and effective binary code constructions. And also make it easy to cope with amplitude scaling. The system outperforms previous watermarking system constructions by exhibiting very good performance also in the presence of gain attack. The encoding process is formed by an outer convolutional encoder CCo, the input to convolutional encoder is watermark w. A random interleaver permuting the outer codewords bits wc and input to the inner convolutional encoder is the permuted outer codewords wc. The outer encoder add some redundancy bits to the watermark message and the inner code embeds the watermark in the host signal by a dirty trellis mechanism. And atlast the sequences are mixed together to form the watermarking signal.

In [9] Yong Xiang.et.al proposes a patch work based audio watermarking method robust to de-synchronization attack. The watermarks are embedded into the host audio signal in the discrete cosine transform (DCT) domain and synchronization bits are implanted in the logarithmic DCT domain. At the decoding stage, find the scaling factor imposed by an attack by analyzing the received audio signal in LDCT domain. Then remove the scaling effect by modifying the received signal, together with the embedded synchronization bits. And extract the watermark from the modified signal.

Michael Arnold.et.al proposes a phase based audio watermarking system robust to acoustic path propagation in [10]. The system embeds information by modulating the phase in the WOLA domain. To improve the overall performances, three complementary modules are introduced: (i) a perceptual model to evaluate the noisiness of individual spectral bins, (ii) a resynchronization module which relies on tracking ther esampling ratio over time and (iii) a statistical detector that aggregates the information from several values of the detection correlation array. The watermark modulation/demodulation framework can then be complemented by any regular channel coding technique. This channel coding is overlaid on top of the system. This can include either expanding the alphabet of symbols, introducing an error correcting code or establishing several communication channels associated to different frequency bands.

In [11] Mehdi Fallahpour and David Megas, proposes a audio watermarking system where embedding is done on the basis of Fibonacci numbers. Watermark extraction is done in a bit-exact manner by changing some of the magnitudes of the FFT spectrum. The main idea of the watermarking system is to divide the FFT spectrum into short frames and the change the magnitude of the selected FFT samples using Fibonacci numbers. First select a frequency part from FFT spectrum. Then divide the frequency band into frames. And a single secret bit is being embedded into each frame in the selected frequency band. On the basis of Fibonacci number and secret bit the magnitude of FFT coefficients in the selected frame is modified. If the bit to be embedded is 0, the coefficient value is changed to Fibonacci number with even index which is closer to the value. If the bit to be embedded is 1, the coefficient value is changed to Fibonacci number with odd index which is closer to the value.Three categories for audio watermarking are summarized in Fig 2.1 which are based on prominent domains for embedding data in an audio signal: temporal, frequency and coded domains.

2.1.1 TEMPORAL DOMAIN

Audio watermarking techniques based on temporal domain are summarized in this section. Famous techniques for temporal domains are including low-bit encoding, echo hiding, and hiding in silence interval. In the following, each technique is fully discussed in detail.

2.1.2 LOW BIT ENCODING

The most applied method for data hiding is called as low-bit encoding or lease significant bit (LSB) . Basically, the least significant bit of the cover audio is utilized for embedding each bit from the message.

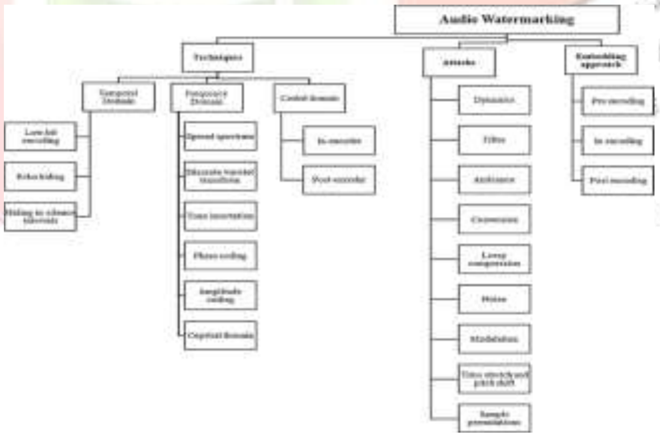


Fig 2.1 Audio Watermarking Techniques [1]

For example, 8 kbps data are hidden in a signal with 8 kHz sampled audio which has 8 bits per sample. This method is relatively simple and has a high capacity for hiding data. The robustness of this method is increased when it is combined with other watermarking methods. Nevertheless, the low-bit encoding method is sensitive to noises, which reduces the security and robustness.

Audio watermarking techniques signal is known which makes this method vulnerable to attacks and an attacker via elimination of entire LSB plane can easily discover a message or destroy the watermark. Basic LSB has been performed for transmission of an audio signal on a wireless network in . The results verified that the method reduces the robustness and security at high rate of embedding data, but it does not harm the imperceptibility of final signal. A method for embedding four bits per sample was presented in that enhanced the hiding capacity. This method reduces the impact of error on the watermarked audio signal by defusing the embedding error on the next four samples. The depth of embedding layer of data increased from 4 layers to

6 and 8 LSB layers with no significant effect on the imperceptibility of the audio signal . The results showed that the methods with higher embedding layer enhanced the robustness of previous method when noise addition and distortion occurs.

As an instance, if four bits “0100” (value 4) are used for embedding data and bit “1” must be embedded in the bit sequence, it is suggested to select the bit sequence of “0011” (that is value 3) instead of having “1100” (that is value 12). The reason is that value 3 is closer to value 4 and the result is a lower embedding error rate. In order to enhance imperceptibility of watermarked signal, the approach avoids hiding data in silent periods of the original signal. Due to assigning 8 bits for LSB embedding, the hiding capacity of the result becomes lower than the previous methods. However, it improves the robustness. The major disadvantage of embedding data in 6th or 8th position of LSB is the difficulty to reveal the original audio signal especially when the bits are shifted or flipped to enhance the embedding error rate.

