International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:3, No:11, 2009

Audio Watermarking Using Spectral Modifications

Jyotsna Singh, *Member, IEEE*, Parul Garg, *Member, IEEE* and Alok Nath De, *Senior Member, IEEE*

Abstract—In this paper, we present a non-blind technique of adding the watermark to the Fourier spectral components of audio signal in a way such that the modified amplitude does not exceed the maximum amplitude spread (MAS). This MAS is due to individual Discrete fourier transform (DFT) coefficients in that particular frame, which is derived from the Energy Spreading function given by Schroeder. Using this technique one can store double the information within a given frame length i.e. overriding the watermark on the host of equal length with least perceptual distortion. The watermark is uniformly floating on the DFT components of original signal. This helps in detecting any intentional manipulations done on the watermarked audio. Also, the scheme is found robust to various signal processing attacks like presence of multiple watermarks, Additive white gaussian noise (AWGN) and mp3 compression.

Keywords—Discrete Fourier Transform, Spreading Function, Watermark, Pseudo Noise Sequence, Spectral Masking Effect

I. Introduction

It is possible to increase the energy present in particular frequencies by exploiting the masking effect in the human auditory systems. Various such techniques have been proposed in the area of multimedia for enhancing the security of digital data [1]. The basic idea is that we can increase the amplitude of the Discrete Fourier Transform (DFT) components of the signal by adding the watermark without giving any perceptual effect if, the increased amplitude is below the Amplitude Spreading Effect of its nearby frequency components. The proposed method process the audio signal frame by frame to spread the amplitude of its DFT coefficients. For this, it incorporates the spectral energy masking function of [2]. The amplitude spreading of $N/2$ DFT components is evaluated and its effect is seen at all the N frequency locations of a frame. This gives us a $N/2 \times N$ matrix. The DFT coefficients of Host Signal are then modified by adding watermark, such that the modified spectra is always below the maximum amplitude spread of original signal. The symmetry of DFT is maintained by adding the N point DFT of watermark to N point DFT of host signal. As a result, the dc and the nyquist components remain real and the complex conjugate symmetry of other terms is also not disturbed. This keeps the sample values of audio real, even after the watermarking process.
Using this watermarking technique successful detection of watermark signal had been accomplished. Also, the robustness of our watermarking approach to different types of attacks like presence of multiple watermark, additive white gaussian noise and MP3 compression has been discussed. We have evaluated the percentage recovery of watermark data in the presence of Additive White Gaussian noise (AWGN), through QPSK modulated Watermark channel [3].

Jyotsna Singh and Parul Garg are with the Division of Electronics and Comm. Engg., Netaji Subhas Institute of Technology, Sector 3, Dwarka, New Delhi 110075, India, email: (jyotsna_nsit@yahoo.co.in).
Alok Nath De is with ST Microelectronics, Plot 1, K. Park-III, Greater Noida, UP-201308, India, email: (aloknath.de@st.com).

II. Bark Scale

The basilar membrane in the hearing mechanism analyzes the incoming sound through the spatial-spectral analysis. This is done in small sectors or regions of the basilar membrane that are called *critical bands*. If all the critical bands are added together in a way that the upper limit of one is the lower limit of next one, the critical band scale is obtained. Also, a new unit has been introduced, the *bark* that is by definition one critical band wide [4]. The formula to convert frequency f in hertz to frequency z in bark scale is

$$z = 13\tan^{-1}\left(\frac{0.76f}{1000}\right) + 3.5\tan^{-1}\left(\left(\frac{f}{7500}\right)^2\right) \quad (1)$$

