ROBUST AUDIO WATERMARKING USING FREQUENCY SELECTIVE SPREAD SPECTRUM THEORY

Hafiz Malik, Ashfaq Khokhar, Rashid Ansari
Dept. of Electrical and Computer Engineering University of Illinois at Chicago, Illinois, USA

ABSTRACT

A new method is proposed for robust audio watermarking using direct-sequence spread spectrum in combination with the subband decomposition of the audio signal. The method exploits the frequency masking characteristics of the human auditory system (HAS) and inserts the watermark into a randomly selected frequency band of the input audio signal. Performance of the proposed system is evaluated for robustness to signal manipulations such as: contamination with additive noise, resampling, compression, filtering, multiple watermark insertion, and random chopping. Experimental results show that the capacity of the proposed watermarking scheme is relatively high compared with existing spread spectrum based audio watermarking schemes.

1. INTRODUCTION

The growth of the Internet, proliferation of the low-cost and reliable storage devices, deployment of long-distance Gbps networks, and the availability of powerful software packages for digital media editing (Adobe PhotoShop, X-wave etc.) has made digital forgeries and unauthorized sharing of digital media a reality. This leads to enormous annual revenue loss due to piracy. It is therefore imperative to have robust technologies to protect copyrighted digital media from illegal sharing and tampering. Traditional digital data protection techniques, such as encryption and scrambling, alone cannot provide adequate protection of copyrighted contents. Digital watermarking technology complements cryptography for protecting content even after it is deciphered [1]. The performance of a given watermarking system depends on the application of interest for which the watermarking system is designed [1].

In this paper an audio watermarking method based on direct-sequence spread-spectrum (DSSS) is proposed. The method is designed to overcome common shortcomings of existing DSSS based audio watermarking systems [5-10] such as vulnerability to desynchronization attacks, poor detection performance, poor fidelity (inaudibility), and limited watermarking capacity. Robustness to desynchronization attacks and reliability of detection performance are improved using content-adaptive features called salient points [11] of the input audio. These salient points are frame level features of the input audio signal that are invariant to common audio processing operations. Only a small fraction of the audible frequency range is used for embedding the watermark in order to reduce the amount of audible distortion. The method exploits the frequency masking characteristics of the human auditory system (HAS) and inserts the watermark into a randomly selected frequency band of the input audio signal. A secret key is used for randomly selecting a frequency band for watermark embedding. The proposed watermarking scheme induces low perceptual as well as mean squared distortion; and is therefore, the proposed scheme has high embedding capacity [4]. Detection performance of the system was investigated for a variety of signal manipulations and attacks on a watermarked audio clip. These attacks include addition of noise, resampling, requantization, filtering, and random chopping. Results show the robustness of the method, with a low detection error rate and a low bit error rate. Moreover, the proposed watermarking scheme is capable of embedding multiple watermarks in the unused frequency bands with the use of separate secret keys.

2. WATERMARKING USING PERCEPTUAL AUDITORY MODEL

The basic idea underlying perception-based watermarking schemes is to incorporate the watermark into the perceptually insignificant region of an audio signal in order to ensure transparency. The perceptually insignificant region is determined using the human perceptual auditory model. Extensive work was done over the years on understanding the properties of HAS and applying this knowledge to audio applications [2]. An important application of perceptual models is in the area of perception-based compression [3]. An important characteristic of HAS is *auditory masking* that has been that has been exploited in audio coding for lossy compression. We consider its use in watermarking.

Human ear performs frequency analysis that maps a frequency to a location along the basilar membrane. The HAS is generally modeled as a non-uniform bandpass filter bank with logarithmically widening bandwidth for higher frequencies [3]. The bandwidth of each bandpass filter is set according to the critical band, which is defined as “the bandwidth in which subjective response changes abruptly” [2]. The critical band rate (CBR) is a measure of location on the basilar membrane just as the frequency gives a measure of location in a spectrum. The unit of critical band rate is Bark. The mapping between CBR and frequency is defined as:

$$z = 13 \arctan(0.76 f) + 3.5 \arctan[(f / 7.5)^2] \qquad (1)$$

where z is CBR in Barks and *f* is frequency in kHz.

Masking is a fundamental property of HAS and is a basic element of perceptual audio coding systems. It is a phenomenon by which a stronger audible signal makes a weaker audible signal inaudible [2], and it occurs both in frequency as well as time domain [2].