2.1.3 ECHO HIDING

An audio effect is known as echo which repeats some parts of the sound by creating delay inside the audio signal. In order to hide an echo, echo-hiding method generates a short echo by using a resonance and adds the echo to the original audio signal. The addition of the short echo is not recognizable by HAS; therefore, this method is not sensitive to noise addition. Other perceptual and statistical properties of original signal are kept in resulted signal.

Three parameters of the echo signal are the candidates for hiding the data. They include the initial amplitude, the delay (or offset), and the decay rate. The data can be successfully hidden in the audio signal if their values are managed to keep the imperceptibility of audio signal . For this reason, the values of amplitude and decay rates should be set below the audible threshold of HAS. As an example, when the time difference between the original signal and the echo stays below 1 ms, there is no annoying effect on the audibility of the signal. Due to the induced size of echo signal, low embedding rate, and security, there are few systems and applications that practically developed this method.

To the best of our knowledge, there is no real system that uses echo hiding in audio watermarking which cannot provide sufficient data for evaluation. An echo-hiding-time spread technique has been introduced to resolve the low robustness of echo-hiding technique in facing with common linear signals . This method spreads the watermark bits all over the original signal and the destination recovers them by using the correlation amount. As a result of being a cepstral content-based method, the cepstral portion of error is detached and the detection rate at the decoder gets higher.

2.1.4 HIDING IN THE SILENCE INTERVEL

Another candidate for embedding data is silence intervals in speech signal. A simple approach for hiding in silence intervals is proposed . Consider n as the number of required bits for denoting a value from the message to hide. The silence intervals in audio signal should be detected and measured in terms of the number of samples in a silence interval. These values are decremented by x, $0 < x < 2n$ bits, where $x = \mathrm{mod}(\mathrm{new\_interval\_length}, 2n)$. As an instance, consider that the value 6 is hided in a silence interval with length 109. Taken 7 samples out from the interval, 102 samples are remained in the new interval. The value x is computed as $x = \mathrm{mod}(102, 8) = 6$. The short length of silence intervals that commonly seen in continuous parts of normal audios is omitted from the portions for hiding data. The perceptual transparency of this method is acceptable, but compression of signal misleads the data extraction process. As a solution for this problem, an approach is presented in which separates the silence intervals from audio intervals so that they are not interpreted as one another. Thus, it reduces the samples in silence intervals and slightly augments the samples of the audio interval. The first and last interval added to the audio during MP3 coding is simply ignored in data hiding and retrieval. As a general conclusion, conventional LSB approach is simpler than other methods; however, its capacity for hiding data is low. Moreover, it is resilient to noise additions and shows higher robustness in comparison with its variants .

2.1.5 FREQUENCY DOMAIN

Main idea behind using the frequency domain (or transform domain) for hidden data is the limitation of HAS when frequency of an audio signal fluctuates very rigid. The “masking effect” phenomenon enables the HAS to mask weaker frequency near stronger resonant frequencies . It provides a time duration that can be utilized for embedding data. The data hidden in this space is not perceptible by HAS. Watermark methods in frequency domain directly manipulate the masking effect of HAS by explicit modification of masked regions or indirectly by slight change of the samples of the audio signals.

2.1.6 SPREAD SPECTRUM

By spreading data in the frequency domain, spread spectrum (SS) technique ensures an appropriate recovery of the watermarked data when communicated over a noise-prone channel. SS utilizes redundancy of data for degrading the error rate of data hiding. An M-sequence of code handles the data and is embedded in the cover audio. This sequence is known to sender and receiver and if some parts of these values are modified by noise, recovery of data is feasible by using other copies . The SS technique was developed in MP3 and WAV signals for the purpose of hiding confidential information in the form of conventional direct sequence spread spectrum (DSSS) technique .A frequency mask was suggested for embedding the data in a watermarked audio signal . When a phase-shifting approach is combined to SS, the result is a watermarked signal with a higher level of noise resistance and robustness. The detection of hidden data is simple in the new method, but the rate of hiding data is low. As a solution, sub-band domain is chosen to provide better robustness and improving the decoder's synchronization uncertainty which require to select proper coefficients in sub-band domain .

2.1.7 DISCRETE WAVELET DOMAIN

Discrete wavelet transform (DWT) is multi-scale and multi-resolution technique to decompose signal to different time-frequency components. A watermarking method is proposed by DWT which hides data in LSB of the wavelet coefficients . The imperceptibility of hidden data is low in DWT. Whenever the integer wavelet coefficients are available, a hearing threshold is useful to improve the audio inaudibility . If a DWT watermarking technique evades embedding data in silent parts, hidden data does not annoy the audience . DWT provides a high rate of data hiding; nevertheless, the procedure for data extraction at the receiver is not always accurate.

2.1.8 PHASE CODING

Another limitation of HAS is its inability to detect the relative phase of different spectral components. It is the basis of interchanging hidden data with some particular components of the original audio signal. This method is called as phase coding and works well on the condition that changes in phase components are retained small. Phase coding tolerates noises better than all other above-mentioned methods. An independent multi-band phase modulation is utilized for phase coding . In phase modulation method, phase alteration of the original audio signal is controlled to obtain imperceptibility of phase modifications. Phase components are determined by quantization index modulation (QIM).

Then, the nearest "o" and "x" points are replaced with phase values of frequency bin to hide "0" and "1," respectively. Therefore, phase coding achieves a higher robustness when perceptual audio compression is applied .

QIM was widely been used that improves the capacity of data hiding of phase coding by replacing the strongest harmonic with step size of $\pi/2n$. Phase coding has zero value of bit error rate (BER) when MP3 encoder is applied that demonstrates the high robustness of this method. As HAS is not sensitive to phase changes, an attacker simply can replace his/ her data with the real hidden data. S/he can apply frequency modulation in an inaudible way and modify the phase quantization scheme.

