Quality and Distortion Evaluation of Audio Signal by Spectrum

Er. Niranjan Singh *enggniranjan@gmail.com*
M-Tech (Computer science and engineering)
RGPV
Bhopal, 462003, India

Dr. Bhupendra Verma *bk_verma3@rediffmail.com*
Director (PG courses) (Computer science and engineering)
RGPV
Bhopal, 462021, India

Abstract

Information hiding in digital audio can be used for such diverse applications as proof of ownership, authentication, integrity, secret communication, broadcast monitoring and event annotation. To achieve secure and undetectable communication, stegano-objects, and documents containing a secret message, should be indistinguishable from cover-objects, and show that documents not containing any secret message. In this respect, Steganalysis is the set of techniques that aim to distinguish between cover-objects and stegano-objects [1]. A cover audio object can be converted into a stegano-audio object via steganographic methods. In this paper we present statistical method to detect the presence of hidden messages in audio signals. The basic idea is that, the distribution of various statistical distance measures, calculated on cover audio signals and on stegano-audio signals vis-à-vis their de-noised versions, are statistically different. A distortion metric based on Signal spectrum was designed specifically to detect modifications and additions to audio media. We used the Signal spectrum to measure the distortion. The distortion measurement was obtained at various wavelet decomposition levels from which we derived high-order statistics as features for a classifier to determine the presence of hidden information in an audio signal. This paper looking at evidence in a criminal case probably has no reason to alter any evidence files. However, it is part of an ongoing terrorist surveillance might well want to disrupt the hidden information, even if it cannot be recovered.

Keywords: Component Steganalysis, Watermarking, Audio Quality Measures, Distortion Metric

1. INTRODUCTION

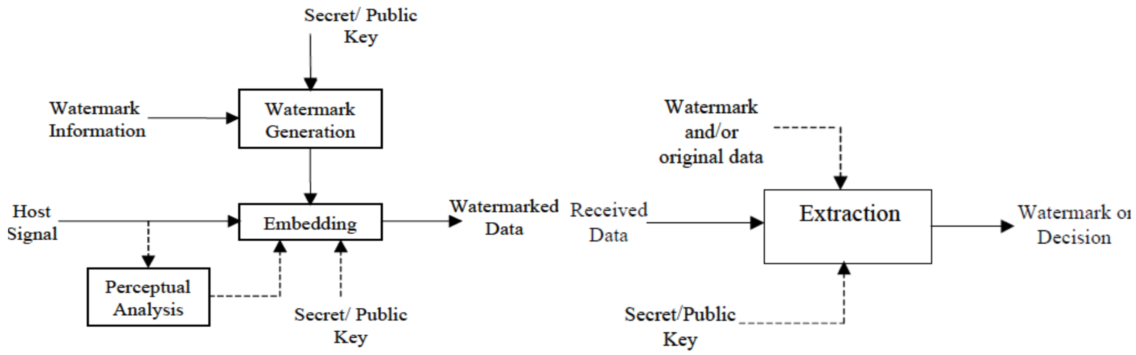
Information hiding in digital audio can be used for such diverse applications as proof of ownership, authentication, integrity, secret communication, broadcast monitoring and event annotation. There are two well-known special cases of information hiding – digital watermarking and steganography. In digital watermarking, the embedded signal depends on a secret key as the threat model includes a malicious adversary who will try to remove or invalidate the watermark. Thus the methods are denominated as "active Steganalysis" since the adversary can actively manipulate the object to alter, invalidate, and obfuscate etc. the watermark. Note that in a digital watermarking application, we always assume that the adversary knows that the content is watermarked and also knows the exact technique that is being used for watermarking. The rapid proliferation of Voice over Internet Protocol (VoIP) and other Peer-to-Peer (P2P) audio services provide vast opportunities for covert communications. By slightly altering the binary sequence of the audio samples with existing steganography tools [2], covert communication channels may be relatively easy to establish. Moreover, the inherent redundancy in the audio signal and its transient and unpredictable characteristics imply a high hidden capacity. This is further aided by the fact that the human ear is insensitive to small distortions in the audio signal.