3. SALIENT POINT EXTRACTION

Spread spectrum techniques have been applied in digital watermarking [5-10] due to their potential for high fidelity, high capacity, robustness, and security. In the proposed scheme, the process of generating a watermark and embedding it into an audio signal is treated in the framework of spread spectrum theory. The original audio signal is treated as noise whereas the message information used to generate a watermark sequence is considered as data. The spreading sequence, also called pseudo-random noise sequence or PN-sequence, is treated as key. This watermarking strategy can be treated in the framework of communication models discussed in [1].

A critical aspect of designing a spread spectrum system is ensuring fast and reliable synchronization at the detector. Synchronization impacts performance as it reduces the overall capacity of the watermarking system, and an active adversary can use explicit synchronization information for de-synchronization attacks. To overcome these problems, synchronization is tied to attack-sensitive locations or salient points for watermark embedding and detection. Salient points are extracted based on the audio features sensitive to the HAS [11], e.g. fast energy transition points, zero crossing rate and spectral flatness measure. If these features are altered then noticeable distortion is introduced. A good salient point extraction method is one that approximately extracts the same salient points before and after common signal manipulations or watermark embedding [11]. Fast energy transition audio feature is used in our method for salient point extraction.

For an audio signal *x(n)*: n = 0,1,2,…N-1, the short time energy ratio at each point is calculated as:

$$Er(n) = \frac{E_{after}(n)}{E_{before}(n)} \qquad (2)$$

where $E_{after}(n)$ and $E_{beforer}(n)$ are defined as:

$$E_{before}(n) = \sum_{i=-r}^{-1} x^2(n+i) \qquad (3)$$

$$E_{after}(n) = \sum_{i=0}^{r-1} x^2(n+i) \qquad (4)$$

Here r is the number of samples before and after *x(n)*. A high energy transition points are defined as:

if: $Er(n) > Th_1$ & $E_{after}(n) > Th_2$

Finally a salient point is decided as follow:

1: If two high energy transition points are separated by less than Th_3 then samples are merged together to form a group.

2: Within each group, the strongest transition point is marked as a salient point.

here Th_1, Th_2 *and* Th_3 are thresholds, these thresholds are set adaptively to ensure 3 - 4 salient points per second.

4. WATERMARK EMBEDDING

To generate and embed a watermark, a list of salient points is first created. A block of *P* samples around each salient point is selected. The block is applied to a 5-level modified wavelet analysis filter bank to generate nine subband signals of unequal bandwidths, as illustrated in Figure1.

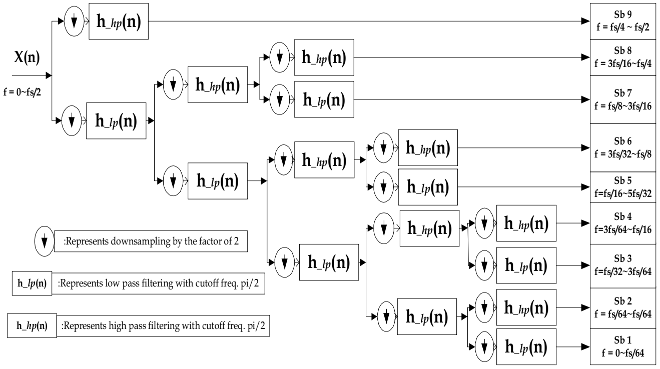

Figure 1: Analysis Filter Bank

The choice of the number of subbands is made based on a compromise between allowing a large choice in random selection and ensuring that the subband bandwidth covers at least three critical bands. A subband from lower eight bands (for n = 5) is selected using the three bit key k_{1i}, for i^{th} salient point, where as the complete key ***K1*** for subband selection is given as:

K1 = $k_{11}| k_{12}|\ldots| k_{1i}|\ldots k_{1M}$

where *M* is the cardinality of the salient point set.