2.1.9 AMPLITUDE CODING

The sensitivity of HAS is high for frequency and amplitude components. Therefore, it is possible to embed hidden data in the magnitude audio spectrum. The capacity of hiding data is high by using this method and the tolerance of the method regarding noise distortion and its security in facing with different attacks is high. Hiding different types of data is feasible by using this method. Encrypted data, compressed data, and groups of data (LPC, MP3, AMR, CELP, parameters of speech recognition, etc.) can be hided by using amplitude coding.

2.1.10 CEPSTRAL DOMAIN

Cepstrum coefficients provide spaces for watermarking. This method is resilient to well-known attacks in signal processing and is also known as log-spectral domain. It locates the hidden data in the portions of frequencies that are inaudible by HAS and obtains a high capacity of hiding data, between 20 and 40 bps . Initially, the domain of original audio signal is modified to cepstral domain. Statistical mean function helps to choose some cepstrum coefficients that are later altered by hidden data. As the masked regions of the majority of cover audio frames are
utilized for data hiding, the imperceptibility of watermarking is relatively high in cepstral domain.

2.1.11 EMBEDDING APPROACH

In covert communication, data is transferred through multiple encoders/decoders. An encoder reduces the size of transmitted data by removing the redundant or unused data. Thus, each coder influences the integrity of data, while the robustness of covert communications requires a high integrity of watermarked data. Although, there are some ways to ensure data integrity in encoder/decoder, it imposes negative impacts on hiding capacity of data. There are three levels for embedding a data-in-audio watermark system . Figure 3.2 summarizes the aforementioned methods for audio steganography according to the occurrence rate. The evaluation of security requires a third-party effort cost to retrieve the hidden data. Each level has some benefits and weaknesses that are discussed as follows.

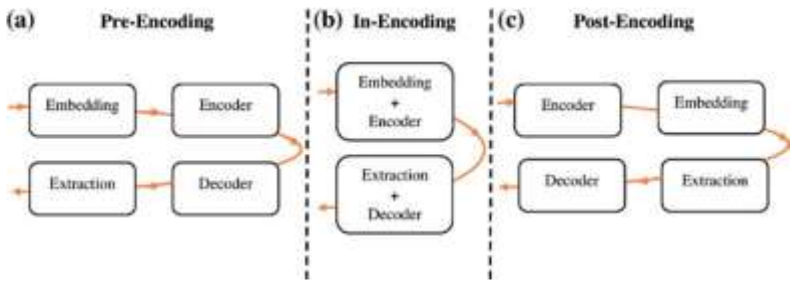


Fig 2.2 : Different approaches for embedding [1]

2.1.12 EMBEDDING BEFORE ENCODING (PRE-ENCODING)

Prior to encoding process, the data is embedded in time and frequency domain. This level is known as pre-encoder embedding. The integrity of data, during transmission over network, is not guaranteed in this level because high degree of data compression in encoders (e.g., in ACELP or G.729) and addition of noise (in any form, e.g., WGN) can compromise the integrity of data. On the other hand, there are some methods that allow a low degree of modifications on the audio signal including resizing, resampling, filtering. Therefore, they are resilient to low degree of noise addition or data compression. Only noise-free environments provide a space for high rate of data hidden.

2.1.13 EMBEDDING DURING ENCODING (IN-ENCODER)

This data embedding level provides a robust data hiding. For this purpose, a codebook of codecs is necessary. The codebook keeps the information of transmitted data once the requantization operation is performed. As a result, for every parameter of audio signal, two important values of embedded-data and codebook parameters are kept. When value of embedded data is manipulated for any reason, this method faces to a severe problem for data extraction. It can occur when the data passes through a voice encoder/decoder in a radio access network (BST, BSC, TRAU) and/or in the core network (MSC) in a GSM network. Similar modifications occur when a voice enhancement algorithm is developed in a radio access network and/or in the core network.

2.1.14 EMBEDDING AFTER ENCODING (POST ENCODER)

This level of embedding data acts on bitstreams rather than the original audio signal. Data is hidden in a bitstream once it passes the encoder and before entering the decoder. Thus, value of data and the integrity of watermarked audio signal are vulnerable to undesirable modifications. Bitstreams are naturally more sensitive to alteration than audio signals and data integrity should be kept small to avoid imperceptibility of audio signal. Nevertheless, post-encoder embedding ensures the correctness of data once it is extracted in tandem-free operations and the message is retrieved in a lossless way.

2.1.15 AUDIO ATTACK

There are many attacks that can degrade the watermark data and as a consequence decrease the robustness of the audio watermarking techniques. In this section, the impact of each attack according to the audibility of hidden data by HAS is measured and the most effective attacks on audio signals are highlighted. Some of the attacks mostly occur in real environments. Suppose an audio signal is prepared to be broadcast on a radio channel. Based on the audience confidence and quality parameters of the radio channel, the audio material is normalized and compressed to fit the necessary level of loudness for transmission. Then, the quality of signal is optimized by equalization; undesired parts are demised or dehisced; useful frequencies are kept and unnecessary ones are omitted by filters.

In some applications, the robustness of watermarked audio signal should be high, e.g., in commercial radio transmission or copyright protection of music. In both examples, the watermark technique should not allow the signal to be destroyed or manipulated by attackers and if an attack occurs, it should not allow the attacker to misuse or reuse the signal. A well-known attack in this situation is lossy compression in MP3 at high rate of compressions. In addition to individual attacks, some attacks act in the form of groups. The group of attacks is also taken into account for performance evaluation of watermarking techniques. Main group attacks are including dynamics, filter, ambience, conversion, loss comparison, noise, modulation, time stretch (pitch shift), and sample permutation.

