AUDIO PREWHITENING BASED ON POLYNOMIAL FILTERING FOR OPTIMAL WATERMARK DETECTION

Nedeljko Cvejic, Tapio Seppänen

MediaTeam
Information processing laboratory
FIN-90014 University of Oulu, Finland
Email: {cvejic, tapio.seppanen}@ee.oulu.fi

ABSTRACT

Watermark detection algorithms in the spread-spectrum audio watermarking systems would have optimal performance if host audio had properties of additive white Gaussian noise. However, statistical properties of the host audio signal are generally different from properties of AWGN. Therefore, there is a need for decorrelation of watermarked audio in order to achieve optimal detection based on correlation. In this paper, we describe whitening procedure for audio using Savitzky-Golay FIR filter, based on polynomial fitting of data. Residual signal has more Gaussian-like distribution and significantly smaller variance compared to the case of unprocessed watermarked audio, which was verified using two hypothesis tests for data distribution normality. The procedure considerably improves detection results and has higher resistance to various watermark attacks in comparison with standard correlation detection.

1. INTRODUCTION

Digital watermarking is a process that embeds an imperceptible and statistically undetectable signature to multimedia content (e.g. images, video and audio sequences). Embedded watermark contains certain information (signature, logo, ID number…) related uniquely to the owner, distributor or the multimedia file itself. Thereby, multimedia data authors and distributors are able to prove ownership of intellectual property rights without need to forbid other individuals to copy multimedia content itself. Watermarking algorithms were primarily developed for digital images and video sequences; interest and research in audio watermarking started slightly later.

In the past few years, several algorithms for embedding and extraction of watermarks in audio sequences have been presented. All of the developed algorithms take advantage of perceptual properties of the human auditory system (HAS), foremost occurrence of masking effects in the frequency and time domain, in order to add watermark into a host signal in a perceptually transparent manner. Embedding of additional bits in audio signal is a more tedious task than implementation of the same process on images and video, due to the dynamical superiority of HAS in comparison with the human visual system. The basic approach to watermarking in the time domain is to embed a pseudo-random noise (i.e. the PN sequence) into the host audio by modifying the amplitudes accordingly. Information modulation is usually carried out using spread-spectrum technique that augments a low-amplitude SS sequence, which is detected by correlation receiver.

Recently, we have developed a spread-spectrum audio watermarking algorithm in time domain for copyright protection of digital audio [9]. The procedure uses a time domain embedding algorithm and properties of spread spectrum communications as well as temporal and frequency-domain masking in HAS. Matched filter technique, based on auto-correlation of embedded PN sequence, is optimal in the sense of Signal to Noise Ratio (SNR) in additive white Gaussian channel. However, the host audio signal is generally far from the additive white Gaussian noise and it leads us to the optimal detection problem using pre-processing of audio by decorrelation of audio samples before detection.

In the present paper, we propose an audio decorrelation algorithm for spread-spectrum watermarking that significantly improves the robustness of watermark detection and demonstrate high resistance to attacks. To remove the cross-correlation of the adjacent samples of host audio, we use residual signal of watermarked audio processed by Savitzky-Golay FIR filter (digital smoothing polynomial filters) and then perform correlation watermark detection.

2. WATERMARK EMBEDDING

Figure 1 gives a general overview of the developed watermark-embedding algorithm. The embedding scheme proposed in this paper modifies the original audio signal, which is represented as a 16-bit sample sequence sampled at 44100 Hz, mono. The PN sequence is obtained from a PN-generator and represented in the bipolar form {-1,1}. Prior to further processing, the PN sequence is filtered in order to adjust it to masking thresholds of the human auditory system (HAS) in the frequency domain. The main goal is to adapt the watermark to such form that the energy of the watermark is maximized under the restriction of keeping auditory distortions to a minimum, although the SNR value is significantly decreased. The frequency characteristic of the filter is an approximation of the threshold in quiet curve of the HAS [4]. Despite the simplicity of the shaping process of the PN sequence in frequency domain, the result is an inaudible watermark as the largest amount of the shaped watermark's power are concentrated in the frequency sub-bands with the lowest HAS sensitivity. In addition, these frequency sub-bands (frequencies below 500 Hz and above 11000Hz) are an essential part of the watermarked audio and cannot be removed from its spectrum without making serious decrease in the perceptual quality. A significant number of computational operations needed for frequency analysis of audio, which have to be run in order to derive global masking thresholds in a predefined time window, are skipped, making this scheme appreciably faster. Although standard frequency analyses obtain more accurate evaluation of perceptual thresholds, simulation tests done with selected audio clips showed a high level of similarity between frequency masking thresholds derived from ISO-MPEG Audio Psychoacoustic Model [4] and our frequency-shaping model.