The selected subband is used to estimate the masking threshold *Tm(k)*, which is calculated as follows:

Let $sb_{i,j}(n)$ *for* $n = 0,1,2\ldots L-1$ be the j^{th} subband of i^{th} frame of the audio data that is selected using key K_{1i}. Its power spectrum is defined as,

$$Psb(k) = |Sb_{i,j}(k)|^2 \qquad (5)$$

here $Sb_{i,j}(k)$ is the discrete fourier transform (DFT) of the $sb_{i,j}(n)$. Now k is wrapped onto Bark scale using Eq 1. The energy in each critical band is calculated as,

$$E(z) = \sum_{k=LBZ}^{UB} Psb(k) / P_z \; : \; for \;\; z = 1,2,\ldots z_t \qquad (6)$$

where z_t is the total number of critical bands in the selected subband, LB and UB are the lower and upper boundaries of the a critical band, and P_z is the total number of points in each critical band. The energy per critical band is used to calculate the masking threshold *Tm(z)* using MPEG layer III psychoacoustic model 1[3].

4.1 Watermark Generation

For each salient point a watermark ***W*** of length L is generated. To generate a watermark ***W***, binary message ***m*** is mapped onto $\boldsymbol{m}^{/}$ using a channel encoder. The channel encoded data is applied to binary phase shift keying

(BPSK) modulator. The output of the BPSK modulator is *Wm(n)* **:** *n = 0,1…q-1*, where *q = L/(spreading factor)*. Maximum length PN-sequence ***p*** of length (L/q) using $log_2 (L/q)$ bit secret key ***K2*** is generated. Finally modulated signal ***Wm*** is spread using PN-sequence ***p*** to generate final watermark ***W***. System key ***K = K1|K2***.

4.1 Watermark Embedding

Spectral shaping based on *Tm(k)* of ***W*** is required to ensure inaudibility of the embedded watermark. For this purpose *W(k) (DFT)* and power spectrum *Pw(k)* of ***W*** is calculated. Now using *Tm(z)* inaudible DFT coefficients of the selected subband $sb_{i,j}$ are removed, i.e.

$$Sbn(k) = \begin{cases} sb_{i,j}(k) & if \quad Psb(k) \geq Tm(z) \\ 0 & if \quad Psb(k) < Tm(z) \end{cases} \qquad (7)$$

similarly unwanted DFT coefficients of *W(k)* are also removed, i.e.

$$Wn(k) = \begin{cases} 0 & if \quad Psb(k) \geq Tm(z) \\ w(k) & if \quad Psb(k) < Tm(z) \end{cases} \qquad (8)$$

The final watermark before embedding is given by

$$W_f(k) = F_z \bullet Wn(k) \qquad (9)$$

where F_z is the shaping factor and defined as,

$$F_z = \frac{A\sqrt{Tm(k)}}{\max(|Wn(k)|)} \qquad (10)$$

where 0 < A <1 is noise gain factor. Finally watermarked output in frequency domain,

$$Wsb_{i,j}(k) = Sbn(k) + W_f(k) \qquad (11)$$

The corresponding time domain watermarked subband signal is obtained by calculating inverse discrete fourier transform (IDFT),

$$Wsb_{i,j}(n) = IDFT\{Wsb_{i,j}(k)\} \qquad (12)$$

This watermarked subband is then use to reconstruct the watermarked audio block data using modified wavelet synthesis filter bank. This process is repeated for the remaining salient points in the salient point list.

Watermark generation and embedding process is illustrated in Figure2.

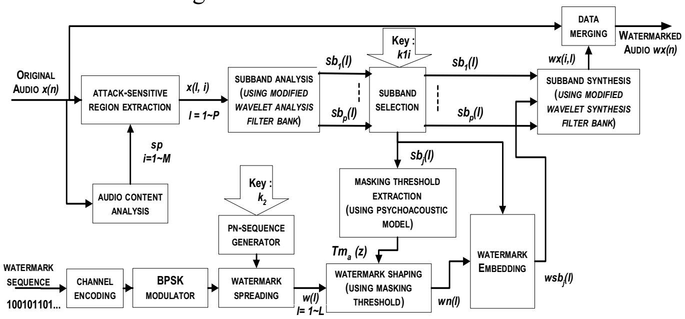

Figure 2: Block Diagram of Watermark Embedding Process

5. WATERMARK DETECTION

In order to be effective a watermarking system should be able to detect/extract the embedded watermark even after the watermarked audio undergoes common signal manipulations and psychoacoustic auditory model based audio processing. An attractive feature of the proposed scheme is that a blind detector can be used for watermark detection/extraction i.e. detector does not require original copy of the audio signal to detect watermark from the received audio signal. The detector has access to the secret key that is the only information that detector has about the embedding. The detector uses salient points for synchronizing the embedded information, so that audio can be analyzed for salient point extraction (as discussed in Section 3). For each point in the salient point list a block of *P* samples is passed through the modified analysis wavelet-filter bank, then using a secret key $\boldsymbol{k_{1i}}$, j^{th} subband $sb_{i,j}$ is selected for watermark detection/extraction. The selected subband is analyzed to extract masking threshold say $Tm_r(z)$. This masking threshold is used to extract the "residual" audio signal, $R_r(k)$, that is defined as,