2.1.16 DENOISER

In some cases, it is essential to find a way for noise removal from the signal. Denoiser acts as a gate. It passes the eligible parts of the signal and blocks the noises. A denoiser needs a value to be used for detection of a noise. A basic denoiser simply considers loudness of signal as a noise, prior that a proper value of the loudness should be set. Here, the setting is assumed as −80 and −60 dB. Indeed, for detection of complicated noises, other techniques, e.g., DE clickers, and advanced tools are required.

2.1.17 NOISE

So far, several attacks have been discussed. The result of most of the attacks is a noise. As already discussed, different sources of noise are known. Hardware components are the most effective sources of noise in audio signals. There is another attack that adds noise to terminate the watermark.

2.1.18 MODULATION

Modulation effect can be considered as attacks, but they usually do not happen in postproduction. Software for processing audio signals can include modulation attacks. They are as follows:

Chorus: Sounds from multiple resources in the form of a modulated echo is added to the original audio signal. The delay time and strength and number of voices are different. Here, 5 voices, 30 mms max. delay, 1.2 Hz delay rate, 10 % feedback, 60 ms voice spread, 5 db vibrato depth, 2 Hz vibrato rate, 100 % dry out (unchanged signal), and 5 % wet out (effect signal)are taken into account.

Flanger: when a delayed signal is added to the original signal, flanger is generated. The delay is short and the length changes constantly.
Enhancer: An audio signal becomes more brilliant or excited if the amount of high frequencies is increased. To simulate the effect of enhancer (or exciter), sound forge is applied and medium setting is used. Detailed information about the parameters is not provided by the program.

2.1.19 TIME STRETCH AND PITCH SHIFT

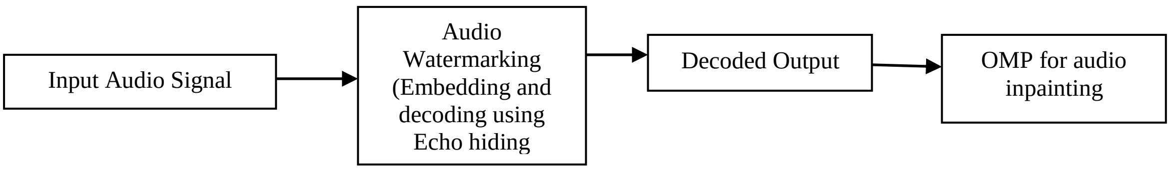
Time stretch and pitch shifts help to fine-tuning or fitting audio into time windows by changing the length of the audio signal with no changes in the pitch or vice versa.

Pitch Shifter: A complicated algorithm for editing audio signals is pitch shifter. This algorithm changes the base frequency of the signal with no modifications in the speed. So far, multiple pitch shifter algorithms have been presented in the literature. Selection of proper algorithm depends on the expected quality of the signal. The sound forge increases the pitch by 5 cent, and this is 480th of an octave.

Time Stretch: Time stretch prolongs or shortens the duration of an audio signal with no modification on the pitch. Here, a sound that forges with a length of 98 % of the original duration is considered.

III METHODOLOGY

This works aims at developing a high end audio system. A new audio watermarking technique based on Echo Hiding and Audio Inpainting is proposed. Human auditory system is less sensitive to echo of itself. An audio segment is divided into two parts, in which the positions for embedding “Zero” and “One” in the echoes are interchanged alternately. The audio inpainting framework that recovers portions of audio data distorted due to impairments such as impulsive noise, clipping, and packet loss. In this framework, the distorted data are treated as missing and their location is assumed to be known. The signal is decomposed into overlapping time-domain frames and the restoration problem is then formulated as an inverse problem per audio frame. Sparse representation modeling is employed per frame, and each inverse problem is solved using the Orthogonal Matching Pursuit algorithm(OMP). The block diagram representation of the proposed method is shown in the figure 3.1.



Watermarking has also many other applications such as copy control, broadcast monitoring and data annotation . For audio watermarking, several approaches have been recently proposed in the literature. These approaches include audio watermarking using phase embedding techniques , cochlear delay , spatial masking and ambisonics , echo hiding , patchwork algorithm , wavelet transform , singular value decomposition and FFT amplitude modification.

As one promising solution, audio watermarking has been proposed for post-delivery protection of audio data. Digital watermarking works by embedding a hidden, inaudible watermark stream into the host audio signal.

A non-blind audio watermarking method is described. The method makes use of echo-hiding to represent watermark bits Assuming data, *I*, hidden in the audio signal is in binary form, i.e.,0,1. For the single channel echo-hiding method, two types of echoes are embedded in the signal to encode two different watermark bits (bit "1" and "0"). Since most high quality music signals are stereo signals, two channels of data are available. For the proposed watermarking scheme, both channels of stereo signals are used for watermarking. For this scheme, one type of echo is sufficient to encode watermark bits "1" and "0". The echoes embedded in one channel are used to encode watermark bit "1" and the echoes embedded in the other channel are used to encode watermark bit "0". The watermark extracting process is performed by detecting the time delay of each embedded echo in Cepstrum domain.

3.1 SINGLE-CHANNEL ECHO-HIDING METHOD

The concept of audio watermarking using echo-hiding was first introduced by Gruhl et al. in 1996 . The echo-hiding method with many outstanding features is adopted in various audio watermarking applications.

An echo can be considered as a delayed version of the signal itself. The delay can be made small such that the echo is not audible. Single-channel echo-hiding method embeds watermarks into an original single-channel signal by hiding inaudible echoes to produce a watermarked signal. The original signal is first divided into many segments. Each segment may be considered as an independent signal for embedding one type of echo. Two types of echoes are used for encoding bit "1" and "0" by using different values of delay, d1 and d2 respectively, as illustrated in Fig. 1. The time delays and amplitudes of the echoes must be chosen such as to make the echoes inaudible and yet the echoes must be readily extracted from the watermarked signal.

