An SVD-Based Audio Watermarking Technique

Hamza Özer
TUBITAK-UEKAE
P BOX 74, 41470, Gebze
Kocaeli, TURKEY
+90-262-648-12-15
hozer@uekae.tubitak.gov.tr

Bülent Sankur
Bogazici University
Dept. of Electronics, Bebek
İstanbul, TURKEY
+90-212-359-54-00/1857
sankur@boun.edu.tr

Nasir Memon
Polytechnic University
Dept. of Comp. and Inf. Science, Six Metrotech Center, 11201, NY, USA
0-718-260-3970
memon@poly.edu

ABSTRACT

We present a non-oblivious, extremely robust watermarking scheme for audio signals. The watermarking algorithm is based on the SVD of the spectrogram of the signal. The SVD of the spectrogram is modified adaptively according to the information to be watermarked. The algorithm is tested for inaudibility performance with audio quality measures and robustness tests with audio Stirmark benchmark tool, which have a variety of common signal processing distortions. The comparison with a DCT based non-oblivious based method shows that the proposed SVD based method performs very satisfactorily.

Categories and Subject Descriptors

H.0 [Information Systems]: General, Security, Copyright.

General Terms

Algorithms, Security.

Keywords

Watermarking, singular value decomposition.

1. INTRODUCTION

Audio watermarking finds applications in various areas such as copyright protection, data authentication, covert communications, addition of metadata, content identification and captioning or labeling of data [7, 11, 20]. In copyright application, the embedded data is intended to be used as a proof of digital media ownership to resolve copyright disputes. Authentication aims to check any evidence of tampering, hence to verify the integrity of the data. Covert communication is the exchange of messages secretly as embedded within host data, in such a way that it does not raise any suspicion that a secret message is being communicated. Insertion of metadata has the purpose of enriching the use of the audio document, for example, it may contain a perceptual hash for database indexing. Obviously these diverse applications have differing robustness, data capacity and imperceptibility requirements [7]. For example, the ability to survive vis-à-vis casual signal processing operations and malicious attacks varies from very low in fragile watermarking to very high in proof of ownership applications. The data embedding capacity similarly varies from one bit per file as in access and copy/not copy applications to one bit per sample, as in covert communication.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MM-SEC'05, August 1–2, 2005, New York, New York, USA.
Copyright 2005 ACM 1-59593-032-9/05/0008...$5.00.

There are several robust audio watermarking methods in the literature. For example, among the time-domain schemes [3, 10, 18, 1, 5, 19, 13], Chen [5] uses quantization index modulation in order to embed the watermark, by selecting one of the two quantizers according to the watermark bit. Other time-domain embedding methods [3, 10, 18, 1, 19] embed watermark by adding some weak noise signal into audio signal, which is modulated by the watermark bit polarity. The Human Auditory System (HAS) [18, 1, 19, 13] is often taken into account and the watermark signal is shaped by a masking function. Finally Gruhl [10] embeds some delayed and attenuated versions of the original samples, where the amount of delay determines the watermark information.

In the transform-domain category, the watermark is embedded into selected transform coefficients [3, 6, 8, 12, 15]. Bender obtains a robust system by modulating the phase spectrum with the watermark signal [3], which, however is observed to produce a small perceptible noise. Cox develops a non-oblivious method by embedding the watermark into most energetic coefficients of DFT or Discrete Cosine Transform (DCT) [6]. In contrast, Lu embeds the watermark into DFT coefficients in an oblivious scheme after pre-shaping it by the Just Noticeable Difference (JND) threshold [15]. Garcia [8] and Kirovski [12] embed HAS-shaped watermark into Short-Time Fourier Transform (STFT) and Modulated Complex Lapped Transform (MCLT) coefficients, respectively. In this method it is always accepted the received signal has a watermark.

In either case, the human auditory system must be taken into account in audio watermarking. It has been observed that HAS has wider dynamic and differential range as compared to other human senses. HAS perceives over a range of power up to one billion to one, and a range of frequencies greater than a thousand to one. It is also quite sensitive to additive random noise. However, despite the large dynamic range of HAS, it actually has a fairly small differential range, and while it is sensitive to amplitude and relative phase, it is unable to perceive absolute phase. As a result, there are some environmental distortions (these distortions can be used for hiding data), which are ignored by the listener.