$$R_r(k) = \begin{cases} 0 & if \quad Ps_r(k) > Tm_r(z) \\ s_r(k) & if \quad Ps_r(k) \leq Tm_r(z) \end{cases} \qquad (13)$$

where $S_r(k)$ is the DFT of $s_r(n)$ the selected subband of the received audio and $Ps_r(k)$ is the corresponding power spectrum.

The residual is transformed into time domain for watermark detection/extraction using IDFT i.e.

$$r_r(n) = IDFT(R_r(k)) \qquad (14)$$

The residual $r_r(n)$ is now used for watermark detection, by using normalized correlation test. The normalized correlation between real sequences $r_r(n)$ and PN-sequence *p(n)* at the detector generated using key ***K2*** is defined as,

$$cor_n(n) = \frac{\sum_{l=-L}^{L} r_r(l)\, p(n+l)}{\sqrt{\sum_{l=0}^{L} r_r(l)^2 \bullet \sum_{l=0}^{L} p(l)^2}} \qquad (15)$$

w

where L is the length of the residual signal. High correlation implies the presence of watermark as illustrated in Figure 3.

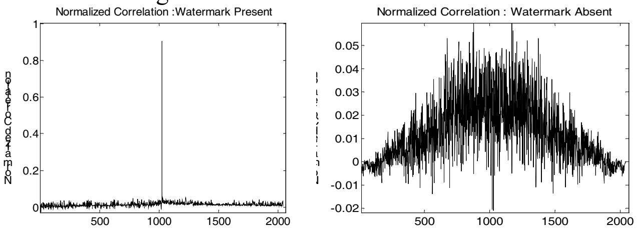

Figure 3: Normalized Correlation for watermarked subband (left) and unwatermarked subband (right).

The normalized correlation is compared with a threshold to determine the presence of a watermark. Let hypothesis H_1 denote the presence of a watermark in a selected subband and H_0 denote the absence of a watermark. The decision criterion is

$$\begin{aligned} H_1 &: \quad if \ \max(cor_n) \geq Th_0 \qquad (16) \\ H_o &: \quad if \ \max(cor_n) < Th_0 \end{aligned}$$

If H_1 is true then the embedded information is recovered by despreading $r_r(n)$ using the PN-sequence generated using same key ***K2***, then demodulating the resulting sequence using BPSK demodulator followed by channel decoding. The detection process is illustrated in Figure4.

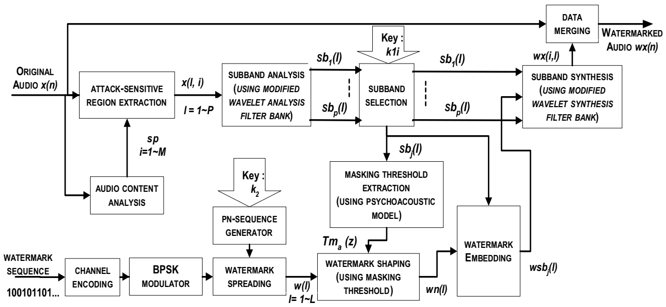

Figure 4: Block Diagram for Watermark Detection

6. EXPERIMENTAL RESULTS

The robustness of the proposed scheme was tested on speech signals and music. The tests included several degradations and distortions, i.e. addition of noise, lossy compression, low pass filtering, resampling, random chopping, and multiple watermarks. The detection performance in each case depends on the following measures, 1) watermark detection rate (WDR) which is a measure of watermark detection, and 2) the bit accuracy rate (BAR) which is a measure of data recovery. The bit accuracy rate is defined as,

$$BAR = \frac{Number \quad of \quad Bits \quad Correctly \quad Detected}{Number \quad of \quad Bits \quad Embedded} \qquad (17)$$

and watermark detection rate:

$$WDR = \frac{Number\ of\ Watermarked\ Frames\ Correctly\ Detected}{Number\ of\ Watermarked\ Frames\ Embedded} \qquad (18)$$

The overall performance of the system is defined as,

$$DPM = BAR \times WDR \qquad (19)$$

where *DPM* stands for detection performance measure. Detection results for degraded watermarked audio based on DPM for a variety of conditions are described below.