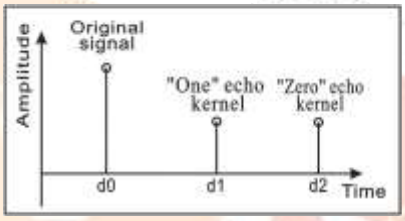


Fig 3.2 : Impulse response of "One" echo kernel and "Zero" echo kernel [2]

The watermarked segment *w*[n], is obtained by taking the convolution of the selected segment of the original signal *s*[n] and the echo kernel e[n]. Mathematically,

$$w(n) = s(n) * e(n) \qquad .... \quad (3.1)$$

The final watermarked signal is a composite of all watermarked segments. The embedding process when "one" echo kernel with time delay "d1" is used to encode watermark bit "1" is illustrated in Fig.3.3.

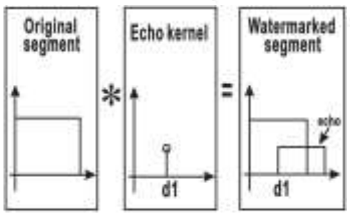


Fig 3.3 : Embedding process [2]

Watermark decoding process is carried out by detecting the delay value of each embedded echo. The Cepstrum computation is used for the extraction process and it can be viewed as conducting de-convolution process to separate embedded echo from watermarked segment.

3.2 Watermark Embedding Process

Original signals are first divided into segments of duration, say 10ms. If echoes are to be embedded in segments with very low amplitude, especially in silent segments, they become audible and the audio quality will be badly affected. In addition, any type of attack can easily 'destroy' the embedded echoes in such low energy segments. Hence, the energy level of each segment must be first determined to assess if it is suitable for embedding. A watermark bit stream in binary format is generated first. The embedding process starts in the first non-silent segment after the header, and the echoes are embedded according to the watermark bit sequence. If the watermark bit is “1”, then the non-silent segment of the left-channel signal is modified by the echo.

$$w_L[n] = s_L(n) * e(n) \quad \text{........................} (3.1)$$

where wL (n)stands for a watermarked segment of the left channel signal; sL(n) stands for an original segment; and e[n] stands for an echo.

No operation is performed on the corresponding segment of the right-channel signal. If the watermark bit to be embedded is “0”, then the non-silent segment of the right-channel signal is modified by the echo. The segment is linearly convolved with the echo.

A non-silent segment on the left-channel signal is identified, if the corresponding segment on the right-channel is also a non-silent segment, then an echo is embedded in each of these two segments. The detection of an echo in both channels simultaneously indicates that the symbol embedded is a header bit. If necessary, more than one header bits may be used to give stronger indication.

3.3 Watermark Decoding Process

The watermark bits are embedded in original audio signals by embedding one type of echo in different segments of stereo signals. Hence it is only necessary to determine if an echo is present.

$$c[n] = F^{-1}(\ln(F(w(n)))) \quad \text{................................} (3.2)$$

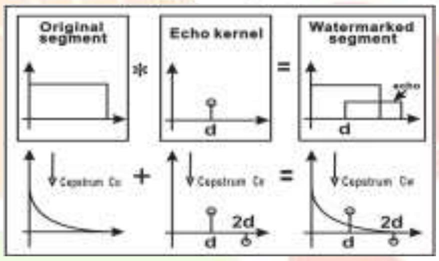


Fig 3.4 : Decoding process [2]

The decoding process is shown in the fig 3.4. The watermark decoding process starts with detecting the header. After the header is detected, Cepstrum computation is applied on each non-silent segment after the header.

If there is an impulse with 0.001s time delay detected in a segment of the left-channel signal and no impulse detected in the corresponding segment of the right-channel signal, a watermark bit “1” is recognized. If there is an impulse with 0.001s time delay detected in the segment of the right-channel signal while no impulse is detected in the corresponding segment of the left-channel, a watermark bit “0” is recognized. No information is decoded if there is no impulse detected in both segments. The final extracted watermark is the sequence of the bits recorded.

3.4 AUDIO INPAINTING

The audio inpainting framework that recovers portions of audio data distorted due to impairments such as impulsive noise, clipping, and packet loss. In this framework, the distorted data are treated as missing and their location is assumed to be known. The signal is decomposed into overlapping time-domain frames and the restoration problem is then formulated as an inverse problem per audio frame. Sparse representation modeling is employed per frame, and each inverse problem is solved using the Orthogonal Matching Pursuit algorithm.

The Signal-to-Noise Ratio performance of this algorithm is shown to be comparable or better than state-of-the-art methods when blocks of samples of variable durations are missing.

a) audio inpainting is defined as an inverse problem, based upon the concept of image inpainting;
b) a framework for audio inpainting in the time domain is proposed, based on sparse representations.

c) the orthogonal matching pursuit (OMP) algorithm for audio inpainting is adapted, in particular to deal with the properties of the Gabor dictionary;
d) a constrained matching pursuit approach is applied to significantly enhance the performance for audio declipping problems

3.5 ORTHOGONAL MATCHING PURSUIT ALGORITHM

Orthogonal Matching Pursuit (OMP) is one of the simplest ways. It is simple and greedy (with some chance to recover). In orthogonal matching pursuit (OMP), the residual is always orthogonal to the span of the atoms already selected. This results in convergence for a *d*-dimensional vector after at most *d* steps.

Conceptually, you can do this by using Gram-Schmidt to create an orthonormal set of atoms. With an orthonormal set of atoms, you see that for a *d*-dimensional vector, you can find at most *d* orthogonal directions.

It is like a discrete L1 vers ion of the technique for computing the SVD (the power method), and can be useful for many other hard optimization problems. We assume we know the measurement (d X N) matrix X, and the N measurements y.