Steganalysis itself can be implemented in either a passive warden or active warden style [3]. A passive warden simply examines the message and tries to determine if it potentially contains a hidden message. If it appears that it does, then the document is stopped; otherwise it will go through. An active warden, on the other hand, can alter messages deliberately, even though there may not see any trace of a hidden message [4], in order to foil any secret communication that can nevertheless be occurring. It should be noted that although there has been quite some effort in the Steganalysis of digital images Steganalysis of digital audio is relatively unexplored. The steganalyzer can be constructed as a "distortion meter" between the test signal and the estimated original signal, $\hat{x}$, using again de-noising.

For this purpose, one can use various "audio signal quality measures [5]" to monitor the extent of steganographic distortion. Here, we implicitly assume that the distance between a smooth signal and its de-noised version is less than the distance between a noisy signal and its de-noised version [6, 7]. The implicit assumption is that any embedding effort will render the signal less predictable and smooth. The perturbations, due to the presence of embedding, translate to the feature space, where the "audio quality features" plot in different parts of the feature space between the marked and non-marked signals. An alternate way to sense the presence of "marking" would be to monitor the change in the predictability of a signal, temporally and/or across scales.

2. PROPOSED TECHNIQUE

It has been observed that filtering an audio signal with no Watermark message causes changes in the quality metrics differently than that of an embedded audio signal [8, 9, 10, 11]. A generic watermarking scheme is shown in Figure 1 (a). The inputs consist of the watermark information, the audio input data and the watermark embedding keys to ensure security. A generic detection process is presented in Figure 1 (b). Depending on the method the original data and watermark may be used in recovery process and also depending on the method the output of recovery may be the watermark itself or some confidence measure [12], which says how likely it is for the given watermark at the input to be present in the data under processing.



(a) (b)

FIGURE 1: Generic watermarking scheme, (a) embedding, (b) recovery

1. Audibility

In order to evaluate the audibility performance of proposed method we have used a perceptual audio quality measure based on psychoacoustic sound representation (PAQM) which have high correlation with subjective measure mean opinion score (MOS) [13]. The ITU has standardized the PAQM as an objective audio quality measure system. In subjective measures the subjects are presented with original and distorted objects (in our case watermarked objects) and give scores for each audio object

2. Robustness

In the robustness experiments, the watermarked object are subjected to a variety of potential signal distortions and watermark detect statistics are computed. The Audio Stir mark [14] Benchmark has been used to simulate the signal attacks. The Benchmark has about 50 distinct distortion tools. The distortion descriptions and the parameters used are presented in Table 1. Some of the attacks such as noise addition, brumm addition and extra-stereo attacks are applied with different strengths

Attack Name	Description / Parameter
Add Brumm	Adds buzz or sinus tone to the sound / 100 to 10100
Add Dyn Noise	Add dynamic white noise to the samples / 20%
AddFFT Noise	Add white noise to the samples in the FFT room /3000
Add Noise	Adds white noise to the samples. The value "0" adds nothing and "32768" the absolute distorted maximum / 100 to 1100
Add Sinus	Adds a sinus signal to the sound file. With it, you can insert a disturb signal in the frequency band where the watermark is located / at 900Hz
Amplify	Changes the loudness of the audio file / 50 (divide the magnitude by 2)
Bass Boost	Increases the bass of the sound file.
Compressor	This works like a compressor. You can increase or decrease the loudness of quietly passages / 2.1
Copy Sample	Is like Flipp Sample but this evaluation process copies the samples between the samples / parameters are the same as Flipp-Sample
Cut Samples	Removes Remove Number (7) of samples ever Remove period (100)
Echo.	Adds an echo to the sound file
Exchange	Swaps two sequent samples for all samples
Extra Stereo	Increases the stereo part of the file / 30,50,70
FFT_HL Pass	Is like the RC-High- and RC-Low-Pass, but now in FFT room / 200 and 9000 Hz.
FFT_Invert.	Inverts all samples (real and imaginary part) in the FFT room
FFT_Real Reverse	Reverses only the real part from the FFT.
FFT_Stat1	Statistical evaluation in FFT room.
FFT_Test I	will do some tests in FFT domain.
Flipp Sample	Swaps samples inside the sound file periodically / number of flipped sample is 2000
Invert	Inverts all samples in the audio file.
LSBZero	Sets all least significant bit's (LSB) to "0" (zero).
Normalize	Normalize the amplify to the maximum value.
Nothing	This process does nothing with the audio file. The watermark should be retrieved. If not, the watermarking algorithm can be a snake oil!
Pitch Scale	Makes a pitch scale
RC-High Pass	Simulates a high pass filter build with a resistance (R) and a capacitor (C).
RC-Low Pass	Simulates a low pass filter like RC-High-Pass.