In this work we propose a semi-oblivious, extremely robust watermarking scheme for audio signals. The watermarking algorithm is based on the Singular Value Decomposition (SVD) of the spectrogram of the signal. The time-frequency of the audio signal is estimated and the resulting magnitude spectrogram is treated as a two-dimensional image or a matrix. The SVD of this matrix provides a medium to embed a 2D watermark pattern directly. In order to ensure the inaudibility (to guarantee that the modifications are below the HAS hearing level) the embedding watermark message is shaped with singular values of original/host audio signal, thus the embedding watermark is modified adaptively with embedded coefficients.

SVD has been employed before for different image applications such as compression, hash extraction and image watermarking. In image watermarking applications [14, 9, 4], the singular values of the host image are adapted in order to embed the watermark. Among alternate techniques, singular values (SVs) of the watermark image is added the singular values of host image in [4]. In [14] the watermark object is first multiplied with the SVs of original image, then decomposed again into its SVs, and finally the thus-modified diagonal matrix is inserted back in the host image. Gorodertski, on the other hand, quantizes the SVs of host image according to watermarking bits [9]. Watermarking of the time-frequency distribution has been considered by Selin [21] in the context of Wigner-Ville distribution.

In our study we first convert the audio sample into a matrix form by using short-time Fourier transform (STFT), obtain its SVD decomposition, and then adaptively modify SVD coefficients with watermark bits. Thus the watermark is embedded in the singular values of STFT coefficients of the host signal.

The rest of the section is organized as follows. Section 2 presents the audio watermarking method. The experiments, conducted to test audibility and robustness of the proposed algorithm are discussed in Section 3. The conclusions are drawn in Section 4.

2. THE AUDIO WATERMARKING METHOD

2.1 Watermark Embedding Method

In our watermark embedding, we first decompose the STFT of the audio signal, organized as a time-frequency matrix, into its singular values, and then mark the singular value matrix. The singular value matrix *D* of the host object is modified according to the watermarking pattern and bit polarity. The general block diagram of embedding procedure is presented in Figure 1.

The watermark is a coded binary sequence and a pseudo-random signal is used to spread the power of every bit to a wide spatial support range. The same pattern is needed to detect the watermark binary string. Therefore the seed of the random sequence can be thought as the secret embedding key.

The audio signal is first converted into a matrix form by STFT. The STFT is a time-frequency analysis that extracts the frequency spectrum of the signal through short-time windows [16]. The STFT is chosen because it is the simplest, invertible time-frequency analysis technique, although other transforms can also be employed. The analysis and reconstruction equations of the STFT are as follows:

$$STFT_x(t,f)=\int x(\tau)g(\tau-t)e^{-j2\pi f\tau}d\tau$$

$$x(\tau)=\iint STFT_x(t,f)g(\tau-t)e^{j2\pi f\tau}dtdf$$

where *g(t)* is some window function. By sliding the function *g(t)* over the signal *x(t),* multiplying them and calculating Fourier transform of the product, we get a two-dimensional representation of the signal. The record of audio signal is processed in overlapping segments. Each frame is windowed to reduce edge artifacts, and then its Discrete Fourier transform (DFT) is calculated. Since STFT is carried out in discrete time and frequency cells, we can consider this density in matrix form (called in the sequel the STFT matrix). A size *FxM* matrix, called the STFT matrix, is obtained, where F is the number of frames, which depends on signal length, and M is the frame size. The phase of the STFT coefficients is preserved, while its magnitude is modified to embed watermark.

The singular value decomposition is a numerical tool, which effectively decomposes a matrix into two orthogonal matrices and its singular values. Thus a matrix *A* is decomposed into $A = U D V^T$, where *A* is the *FxM* matrix that we want to summarize, *D* is *FxM* matrix with only min(F,M) diagonal elements, *U* is an *FxF* orthogonal matrix, and *V* is an *MxM* orthogonal matrix. One useful property of the SVD is that the singular values are invariant under orthogonal transformations.

We obtain the SV decomposition of the STFT matrix of each frame, and then embed the watermark bits in the *D* singular value matrix. Thus each STFT matrix *A*, carries a watermark bit or the STFT matrices form the footprints of bits. The size of the matrices (in turn related to the frame size) determines the watermark payload, and we have experimented with different payload rates.