The OMP algorithm can be stated as follows :

Step 1: Initialize the residual r_o = y and initialize the set of selected variables X(co)=φ. Let the iteration counter i=1.
Step 2: Find the variable X_{ti} that solves the maximization problem

$$\max |X_t' r_{i-1}|$$

and add the variable X_{ti} to the set of selected variables. Update $c_i = c_{i-1} \cup \{t_i\}$

Step 3 : Let $P_i = X(c_i)(X(c_i)'X(c_i))^{-1} X(c_i)'$ denote the projection onto the linear space spanned by the elements of $X(c_i)$. Update $r_i = (I - P_i)y$

Step 4: If the stopping condition is achieved, stop the algorithm. Otherwise, set i=i+1 and return to step 2

IV. RESULTS AND DISCUSSION

4.1 DATABASE DETAILS [22]

Collected a set of audio database. The recording was done at Carnegie Melon University. The room size is 6.0m(L) x 3.7m(W) x 2.8m(H) . The speaker was sitting on a chair , facing the wall. The distance to the wall is about .75m. The distance to the wall on the left side of the speakers was about 2.0m. When the recording session was done there were three computers , printer and humidifier. The voice was recorded with Compaq iPAQ 3630 built in microphone and an optimums nova 80 cross talk microphone. The channels were recorded with 11.025khz sampling frequency. The carrier signal is of duration 54 seconds and the hiding audio is of duration less than 54 seconds.

4.2 WORK DONE

4.2.1 Analysis of Audio Inpainting

Made following assumptions on the input audio signal to analyze audio inpainting technique

1) Added white Gaussian noise
2)Zero padding to introduce missing samples in the audio
3)Clipping at magnitude 0.4

Restored the missing portion using audio inpainting technique and calculated signal parameters. Worked on 20 samples with 5 audio signal in each. Calculated the average values. Here I have tabulated values of 10 samples with 5 audio each. The values are tabulated in table 4.1.

PSNR value gives an overview of the quality of the restored signal. The mean squared error measures the average of the squares of the error , that is the difference between estimator and what is estimated. The MSE values close to zeros are better. The max error signifies the maximum amount of error that can occur in a signal. The experiment is conducted on 100 samples and values are tabulated.

Table 4.1 Extracted signal parameters

SAMPLES	PSNR	MSE	MAX ERROR
SAMPLE 1	84.95	2.0778e-04	.066
SAMPLE 2	84.96	2.1773e-04	.0678
SAMPLE 3	84.89	2.0649e-04	.0684
SAMPLE 4	84.85	2.1625e-04	.0635
SAMPLE 5	84.94	2.0956e-04	.0659
SAMPLE 6	84.95	2.0260e-04	.0668
SAMPLE 7	84.78	2.1444e-04	.0691
SAMPLE 8	84.53	2.4441e-04	.1092
SAMPLE 9	84.96	2.0439e-04	.066
SAMPLE 10	84.84	2.1128e-04	.07
AVG VALUE	84.96	2.1893e-04	.071

From the above table we can infer that the signal parameters ensure the higher quality of the reconstructed signal.

The value range is as follows :

PSNR : 84 TO 85
MSE : 2.0260e-04 TO 2.1893e-04
MAX ERROR : .06 TO .1

Fig 4.1 is a sample of the audio input from the sample 1 and the signal 001.

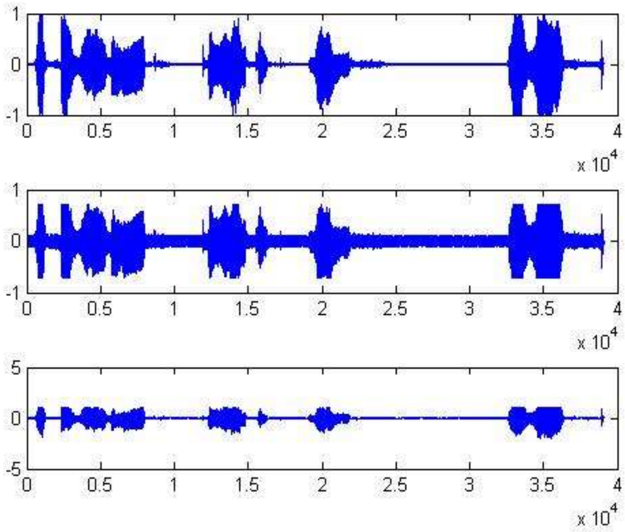


Fig 4.1 Visual representation of audio inpainting (Input signal, distorted signal, Inpainted signal respectively)

From the experiment its is clear that the recovered audio signal is of high quality . The PSNR , MSE and Max Error values are above the standard values . So audio inpainting is applied with Echo hiding to ensure the quality of the signal.

4.2.2 Audio watermarking using echo hiding and audio inapinting

The audio signal is watermarked into the carrier audio using echo hiding technique. The watermarked audio signal is decoded at the output. To improve the quality of the decoded output audio inpainting is done. This will improve the quality of the decoded signal. The audio plot is shown in the fig 4.2

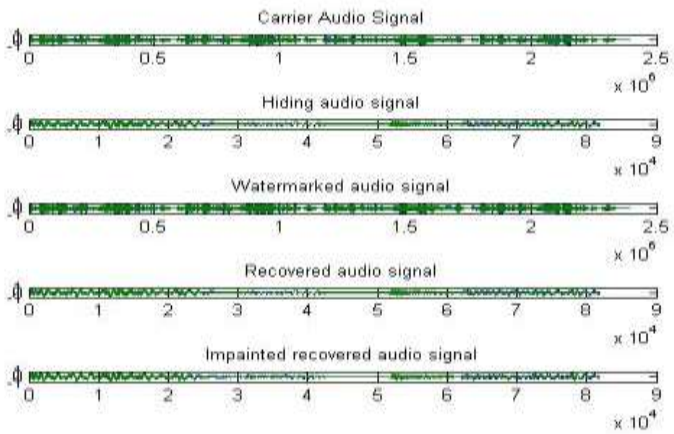


Fig 4.2 Audio signal representation (Carrier audio, Hiding audio , Watermarked audio, Recovered audio, Inpainted audio)

The pitch of each audio signals are estimated and plotted.