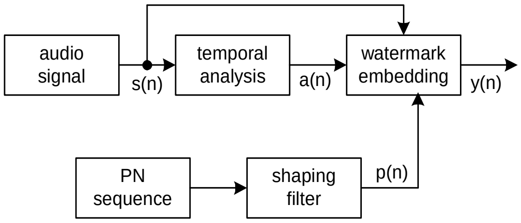

Figure 1: Watermark embedding scheme

Host audio sequence is also analysed in the time domain, where a minimum or a maximum is determined in the block of host audio signal with 335 samples (length of 7.6 ms). The masker value determined from the temporal analysis is used as a reference point to determine the level of power of the watermark sequence in the analysed block. With reference to temporal masking curves and the length of analysed audio frames, it was concluded that the added information should be at least 26 dB below the power level of the audio maximum in the frame [1]. As the result temporal analysis, the watermark samples are weighted by the coefficients a(n) in order to be below masking threshold.

Therefore, the watermark signal is embedded into host audio using two time-aligned processes. In the first process, the PN sequence has been filtered with the shaping filter, where sequence p(n), adapted to psychoacoustic properties of HAS in frequency domain, is the output. At the output of the watermark embedding scheme sequence p(n) is being weighted using coefficients a(n) and parameter α and added to the original audio signal:

$$y(n)=s(n)+\alpha\cdot a(n)\cdot p(n)=s(n)+w(n)$$

where s(n) denotes input audio signal, a(n) are coefficients from the temporal analysis block, α parameter represents the trade-off between perceptual transparency and detection reliability and w(n) stands for final watermark sequence. Parameter α can always be set to the value that places the masking curve of the algorithm near or under the most stringent local threshold value defined by standard masking model. Subjective listening tests were performed on different audio clips in order to experimentally determine the maximum value for α. High perceptual transparency was achieved for $\alpha\in(0.1,0.4)$, depending on the type of music.

3. WATERMARK DETECTION

In a correlation detection scheme, usually used for watermark extraction process in spread-spectrum watermarking algorithms, it is often assumed that the host audio signal is white Gaussian process. However, real audio signals don't have white noise properties as adjacent audio samples are highly correlated. Therefore, presumption for optimal signal detection in the sense of signal to noise ratio is not satisfied, especially if extraction calculations is performed in short time windows of audio signal. Figure 2a depicts probability density function (pdf) of 5000 successive samples of a short clip of the watermarked audio signal. It is obvious that the pdf of watermarked audio is not smooth and has a large variance.

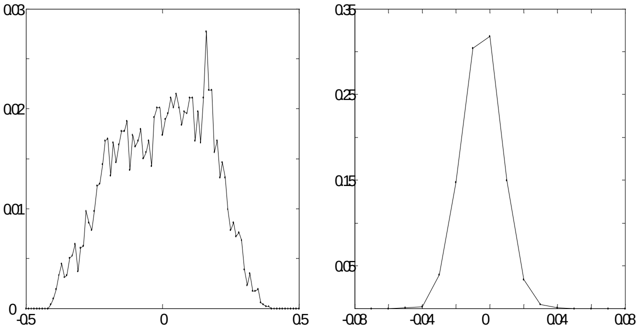

Figure 2. a) Pdf of watermarked audio; b) Pdf of residual signal after Savitzky-Golay filtering

3.1. Whitening of audio signal component

However, in the case of coloured Gaussian channel, it is possible to significantly increase the detection performance by pre-processing of the watermarked audio before correlation calculation. In order to decrease correlation between samples of the audio signal, we use least squares (Savitzky-Golay) smoothing filters (with different polynomial order and window length), which are typically used to "smooth out" a noise signal whose frequency span is large [6]. Rather than having their properties defined in the Fourier domain, and then translated to the time domain, Savitzky-Golay filters derive directly from a particular formulation of the data-smoothing problem in the time domain. Savitzky-Golay filters are optimal in the sense that they minimize the least square errors in fitting a polynomial to frames of noisy data.