We first decompose a block *A* via SVD to a diagonal matrix form, where *D* contains zeroes in the off-diagonal positions. Then the watermark carrier $W=\{w(i,j)\}$ (an *FxM* matrix with random noise-like elements) is added with a scaling or strength factor a as follows:

$$w_D(i,j)=\delta_i \;+\; ab\delta_i w(i,j) \quad for \;\begin{cases} i=1,2,\ldots,F \\ j=1,2,\ldots,M \end{cases}$$

where $\delta_i;\ i=1\ldots\min(M,F)$ are the singular values of the matrix *A* (diagonal elements of D), $b\in\{0,1\}$ is the polarity of the watermark sequence to be embedded, $w(i,j);\ i=1\ldots F,\ j=1\ldots M$ represent the random carrier elements, and *a* is the watermark strength factor. Before embedding, the watermark is also shaped by the original singular value coefficients in order to enable object-dependent watermarking. In other words, we adaptively scale the watermark strength according to the characteristics of the cover medium with perceptual concerns. The resulting watermark matrix is subjected to a new SVD operation and it results in the new set of left-, right-singular vectors and SV matrix.: $W_D=U_W D_W V_W^T$. Finally, the watermarked message block, A_W , is obtained by inverse SVD operation ($A_W=UD_WV^T$). In other words, the message block (the STFT block) is reconstituted with its original right and left eigenvectors (the columns and rows of matrices *U* and *V*) and the new non-diagonal matrix D_W . The matrices U_W , V_W and D_W

must be preserved for detection. The embedding steps can be summarized as follows:

$$\Rightarrow \quad A = UDV^T$$

$$\Rightarrow \quad w_D(i,j) = \delta_i \; + \; ab\delta_i w(i,j) \qquad for \quad \begin{cases} i = 1,2,\ldots,F \\ j = 1,2,\ldots,M \end{cases}$$

$$\Rightarrow \quad W_D = U_W D_W V_W^T$$

$$\Rightarrow \quad A_W = UD_W V^T$$

In the last step, the time-frequency plane is tiled back with the watermarked magnitude components and the original phase. The watermarked audio signal results from the inverse STFT operation.

2.2 Watermark Detection

The received audio signal is transformed into matrix form and blocked in the same way. It is assumed that the matrices *Uw, Vw, D* and the key to the pseudo-random signal are all known at the receiver side. For an analyzed object *A'*, the detection/synthesis procedure is the reverse of the embedding/analysis procedure, as follows:

$$\Rightarrow \quad A' = U' D'_W V'^T$$

$$\Rightarrow \quad W'_D = U_W D'_W V_W^T$$

$$\Rightarrow \quad W' = \frac{D^{-1}\left(W'_D - D\right)}{a}$$

Eventually the received *W'* is compared with the key signal, in other words, the pseudo random signal *W*. We have used the normalized correlation as similarity measure: if the inner product, that is, the term-by-term multiplication of the two matrices

$W' \bullet W = \sum_{i,j} w_{ij} w'_{ij}$, is positive then one decides for bit 1, otherwise for bit -1. The viability of the scheme is illustrated in the detector outputs of the correlation receiver as in Figure 2. In order to test the behavior of the correlation receiver, the extracted watermark message compared with 1000 different random signals and similarity scores in the presence of four distortion types (distortions are described in Section 3.2) are plotted. It can be observed that the response due to correct watermark key is much above the impostor sequences.