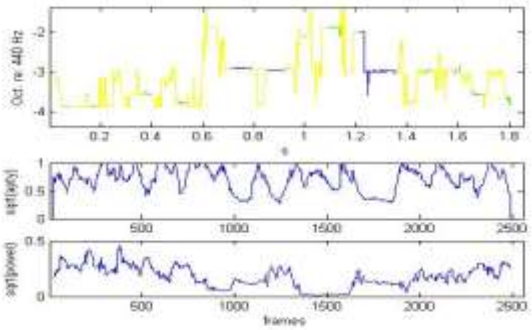


Fig 4.3 Pitch plot of carrier audio

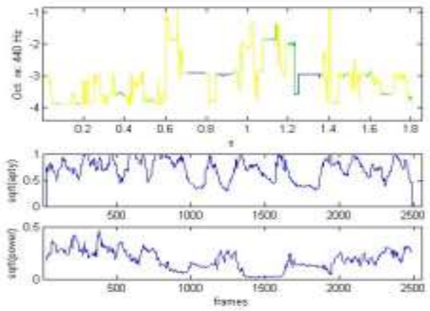


Fig 4.4 Pitch plot of hiding audio

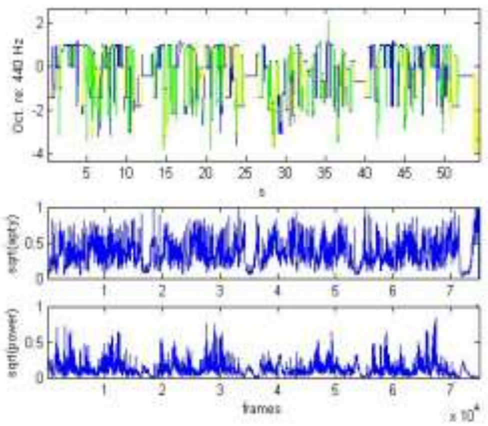


Fig 4.5 Pitch plot of watermarked audio

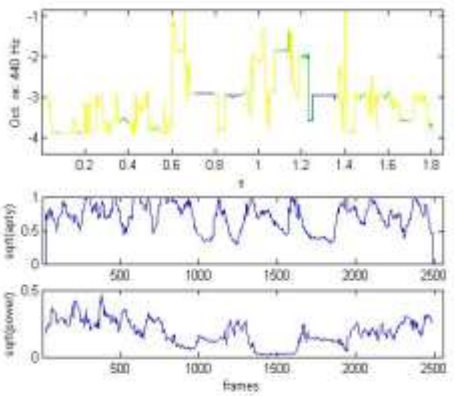


Fig 4.6 Pitch plot of decoded output

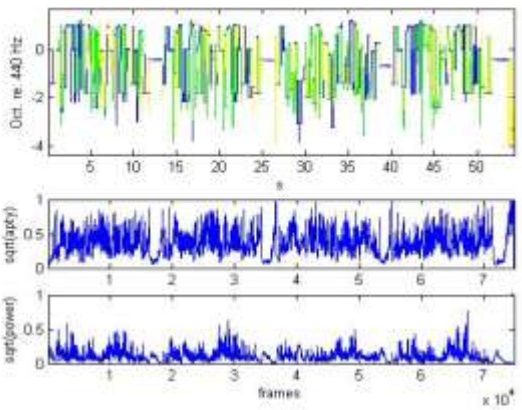


Fig 4.7 Pitch plot of inpainted output

PITCH MEASUREMENT

- Carrier audio pitch - 330.1471Hz
- Hiding audio pitch - 57.7225Hz
- Watermark audio pitch- 330.092Hz
- Recovered audio pitch - 57.7225Hz
- Recovered inpainted audio pitch - 57.7225Hz

From the result analysis Pitch remain unaltered.

Following are the values of signal parameters obtained from this experiment :

- PSNR : 80.2518
- MSE : 0.00061362
- MAX ERROR : 0.082328

- L2RAT : 0.95856

In the experiment conducted on 100 audio samples the average value of peak signal to noise ratio (PSNR) obtained is 80.2518. Peak signal-to-noise ratio, often abbreviated **PSNR**, is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. The general acceptable PSNR value ranges from 60db to 80 db. This ratio is often used as a quality measurement between the original and a compressed image. The higher the **PSNR**, the better the quality of the compressed, or reconstructed image. The Mean Square Error (MSE) and the Peak Signal to Noise Ratio (**PSNR**) are the two error metrics used to compare compression quality. The MSE value obtained from the experiment is 0.00061362 whereas the expected value is < 0.05. The max error signifies the maximum amount of error that can occur in a signal. The value obtained is less than 0.1 which implies that the quality of the output signal is high.

Pitch is a perceptual property of sounds that allows their ordering on a frequency-related scale, or more commonly, pitch is the quality that makes it possible to judge sounds as "higher" and "lower" in the sense associated with musical melodies. Pitch can be determined only in sounds that have a frequency that is clear and stable enough to distinguish from noise. Pitch is a major auditory attribute of musical tones, along with duration, loudness, and timbre. From the pitch extraction it is clear that the pitch of the input signal is not changed during the entire experiment. It ensure that the signal is not distorted and high quality is maintained.

V CONCLUSION

This project aims at developing an efficient audio security system based on Echo Hiding Watermarking and Audio inpainting. From the experiment conducted it is ensured that the signal parameter values exceeds the standard values. Thus this is an efficient method in audio security. Digital audio watermarking is more challenging as compared to watermarking other media. Majority of the schemes in literature use transform domain as they show good imperceptibility and robustness against different attacks. The proposed digital audio watermarking schemes are a small contribution to the field of digital audio watermarking. We analyzed the classical audio watermarking schemes and their limitations. The proposed scheme also shows significant improvement over the single-channel echo-hiding method in the extraction rates of watermark bits when the watermarked signals are put through various forms of attacks. The simulation results demonstrate that the proposed scheme is superior to the conventional schemes in the perceptual quality of the watermarked audio signals, the robustness against various attacks and the security in terms of preventing unauthorized users from detecting embedded data.