Re-sampling	Changes the sample rate of the sound file / half the sampling rate
Smooth	This smoothes the samples.
Smooth2	Is like Smooth, but the neighbor samples are voted a little bit different.
Stat1	Statistical distortion 1
Stat2	Statistical distortion 2
Voice Remove	Is the opposite to Extra-Stereo. This removes the mono part of the file (mostly where the voice is). If the file does not have a stereo part (expl. only mono) then everything will be removed.
Zero Cross	This is like a limiter. If the sample value is less the given value (threshold), all samples are set to zero / 1000
Zero Length	If a sample value is exactly "0" (zero) then it inserts more samples with the value "0" (zero) / 10 samples are included
Zero Remove	This removes all samples where the value is "0" (zero).

TABLE 1: Attacks applied by Audio Stir mark Benchmark tool

3. Comparison Tests

We have conducted some more experiments in order to compare the proposed approach with a DCT based audio watermarking technique [15], which is one of the leading non-oblivious watermarking techniques proposed in the literature. In this technique, the watermark is embedded by modifying the largest coefficients of DCT (excluding DC term). Their conjecture is that, these components are heuristically perceptually more significant than others. In the decoding phase, they use the original cover data, extract it from the received object, and compare the residual with the original watermark and make a decision

4. Regression Analysis Classifier

In the design of a regression classifier, we regress the distance measure scores to, respectively, -1 and 1, depending upon whether the audio did not or did contain a hidden message. In the test phase, the incoming audio signal is de-noised and the selected quality metrics are calculated, then the distance measure is obtained by using the predicted regression coefficients [16]. If the output exceeds the threshold 0, then the decision is that the audio contains message, otherwise the decision is that the audio dos not contain any message

5. Algorithm

The proposed feature calculation algorithm proceeds along the following steps:

- Step 1. For a given audio file x(n), apply wavelet de-noising to get its de-noised version $\tilde{X}(n)$.
- Step 2. Partition the signal x(n) and $\tilde{x}(n)$ with pre-defined segment length M. Calculate the wavelet coefficients $\hat{C}_m^P$ and C_m^P at different levels p for segment m.
- Step 3. For each wavelet decomposition level p, calculate the distortion measure H_m^P. signal spectrum.
- Step 4. Set up the feature vector V^P by calculating the moments of D^P for each wavelet decomposition level p.
- Step 5. Set up the high-dimensional feature V.
- Step 6. Generate signal spectrums of x(n) H_m^P

3. EXPERIMENTAL RESULT

In the experiments, the signal, sampled at 16 kHz, is segmented into 25 ms frames, which are weighted with a hamming window. There exists 50% overlap between segments. The tests are run for three sets of data, namely, speech, pure instrumental audio and song records. There is overall 156 speech records, 112 music excerpts and 86 instrumental records used. The speech segments have durations of three to four seconds, and recorded in acoustically shielded medium. In the audio repertoire, three different instrumental sources and three different song records are used.