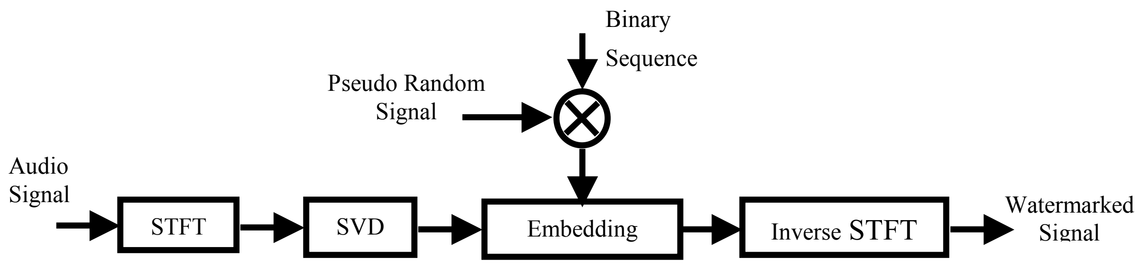

Figure 1. SVD-based audio watermarking procedure

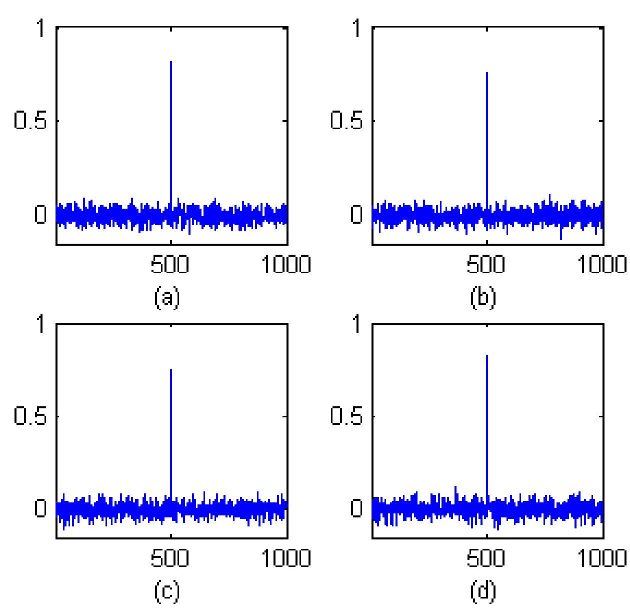

Figure 2: Detector response to 1000 randomly generated watermarks: the abscissa denotes the detector response, the indexes of 500 denotes the watermarked objects after the attacks (a) copysample, (b) fft_HLPass, (c) flipsample, and (d) zerocross.

3. EXPERIMENTAL RESULTS

We performed extensive experiments in order to test the imperceptibility and robustness characteristics of the proposed audio watermarking method. The compromises between audibility of watermarking artifacts and robustness requirements have been discussed in several papers [12, 13].

In the experiments we used speech signal sampled at 16 kHz and music signals, which were originally at 44100 kHz rate, converted down to 16 kHz sampling rate. These signals are segmented into 25 ms frames, which are weighted with a hamming window, with 50% overlap between segments. The tests are run for three sets of data, namely, speech, pure instrumental audio and song records. Overall 200 speech records, 142 music excerpts and 105 instrumental records are 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. The music records are taken from the songs of famous music groups U2 and Rolling Stones. The songs are ‘One’ (a slow song), ‘Even Better Than The Real Thing’ of U2 and ‘Paint It, Black’ of Rolling Stones. The audio records (songs and instrumentals) are separated into 10-second long segments and processed as individual objects.

3.1 Audibility Tests

In order to evaluate the audibility performance of the proposed method, we have used a perceptual audio quality measure based on psychoacoustic sound representation (PAQM). PAQM is claimed to be highly correlated with subjective measures such as the mean opinion score (MOS) [2]. In fact, Beerends has shown that the correlation between PAQM and MOS is about 0.98 [2]. International Telecommunication Union (ITU) has standardized PAQM as an objective audio quality measure system. We have used the ITU implementation tool of perceptual speech quality measure to develop the PAQM function. Recall that the grading scale of MOS is as follows: 5.0 for imperceptible, 4.0 for perceptible but not annoying, 3.0 for slightly annoying, 2.0 for annoying, 1.0 for very annoying. We have adjusted the watermark strength a to achieve satisfactory audibility scores, chosen as 0.15. This yields PAQM scores of about 0.01, its MOS equivalent being about 4.7, which, in turn is nearly imperceptible.

3.2 Robustness Tests

In the robustness experiments, the watermarked objects are subjected to a variety of attacks and watermark detection statistics are computed. The Audio Stirmark Benchmark [17] has been used to simulate the signal processing attacks. The Benchmark has about 50 distinct distortion tools. The distortions are used with default parameters given in the batch file of the packet. Some of the attacks are noise addition, filtering, compression, sample addition/cutting/flipping, pitch scale changing, statistical attacks, resampling, echo addition, smoothing, zero sample addition/removing etc.

We have conducted the experiments with different watermarking rates (8, 16 and 32 bits per second) on the three types of data, namely speech, pure instrumental and complex music. The attacks are applied one at a time, in other words the combined attacks are not considered.

In Fig. 3, the impacts of some attacks on original wave sound are presented. In this figure, copysample, flipsample, fft_hlpass, and zerocross attacks are shown. One can observe in Fig. 3 that these attacks generate visible waveform distortions.