Using a frame-based processing of the audio signal, I have adapted the Orthogonal Matching Pursuit algorithm to address the Audio Inpainting problem. The PSNR performance of this algorithm has been shown to be comparable to or better than conventional methods.

5.1 SCOPE FOR FUTURE WORK

Audio watermarking based on Spikegram which is based on a perceptual kernel representation of audio signals can be used. The proposed method is a non blind audio watermarking method, to overcome this disadvantage spikegram based blind audio watermarking can be used. The work can be extended to real talk audio signals.

REFERENCES

[1] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, “Techniques for data hiding,” *IBM Syst. J.*, vol. 35, no. 3&4, pp. 313–336, 1996.
[2] D. Gruhl, A. Lu, and W. Bender, “Echo hiding,” in *Proc. Inf. Hiding Workshop*, Cambridge, U.K., 1996, pp. 295–315.
[3] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, “Secure spread spectrum watermarking for multimedia,” *IEEE Trans. Image Process.*,
vol. 6, no. 12, pp. 1673–1687, Dec. 1997.
[4] S. Craver, N. Memon, and M. M. Yeung, “Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks, and
implications,” *IEEE J. Sel. Areas Commun.*, vol. 16, no. 4, pp. 573–586, May 1998.
[5] D. Kirovski and H. Malvar, “Robust spread-spectrum audio watermarking,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*,2001, pp. 1345–1348.
[6]. Bender, W., et al. 1996. Techniques for data hiding. *IBM Systems Journal*. 35(3.4): 313–336.
[7] Arnold, M. 2000. Audio watermarking: features, applications, and algorithms. In *IEEE International Conference on Multimedia and Expo (II)*. Citeseer.
[8]. Yeo, I.-K., and H.J. Kim. 2003. Modified patchwork algorithm: A novel audio watermarking scheme. *IEEE Transactions on Speech and Audio Processing* 11(4): 381–386. 4. Cox, I.J., et al. 1997. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image 0Processing* 6(12): 1673–1687.

[9]. Xu, C., et al. 1999. Applications of digital watermarking technology in audio signals. *Journal of the Audio Engineering Society* 47(10): 805–812.
[10]. Oh, H.O., et al. 2001. New echo embedding technique for robust and imperceptible audio watermarking. In *2001 IEEE international conference on acoustics, speech, and signal processing, 2001. Proceedings. (ICASSP'01)*. IEEE.
[11]. Kim, H.J., and Y.H. Choi. 2003. A novel echo-hiding scheme with backward and forward kernels. *IEEE Transactions on Circuits and Systems for Video Technology* 13(8): 885–889.
[12] Ko, B.-S., R. Nishimura, and Y. Suzuki. 2005. Time-spread echo method for digital audio watermarking. *IEEE Transactions on Multimedia* 7(2): 212–221.
[13] Chowdhury, R., et al. 2016. *A view on LSB based audio steganography.*
[14]. Gopalan, K. 2003. Audio steganography using bit modification. In *ICME'03. Proceedings. 2003 International Conference on Multimedia and expo, 2003.* IEEE.
[15] Cvejic, N., and T. Seppanen. 2002. Increasing the capacity of LSB-based audio steganography. In *2002 IEEE workshop on multimedia signal processing*. IEEE.
[15] Ahmed, M.A., et al. 2010. A novel embedding method to increase capacity and robustness of low-bit encoding audio steganography technique using noise gate software logic algorithm. *Journal of Applied Sciences* 10(1): 59–64.
[16] Cvejic, N., and T. Seppanen. 2004. Reduced distortion bit-modification for LSB audio steganography. In *2004 7th international conference on signal processing, 2004. Proceedings. ICSP'04*. IEEE.
[17] S.E. Siwek, "True Cost of Recorded Music Piracy to the U.S. Economy", Published by IPI, Aug. 2007.
[18] I. Cox, M. Miller, J. Bloom, J. Fridrich and T. Kalker, "Digital Watermarking and Steganography", San Francisco, USA: Morgan Kaufmann Publishers Inc., 2nd ed., 2007.
[19] M. Steinebach and J. Dittmann, "Watermarking-based digital audio data authentication", Eurasip J. Appl. Signal Process., pp.1001-1015, 2003.
[20] A. Boho, G. Van Wallendael, A. Dooms, J. De Cock, et al., "End-To- End Security for Video Distribution", IEEE Signal Processing Magazine, vol.30, no.2, pp.97-107, 2013.
[21] S. Majumder, K.J. Devi, S.K. Sarkar, "Singular value decomposition and wavelet-based iris biometric watermarking", IET Biometrics, vol.2, no.1,
pp.21-27, 2013.
[22] http://www.speech.cs.cmu.edu/databases/pda/README.html
[23] A. Adler, V. Emiya, M. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "A constrained matching pursuit approach to audio declipping,"
in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Prague, Czech Republic, May 2011, pp. 329–332.
[24] A. Janssen, R. Veldhuis, and L. Vries, "Adaptive interpolation of discrete- time signals that can be modeled as autoregressive processes,"
IEEE Trans. Acoustics, Speech and Signal Process., vol. 34, no. 2, pp.317–330, Apr. 1986.
[25] W. Etter, "Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters,"
IEEE Trans. Signal Process., vol. 44, no. 5, pp. 1124–1135, May 1996.
[26] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration—A Statistical Model-Based Approach*. Berlin, Germany: Springer-Verlag,
1998.