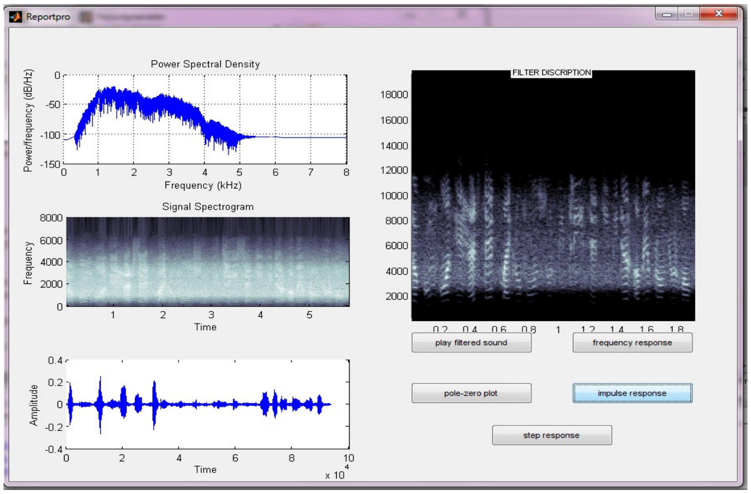


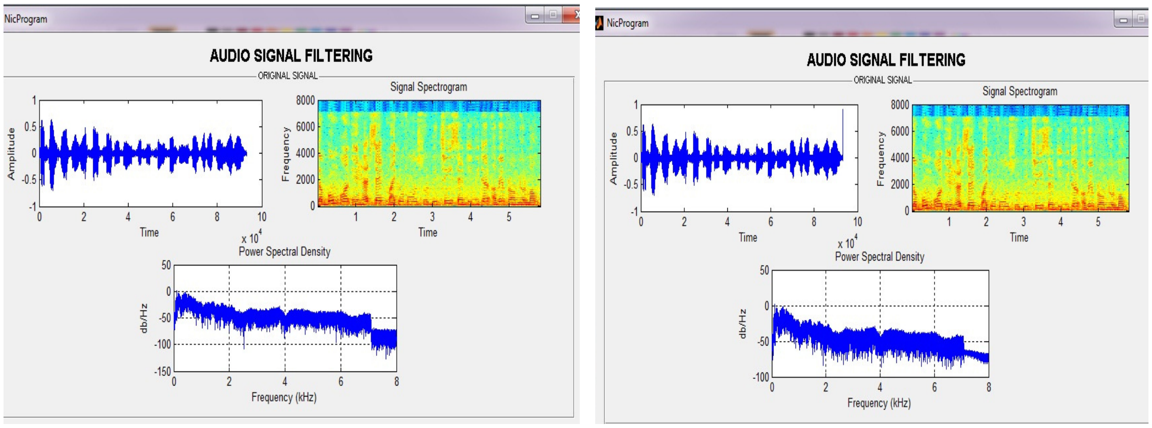
FIGURE 2: A complete report plot of original and plain audio file

The audio records (songs and instrumentals) are separated into 10-second long segments and processed as individual objects. That is for speed up the experiments because there are lots of experiments to do. We have conducted the experiments with different watermarking rates (8, 16 and 32 bits per second) on the three types of data types, which are speech, pure instrumental, and music. The attacks are applied one at a time, in other words the combined attacks are not considered.

In Figure 3, the impacts of some attacks on original wave sound are presented. In the figure 3 the attacks can be deduced from the figure that the attacks generate visible distortions and the distortions on the wave shapes can easily be observed.
We are take a plain audio file and check it signal spectrogram, frequency response, pole-zero, impulse response and step response. We also plot graph between Time and amplitude. We are not get any mixed sound or distraction.

First we have taken original audio file without any hidden massage. And apply different method to check hidden file. This section will show some examples of audio file that can hide the massage. But we detect the presence of steganography programs, detect suspect carrier files, and disrupt stegano-graphically hidden messages.

The detection of steganography file on a suspect computer is important to the subsequent forensic analysis. As the research shows, many steganography detection algorithm work best when there are clues as to the type of steganography that was employed in the first place. Finding steganography file on a computer would give rise to the suspicion that there are actually steganography files with hidden messages on the suspect computer.