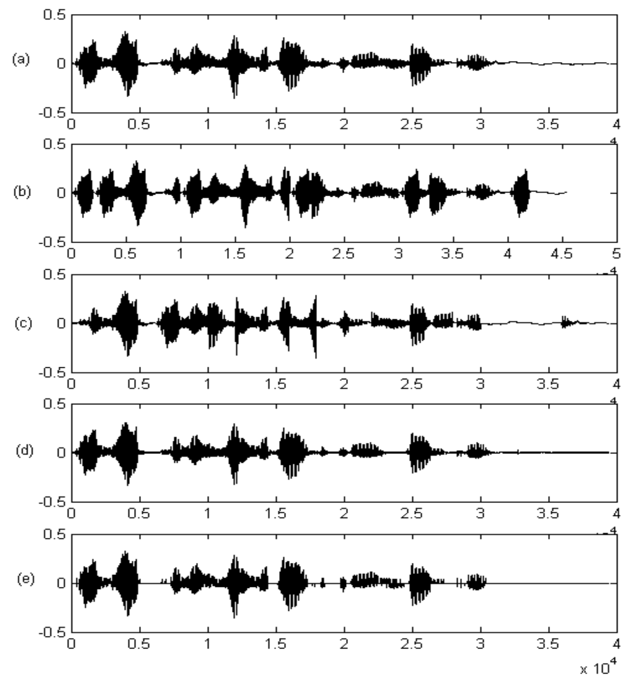

Figure 3. The original record and attacked versions, (a) original, (b) copysample attack, (c) flipsample attack, (d) fft_hlpass attack, (e) zerocross attack.

The watermark detection performance results are given in Table 1, where the misdetection percentage of watermark (Bit Error Rate: BER) is given. One can conclude from Table 1 that the proposed method works satisfactorily at embedding rates of 8 and 16 bps for speech, at rates of 8, 16, and 32 bps for music. Actually for a given attacking strength, the resulting distortions on different records (speech, music etc) are different. For example, the objective quality measure (PAQM) scores after zerocross attack are 0.08 (MOS equivalent is about 3.9) for speech records, 0.002 (MOS equivalent is about 4.9) for pure instrumental records, and 0.0022 (MOS equivalent is about 4.8) for music records, which in turn is the reason for different detection performances. A better approach would have been to gauge the attack strength to result in the same PAQM scores for music and speech and possibly accompanied by subjective tests, and then run the watermark detection tests. Nevertheless, we leave it for a future study due to the length of the experimental procedure involved, and we give here fixed-strength, variable-PAQM performance results.

Table 1. Bit error rates after the attacks applied by Audio Stirmark Benchmark. The total number of watermark bits tested is 12800 for speech and 22720 for music data.

Attack Name	Speech Data			Pure Instrumental			Music		
	8 bps	16 bps	32 bps	8 bps	16 bps	32 bps	8 bps	16 bps	32 bps
AddBrumm	0	0	0	0	0	0	0	0	0
AddDynNoise	0	0	0	0	0	0	0	0	0
AddFFTNoise	0	0	0	0	0	0	0	0	0
AddNoise	0	0	0	0	0	0	0	0	0
AddSinus	0	0	0	0	0	0	0	0	0
Amplify	0	0	0	0	0	0	0	0	0.75
BassBoost	0	0	0	0	0	0	0	0	0
Compressor	0	0	0	0	0	0	0	0	0
CopySample	2	4	5	0	0	0.75	0	0	0.5
CutSamples	0	1	3	0	0	0	0	0	0
Echo	0	0	0	0	0	0	0	0	0
Exchange	0	0	0	0	0	0	0	0	0
ExtraStereo	0	0	0	0	0	0	0	0	0
FFT_HLPass	0	1	2	0	0	0	0	0	0
FFT_Invert	0	0	0	0	0	0	0	0	0
FFT_RealRev	0	0	0	0	0	0	0	0	0
FFT_Stat1	0	0.5	2	0	0	0	0	0	0.5
FFT_Test	0	0.25	1.5	0	0	0	0	0	0.4
FlippSample	1	1	2.5	0	0	0	0	0	0.75
Invert	0	0	0	0	0	0	0	0	0
LSBZero	0	0	0	0	0	0	0	0	0
Normalize	0	0	0	0	0	0	0	0	0
Nothing	0	0	0	0	0	0	0	0	0
PitchScale	0	0	0	0	0	0	0	0	0
RC-HighPass	0	0	0	0	0	0	0	0	0
RC-LowPass	0	0	0	0	0	0	0	0	0
Resampling	0	0	0	0	0	0	0	0	0
Smooth	0	0	0	0	0	0	0	0	0
Smooth2	0	0	0	0	0	0	0	0	0
Stat1	0	0	0	0	0	0	0	0	0
Stat2	0	0	0	0	0	0	0	0	0
VoiceRemove	0	0	0	0	0	0	0	0	0
ZeroCross	3	3.75	6	0	0	0	0	0	0
ZeroLength	0	0	0	0	0	0	0	0	0
ZeroRemove	0	0	0	0	0	0	0	0	0
Average of all attacks	0.17	0.31	0.63	0	0	0.02	0	0	0.1