Equivalently, the idea is to approximate the underlying function within a moving window by a polynomial, typically quadratic. Figure 2b shows the pdf of the 5000 consecutive samples of the residual signal after applying Savitzky-Golay filters, with fourth order polynomial and 21 samples long time windowing. It can clearly be seen that pdf of residual signal has more Gaussian-like distribution and significantly smaller variance compared to the case of pdf of watermarked audio signal. We verified Gaussian-like distribution of the residual signal using Bera-Jarque parametric hypothesis test of composite normality [7] and single sample Lilliefors hypothesis test [8]. Both tests have rejected hypothesis that watermarked audio has white Gaussian distribution, with significance level of 5%. On the other hand, both tests showed that we cannot reject the hypothesis that residual signal has white Gaussian distribution, using the same significance level.

3.2. Optimal watermark detection

Pre-processed audio test sequence $\hat{y}$ may have an embedded watermark ($\hat{y}(i)=\hat{s}(i)+\hat{w}(i)$, $0 \leq i \leq N-1$), on the other hand it may be an unwatermarked audio sequence ($\hat{y}(i)=\hat{s}(i)$, $0 \leq i \leq N-1$). The detection process verifies two hypotheses on the received content [2]:

- H_0: $\hat{y}$ represents a non-watermarked audio content, so it is Gaussian white noise – residual signal of host audio after decorrelation process
- H_1: $\hat{y}$ represents a watermarked audio sequence; it consists of decorrelated host audio and watermark.

As decorrelation pre-processing was implemented, we can assume that output of decorrelation filter $\hat{y}$ for a given $\hat{w}$ has the white Gaussian distribution and the Likelihood Ratio Test may be performed.

In addition, ”watermark part” of the residual signal ($\hat{w}$) is a sequence of samples $\hat{w}(i)$ with two equiprobable values, for example $\hat{w}(i) \in \{-\varepsilon, +\varepsilon\}$, generated independently with respect to $\hat{s}$. Parameter ε is set based on temporal analysis within one block of host audio. As same PN generation and perceptual shaping of the PN sequence can be done on the “receiver side”, correlation detector performs the simple correlation calculation between pre-processed audio and whitened watermark sequence:

$$C= \hat{y} \cdot \hat{w} = (\hat{s}+\hat{w}) \cdot \hat{w}= \hat{s} \cdot \hat{w} + \hat{w} \cdot \hat{w} = \hat{s} \cdot \hat{w} + N \cdot \varepsilon^2$$

where N is cardinality of involved vectors, and the correlation between two vectors $\hat{u}$ and $\hat{g}$ is defined as $\hat{u} \cdot \hat{g} = \Sigma\, \hat{u}(i) \cdot \hat{g}(i)$. Since the host audio signal part of the residual audio clip - $\hat{s}$ can be approximated as a Gaussian random vector

$$\hat{s} \sim N(\mu_x, \sigma_x);\ \sigma_x >> \varepsilon,$$

the normalized value of correlation can be written as:

$$Q = \frac{C}{N\varepsilon^2} = \rho + \frac{1}{\varepsilon} N\left(0, \sigma_X / \sqrt{N}\right)$$

where ρ=1 if watermark is present and ρ=0 if there is no watermark. The optimal detection rule is to declare that watermark is embedded in host audio if value of Q exceeds given threshold value T. The selection of the threshold T controls the trade-off between false alarm probability and probability of detection. Using derivations from the Central Limit Theorem, probability that Q>T is equal to:

$$\lim_{N \to \infty} \Pr\left[Q > T\right] = \frac{1}{2} erfc\left(\frac{T\sqrt{N}}{\sigma_x \sqrt{2}}\right)$$

It is obvious that decorrelation of audio sequence leads to decrease in variance value of signal σ_x (Figure 2), which again, according to the formulas given above should lead towards better detection performance and smaller false alarm probability [3]. Dominant factor of the detection algorithm is determined by the autocorrelation of the whitened watermark sequences, while “noise” associated with audio covert communications channel is additive white Gaussian.