(a)

(b)

FIGURE 3: The original record and attacked versions, (a) original, (b) Add Noise attack

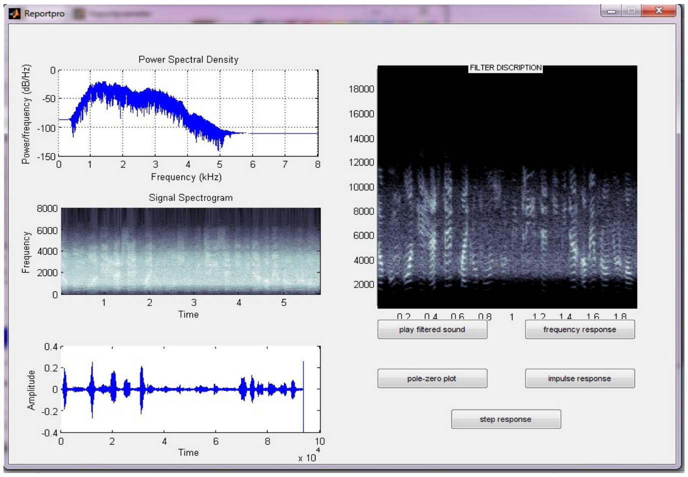


FIGURE 4: A complete report plot of audio file with hidden massage

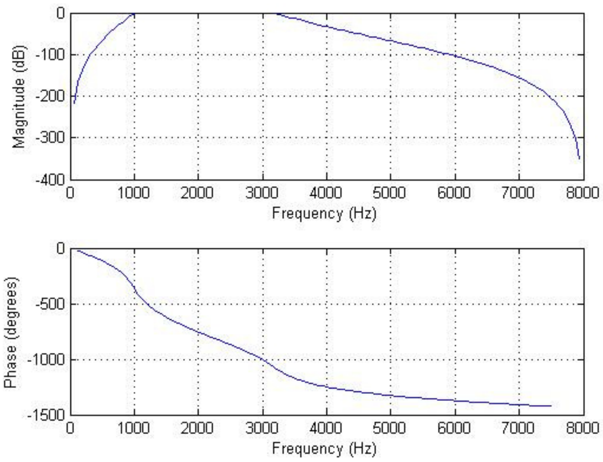


FIGURE 5: A frequency response graph plot of original and plain audio file

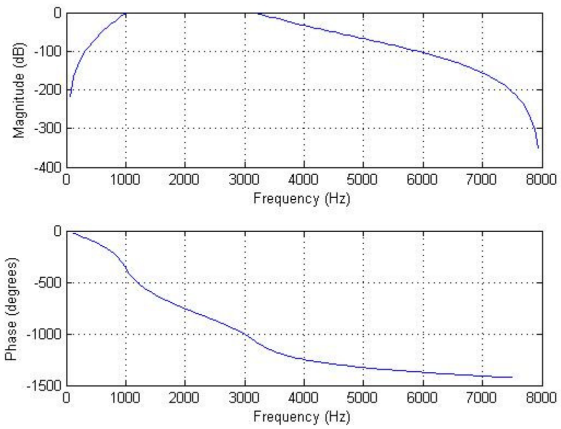


FIGURE 6: A graph plot of Frequency Response of audio file with hidden massage

This paper has been tested on 10 audio file. Our steganography algorithm was able to detect the presence of hidden messages with 65 percent accuracy with a false-positive rate less than 0.001 percent.

4. CONCLUSION

A distortion metric based on Signal spectrum was designed specifically to detect modifications and additions to audio media. We used the Signal spectrum to measure the distortion. The distortion measurement was obtained at various wavelet decomposition levels from which we derived high-order statistics as features for a classifier to determine the presence of hidden information in an audio signal.

In this study, an audio Steganalysis technique is proposed and tested. The objective audio quality measures, giving clues to the presence of hidden messages, are searched thoroughly. The distortion measurement was obtained at various wavelet decomposition levels from which we derived high-order statistics as features for a classifier to determine the presence of hidden information in an audio signal.