3.3 Comparison Tests

We have conducted additional experiments in order to compare the proposed approach with a DCT-based audio watermarking technique [6], which was a direct adaptation of Cox's oblivious image watermarking. In this technique, the watermark is embedded by spreading watermark bits to the largest DCT coefficients (excluding DC term) of the audio segments. In other words, one takes the DCT transform of the audio segments, ranks them in magnitude excluding the DC term, and marks a number of DCT coefficients commensurate with the proposed method. Therefore, the bit footprints in the proposed STFT-SVD method and the Cox' method for audio are equal in size, and the embedding strength is adjusted to achieve the same PAQM figure. In the decoding phase, the original cover data is used. The experiments were conducted on speech and music data at the embedding rate of 32 bps with the same embedding and attack strengths as the SVD-based algorithm. In both embedding methods (the proposed SVD-based one and DCT-based method) the embedding strength is adjusted to give approximately the same inaudibility score. More specifically, the inaudibility measured with the PAQM has a MOS equivalent score of 4.7, which is nearly imperceptible. The results, in term of percentage bit error scores, are tabulated in Table 2.

Table 2: Comparison results of the DCT and SVD based methods (32 bit/s). Overall 25640 bits added to the speech and 45440 bits to the music signals.

Attack Name	Speech Data Set		Music Data Set	
	SVD Based M.	DCT Based M.	SVD Based M.	DCT Based M.
AddBrumm	0	0.995	0	1.25
AddDynNoise	0	0	0	1.56
AddFFTNoise	0	50.34	0	51.25
AddNoise	0	0	0	0.78
AddSinus	0	4.97	0	0.77
Amplify	0	49.6	0.75	52.32
BassBoost	0	0	0	0
Compressor	0	0	0	0
CopySample	5	100	0.5	100
CutSamples	3	100	0	100
Echo	0	48.96	0	23.43
Exchange	0	0	0	0
ExtraStereo	0	0	0	0
FFT_HLPass	2	0	0	0.31
FFT_Invert	0	49.65	0	52.6
FFT_RealReverse	0	0	0	0.78
FFT_Stat1	2	39.31	0.5	19.84
FFT_Test	1.5	35.44	0.4	19.80
FlippSample	2.5	15.42	0.75	21.66
Invert	0	48.75	0	52.42
LSBZero	0	0	0	0
Normalize	0	51.24	0	0
Nothing	0	0	0	0
PitchScale	0	0	0	0
RC-HighPass	0	2.48	0	2.03
RC-LowPass	0	0	0	0
Resampling	0	0.41	0	0.62
Smooth	0	0	0	0
Smooth2	0	0	0	0
Stat1	0	0	0	0
Stat2	0	0	0	0

Table 2 (Continued): Comparison results of the DCT and SVD based methods (32 bit/s). Overall 25640 bits added to the speech and 45440 bits to the music signals				
VoiceRemove	0	49.7	0	52.1
ZeroCross	6	0	0	0
ZeroLength	0	100	0	60.5
ZeroRemove	0	59.6	0	100
Average of all	0.629	23.03	0.09	20.4

It can be observed that the DCT-based method (adaptation of Cox [6]) fails in the case of sample duplication-type of attacks, such as copying and cutting of randomly selected samples, removing the zero valued samples or adding more zero valued samples. Moreover it has quite poor performance in the case of amplitude-type distortions, such as amplifying, noise addition in the FFT domain, echo addition, inverting the spectral or time domain samples, and voice removing. On the other hand, the DCT-based method performs slightly better only in the DFT-domain high-pass and low-pass filtering attack and in the thresholding-to-zero (ZeroCross) attack. Let's recall that SVD-based method comes at the price of storing the right- and left eigenvectors (*U, V* matrices in Eq. 2) . We can offer some explanation to this performance differential between SVD-based and DCR-based watermarking schemes. We have observed that in the case of sample deletion and addition attacks, the envelope of the DCT coefficients change, while in the case of the FFT spectrum one observes only a scaling of the spectrum, but not any spectral envelope distortion.