4. EXPERIMENTAL RESULTS

After inserting a watermark as described in Section 2, we compared decorrelation filter detection from Section 3 with plain matched filter detection process. The embedded watermark should be extractable even if common signal processing or compression attacks are applied to host audio. An attacker may attempt blind watermark destruction, for example requantization, low pass filtering, dynamic compression, equalization and mp3 compression etc. As the test audio sequences we used audio clips from broad range of different music styles, results presented below are for Celine Dion’s “My heart will go on” audio segment, as mono signal, 11.61 seconds long, sampled at 44.1 KHz with 16 bits per sample resolution. Averaged SNR of inserted watermark is –26 dB, as meaningful range of SNR in audio watermarking systems should be below –26 dB [1]. Processing was performed in MathWorks’ MatLab and Syntrillium’s CoolEdit 2000. Detection performance of the algorithm was tested against common watermarking attacks:

1. MPEG layer-3 coding, at a rate of 64 kb/s using Syntrillium’s commercial mp3 coder licensed from the Fraunhofer IIS.
2. Low-pass filtering using Blackman-Harris window with FFT length of 2048 samples with cut-off frequency 3000 Hz and 40 dB stop band attenuation.
3. Amplitude compression (8.91:1 for A>-29dB, 1.73:1 for –46dB<A<-29dB and 1:1.61 for A<-46dB)

Watermark detection results are obtained for 512*100 detection interval [2], where hundred times redundant watermark detection is performed (the length of audio clip used for one observation is 1.161 sec). The test is performed on ten consecutive time frames of watermarked audio. During the test value ρ was evaluated, therefore the ideal test value should be ρ=1, because we tested watermarked audio clip. However, watermark attacks significantly change spectral and temporal characteristics of watermarked audio in order to remove embedded watermark. Figure 3 shows detection results after mp3 compression, with bitrate 64 kbps/mono.

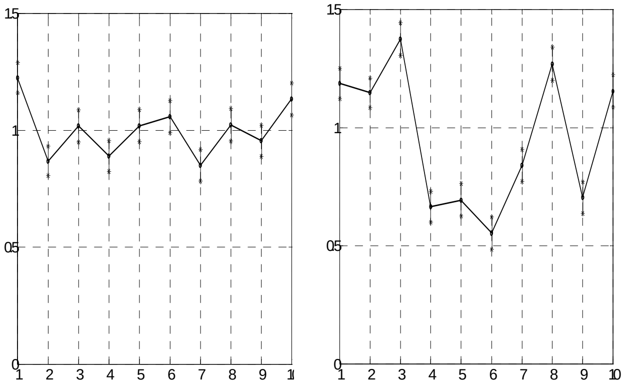

Figure 3. a) Detection after decorrelation (mp3, 64 kbps); b) Detection (no decorrelation, mp3, 63 kbps)

‘O’ denotes the averaged correlation coefficient within one time frame and ‘*’ indicates the estimated standard deviation value of ‘O’ within given time frame. It is obvious that decorrelation algorithm considerably improves watermark detection; ρ values are closer to ideal value ρ=1, and variations across ten time frames are remarkably smaller.

In the next experiment, we tested detection performance of the algorithm when low pass filtering attack on watermarked audio is used. Again, significantly better detection performance is attained when decorrelation algorithm is done prior to watermark detection. The reason for better performance is that watermarked audio sequence after mp3 compression and low pass filtering attacks still keep their amplitude-pdf different from white Gaussian pdf. Therefore, correlation detection is not optimal in the sense of Signal to Noise Ratio (SNR) as channel is far from additive white Gaussian channel. Residual signal has in both cases properties considerably more similar to AWGN and detection is accordingly more precise and stable.