In this paper, an audio Steganalysis technique is proposed and tested. The audio Steganalysis algorithms exploit the variations in the characteristic features of the audio signal as a result of message embedding. Audio Steganalysis algorithms that detect the discontinuities in phase (as a result of phase coding), variations in the amplitude (as a result of Echo hiding) and the changes in the perceptual and non-perceptual audio quality metrics as a result of message embedding have been proposed. In summary, each carrier media has its own special attributes and reacts differently when a message is embedded in it. Therefore, the Steganalysis algorithms have also been developed in a manner specific to the target stegano file and the algorithms developed for one cover media are generally not effective for a different media.

5. REFERENCES

[1] Er. Niranjan Singh and Dr. Bhupendra Verma, “Steganalysis of Audio Signals, Audio Quality and Distortion Measures” *ICCET 2010 - International Conference on Computer Engineering and Technology* CET6011.0.607 ISBN No 978-81-920748-1-8.

[2] Johnson, N.F., S. Jajodia, “Steganalysis of images created using current steganography software”, in David Aucsmith (Ed.): *Information Hiding, LNCS 1525*, pp. 32-47. Springer-Verlag Berlin Heidelberg, 1998.

[3] Westfeld, A. Pfitzmann, “Attacks on steganographic systems*”, in Information Hiding, LNCS* 1768, pp. 61-66, Springer-Verlag Heidelberg, 1999.

[4] Bender, W., D. Gruhl, N. Morimoto, and A. Lu, “Techniques for data hiding”, *IBM Systems Journal,* vol. 35, no: 3&4, pp. 313-336, 1996.

[5] Kitawaki, N., H. Nagabuchi, and K. Itoh, “Objective quality evaluation for low-bit-rate speech coding systems,” *IEEE J. Select. Areas Commun.,* vol. 6, pp. 242-248, Feb. 1988.

[6] Coifman, R. R., and D. L. Donoho, “Translation-invariant denoising,” in *Wavelets and Statistics A. Antoniadis and G. Oppenheim, Eds, Springer-Verlag lecture notes*, San Diego, 1995.

[7] Yang, W, M. Dixon, and R. Yantorno, “A modified bark spectral distortion measure which uses noise masking threshold,” *IEEE Speech Coding Workshop,* pp. 55-56, Pocono Manor, 1997.

[8] Voloshynovsky, S., S. Pereira, V. Iquise, and T. Pun, “Attack modeling: towards a second generation watermarking benchmark”, *Signal Processing,* vol. 81, pp. 1177-1214, 2001.

[9] Swanson, M. D., Bin Zhu, Ahmed H. Tewfik, and Laurence Boney, “Robust Audio Watermarking Using Perceptual Masking”, Signal Processing 66, pp. 337-355, 1998.

[10] Bassia, P. and I. Pitas, “Robust Audio Watermarking in the Time Domain*”, in 9th European Signal Processing Conference (EUSIPCO’98), Island of Rhodes, Greece*, 8–11 Sept. 1998.

[11] Chen, B. and G. W. Wornell, “Quantization Index Modulation: a Class of Probably Good Methods for Digital Watermarking and Information Embedding”, *IEEE Trans. on Information Theory*, Vol. 47, No. 4, pp. 1423-1443, May 2001.

[12] Wang, S., A. Sekey, and A. Gersho, “An objective measure for predicting subjective quality of speech coders”, *IEEE J. Select. Areas Commun*., vol. 10, pp. 819-829, June 1992.

[13] Beerends, J. G. and J. A. Stemerdink, “A Perceptual Audio Quality Measure Based on a Psychoacoustics Sound Representation,” *J. Audio Eng. Soc.,* Vol. 40, pp.63- 978, Dec. 1992.

[14] Stirmark,http://amslsmb.cs.unimagdeburg.de/smfa/main.asp, 2004.

[15] Cox, I., J. Kilian, F. T. Leighton, and T. Shamoon, “Secure Spread Spectrum Watermarking for Multimedia”, *IEEE Trans. on Image Process*., Vol. 6, No. 12, pp. 1673-1687, Dec 1997.

[16] Voran, S., “Objective estimation of perceived speech quality, part I: development of the measuring normalizing block technique”, *IEEE Transactions on Speech and Audio Processing*, in Press, 1999.