4. CONCLUSIONS

A novel audio watermarking method is proposed based on embedding in the spatio-temporal domain. The method uses the STFT variety of time-frequency decomposition for audio signals, and modifies its singular value matrix with a watermark matrix before re-constituting the signal. The method is semi-blind, in that there is no need to know the watermark sequence at the receiver, but the left and right singular vectors as well as the modified singular value matrix must be available at the receiver.

Experiments have shown that the inaudibility and robustness performance goals can be achieved. When the audio Stirmark benchmark tool is used to evaluate the robustness performance against signal distortions, the algorithm outperformed definitely its competitors.

The method is presently semi-blind and needs storage amount commensurate with the audio file. Alternative spectral domain techniques, where there is no need to store the transform matrices is being investigated. Similarly, time-frequency analysis tools other than STFT should be considered.

5. REFERENCES

[1] P. Bassia 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.

[2] J. G. Beerends, J. A. Stemerdink, A perceptual audio quality measure based on a psychoacoustics sound representation, *J. Audio Eng. Soc.,* vol. 40, pp. 963-978, Dec. 1992

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

[4] D. V. S. Chandra, Digital image watermarking using singular value decomposition, *Proceedings of 45th IEEE Midwest Symposium on Circuits and Systems,* Tulsa OK, pp. 264-267, August 2002.

[5] B. Chen 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.

[6] I. J. Cox, 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, December 1997.

[7] J. Dittmann, M. Steinbach, T. Kunkelmann, and L. Stoffel, Watermarking for media: Classification, quality evaluation, design improvements, *In Proceedings ACM Multimedia 2000,* Marina Del Rey, USA, pp. 107-110, November 2000.

[8] R. A. Garcia, Digital watermarking of audio signals using a psychoacoustic auditory model and spread spectrum theory, *107th Convention: Audio Engineering Society,* New York, 1999.

[9] V. I. Gorodertski, L. J. Popyack, V. Smailov, and V. A. Skormin, SVD-based approach to transparent embedding data into digital images, *MMM-ACNS2001,* St. Petersbug, Russia, May 2001.

[10] D. Gruhl, A. Lu, and W. Bender, Echo hiding, *Information Hiding 1st International Workshop,* June 1996.

[11] S. Katzenbeisser and F. A. P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, 2000.

[12] D. Kirovski and H. Malvar, Spread-spectrum watermarking of audio signals, *IEEE Transaction on Signal Processing,* vol. 51, No.4, pp. 1020-1033, Appril 2003.

[13] A.N. Lemma, J. Aprea, W. Oomen, and L. Kerkhof, A temporal domain audio watermarking technique, *IEEE Transaction on Signal Processing,* Vol. 51, No.4, pp. 1088-1097, Appril 2003.

[14] R. Liu and T. Tan, A SVD-based watermarking scheme for protecting rightful ownership, *IEEE Transaction on Multimedia,* 4(1), pp. 121-128, March 2002.

[15] C. S. Lu, H. Mark, and L. H. Chen, Multipurpose audio watermarking, *Proc. of the 15th IAPR : Int. Conf. on Pattern Recognition,* Barcelona, 2000.

[16] S. Qian and D. Chen, *Joint Time-Frequency Analysis,* Prentice Hall, 1996.

[17] Stirmark, http://amsl-smb.cs.uni-magdeburg.de/smfa/main.php.

[18] M. D. Swanson, Bin Zhu, Ahmed H. Tewfik, and Laurence Boney, Robust audio watermarking using perceptual masking, *Signal Processing 66,* pp. 337-355, 1998.

[19] M. Veen, F. Bruekers, J. Haitsma, T. Kalker, A. N. Lemma, and W. Oomen, Robust, multi-functional and high-quality audio watermarking technology, *Audio Engineering Society 110th Convention,* May 2001.

[20] V. Venkatachalam, L. Cazzanti, N. Dhillon, and M. Wells, Automatic identification of sound recordings, *IEEE Signal Processing Magazine,* pp. 92-99, March 2004.

[21] M. Al-Khassaweneh and S. Aviyente, "Robust Watermarking on the joint spatial-spectral domain", in *Proceedings of IEEE DSP Workshop,* August 2004.