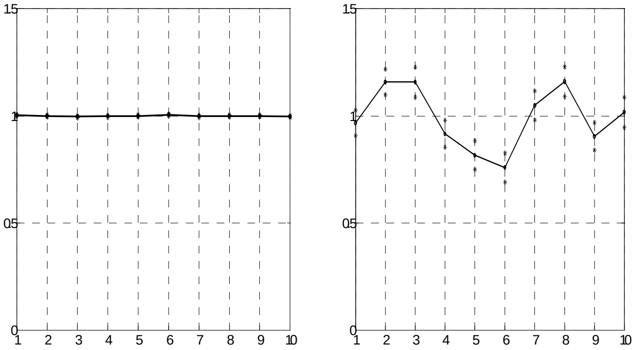

Figure 4. a) Detection after decorrelation (low pass filter); b) Detection (no decorrelation, low pass filter)

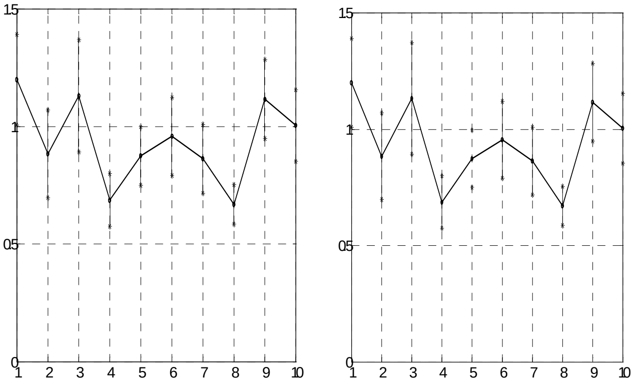

Figure 5. a) Detection after decorrelation (amplitude compression); b) Detection (no decorrelation, amplitude compression)

Attack on our watermarking algorithm with severe amplitude compression destroys dynamic properties of the watermarked audio by attenuating all large amplitudes of host audio sequence, causing far more perceptual distortion in comparison with the other watermark attacks introduced. Test results, given in Figure 5, confirm that no significant improvement in detection results is possible using decorrelation filter, as watermarked audio already has Gaussian-like pdf of amplitudes after amplitude compression attack. In general, the decorrelation algorithm described in Section 3 improved performance and stability of watermark detection correlation test. Similar test results are obtained also in the case of echo addition, equalization, noise addition, etc.

5. CONCLUSION

In order to achieve optimal watermark detection based on correlation, statistical properties of the host audio signal must be similar to the statistical properties of AWGN. In this paper, we describe whitening procedure for audio using Savitzky-Golay FIR filter, based on polynomial fitting of data. Using this procedure for pre-processing of watermarked audio significantly better detection results have been achieved without imposing large additional computational complexity in the detection algorithm. Experimental results show higher resistance to common watermark attacks if the whitening procedure is used, in comparison with watermark detection of unprocessed audio signal.

AKNOWLEDGMENTS

The work is part of the Duchess project, which is supported by the Finnish National Technology Agency, TEKES. The research topic has been supported by Nokia Oyj Foundation.

REFERENCES

[1] P. Bassia, I. Pitas. "Robust audio watermarking in the time domain", *IEEE Transactions on Multimedia, Vol.3, No 2.,* pp. 232-241, June 2001.

[2] Mitchell D. Swanson, Bin Zhu, Ahmed H. Tewfik. "Robust audio watermarking using perceptual masking", *Signal Processing, Vol. 66*, pp. 337-355, 1998.

[3] D. Kirovski and H. Malvar “Robust covert communication over a public audio channel using spread spectrum”, *4th International Information Hiding Workshop,* Pittsburgh, PA, April 2001.

[4] ISO/IEC IS 11172, Information technology – coding of moving pictures and associated audio for digital storage up to about 1.5 Mbits/s

[5] H. Kim ”Stochastic model based audio watermark and whitening filter for improved detection”, *IEEE International Conference on Acoustic, Speech and Signal Processing*, Istanbul, Turkey, pp. 1971-1974, June 2000.

[6] Orfanidis, S.J., Introduction to Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1996.

[7] J. B. Cromwell, W. C. Labys and M. Terraza: *Univariate Tests for Time Series Models*, Sage, Thousand Oaks, CA, 1994.

[8] Conover, W. J. (1980). Practical Nonparametric Statistics. New York, Wiley.

[9] Cvejic N. and Seppänen T. “Improving audio watermarking scheme using psychoacoustic watermark filtering, *IEEE International Symposium on Signal Processing and Information Technology,* Cairo, Egypt, 2001, accepted