- White Gaussian noise is added to the watermarked audio; the *DPM* (as defined in Eq. 19) values in the presence of white gaussian noise with power from 0 to 50% of the signal power are shown in Figure 5.

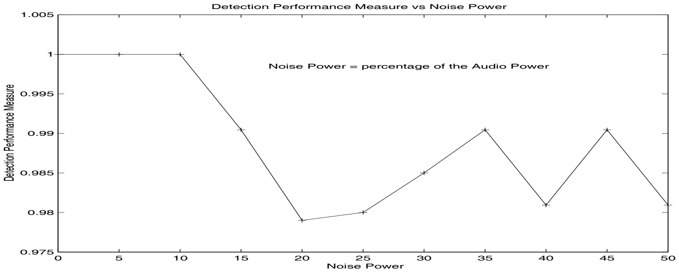

Figure 5: DPM for different values of noise power (P_n)

- Watermarked audio is down-sampled to 22.05 kHz and then interpolated to 44.1 kHz. The *DPM* value for this test is 1.
- Watermarked audio undergoes ISO/MPEG-1 Audio Layer III encoding/decoding at a bit rate of 128 kbs. The *DPM* value for compression test is 1.
- Watermarked audio signal is lowpass filtered with 4 kHz cutoff frequency, Detection of resulting audio gives a *DPM* of .995. Detection performance is still acceptable despite severe audible distortion
- To investigate desynchronization attacks, one out of every 100 samples of watermarked signal was randomly dropped. Detection applied to this signal gave a *DPM* of 1.
- Three watermarks simultaneously embedded in the audio, with a unique sectary key assigned to and a unique subband selected for each watermark. The DPM was 1 as long as the number of watermarks is less than the number of analysis subbands.

7. CONCLUSION

A novel watermarking scheme for audio based on FS-DSSS is proposed. The technique introduces lower mean square as well as perceptual distortion compare to existing schemes [5-10] this is due to that fact that a watermark is embedded in a small frequency band of complete audible frequency range. The watermarking capacity theory presented in [4] suggests that the proposed scheme can embed more information. The proposed method is also robust to standard data manipulations i.e. noise addition, compression, random chopping and resampling.

8. REFERENCES

[1] I. J. Cox, M. L. Miller, and J. A. Bloom, "Digital Watermarking," *Morgan Kaufmann,* 2002.

[2] E. Zwicker, and H. Fastl, "Psychoacoustics: Facts and Models," *Springer-Verlag, Berlin,* 1999.

[3] P. Noll, "MPEG Digital Audio Coding," *IEEE Sig. Proc. Mag.* vol. 14, no. 5, pp. 59-81, September 1997.

[4] P. Moulin, and J. A. O'Sullivan, "Information-Theoretic Analysis of Information Hiding," *IEEE Trans. Info. Theory*, vol. 49, No. 3, pp. 563-593, March 2003.

[5] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, Vol.35, nr. ¾, 1996

[6] I. J. Cox, J. Kilian, and T. Leighton, "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Trans. Image Proc.*, vol. 6, pp. 1673-1687, 1997.

[7] M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, "Robust Audio Watermarking Using Perceptual Masking," *Signal Processing*, vol. 66, pp. 337-355, 1998.

[8] C. I. Podilchuk, and E. J. Delp, "Digital Watermarking Algorithms and Applications," *IEEE Signal Processing Magazine*, pp. 33-45 July, 2001.

[9] D. Kirovski, and H. S. Malvar, "Spread Spectrum watermarking of Audio Signals," *IEEE Trans. Signal Proc.* Vol. 51, no. 4, pp. 1020-1033, April, 2003.

[10] R. A. Garcia, "Digital Watermarking of Audio Signals using Psychoacoustic Auditory Model and Spread Spectrum Theory," *107th Convention, AES*, New York, September, 1999.

[11] C.-P. Wu, P.-C. Su, and C.-C. J. Kuo, "Robust Audio Watermarking for Copyright Protection," *SPIE's 44th Anl. Met. Adv. Sig. Proc. Alg. Arch. Imp.,* July, 1999.