III. The Basilar Membrane Spreading Function

Let $\overline{s} = (s_0, \cdots, s_{N-1})$ be a discrete signal, and S_k its DFT

$$S_k = \sum_{n=0}^{N-1} s_n e^{-i(2\pi nk/N)} \quad (2)$$

Here $k = 0, 1, \cdots, N-1$. The inverse of the DFT gives back the samples in time domain

$$s_n = \frac{1}{N}\sum_{k=0}^{N-1} S_k e^{i(2\pi nk/N)} \quad (3)$$

The real nonnegative energy spreading function $SF_{dB}(i,j)$ approximates the basilar spreading as a triangular spreading function [4], given as

$$\begin{aligned} SF_{dB}(i,j) &= 15.81 + 7.5(\Delta_z + 0.474) \\ &\quad -17.5\sqrt{1 + (\Delta_z + 0.474)^2} \end{aligned} \quad (4)$$

$SF_{dB}(i,j)$ is the masking spread in decibels (dB) from i^{th} frequency to j^{th} frequency. The bark separation between i^{th} and j^{th} DFT components is $\Delta_z = z_j - z_i$, where z_i denote the bark frequency of i^{th} frequency location.

IV. Watermarking Scheme

A. Amplitude Spreading Function

The DFT of real vector $\overline{S}$ satisfies the symmetry property $S_k = S^*_{N-k}$, where $k = 1, ..., N-1$. Let the discrete signal $\overline{s}$ is sampled at frequency, f_s Hertz. The DFT (S_k, for $0 \leq k \leq N/2$, N a power of 2) corresponds to frequency in Hz

$$f_k = F_s \times k/N\,, \qquad k \leq N/2 \quad (5)$$

Considering the duplication in the spectra for $k \geq N/2$, we evaluate the masking spread for amplitude of $N/2$ components only

$$A1(i,j) = \sqrt{SF(i,j)}\,, \qquad 0 \leq i \leq N/2 - 1 \tag{6}$$

The square root is to convert the masking spread from energy scale to amplitude scale. $SF(i,j)$ is obtained by taking the inverse decibel of $SF_{dB}(i,j)$. Now respecting the symmetry property of DFT components, we define $A(i,j)$ as,

$$\begin{aligned} A(i,j) &= A1(i,j), && 0 \leq j \leq N/2 \\ &= A(i,N-j), && N/2+1 \leq j \leq N-1 \end{aligned} \tag{7}$$

The amplitude spread of i^{th} DFT component is then defined as,

$$A'(i,j) = A(i,j)S(i) \tag{8}$$

This gives $N/2 \times N$ matrix showing amplitude spread of each of the $N/2$ DFT components at $N = 512$ frequency locations. This is the convolution of basilar membrane spreading function with the amplitude of audio signal. Figure 1 shows the amplitude spread of 17^{th} and 20^{th} DFT components and magnitude of 19^{th} DFT component. Figure implies that the magnitude of 19^{th} DFT component can be modified till the point of intersection of 17^{th} and 20^{th} amplitude spreads.

B. Maximum Amplitude Spread

Using(8), evaluate the Maximum Amplitude Spread $Y(j)$ at frequency j, due to $i = 0, 1, ...N/2-1$ frequency components.

$$Y(j) = \max(|A'(i,j)|) \tag{9}$$

for $0 \leq j \leq N-1$. The amplitude spreads of neighboring DFT components overlap each other. We consider the amplitude spread at j, whose absolute value is maximum of all the overlapping spreads due to $N/2$ DFT components. Figure 2 shows the plot between the absolute value of maximum amplitude spread and their DFT coefficients at all the locations $j = 0, 1, ..., N-1$.

C. Threshold Computation

The Threshold value of scaling coefficient α decides, that by what amount the watermark is to be suppressed so that its maximum data can be embedded into the spectra of original signal with least perceptual distortion. In our proposed method α is evaluated from the difference of maximum amplitude spread and the spectral amplitude of original signal. This is the maximum coefficient value which inserts watermark sample on every spectral component. An uninterrupted thin layer of watermark floats on spectra i.e. 512 samples of watermark float on same number of spectral components of host signal. If the coefficient value exceeds this threshold, then the watermark layer gets disturbed. Going below this value will not disrupt the watermark layer, but its magnitude then becomes so low to detect.

D. AWGN Watermark Channel

Digital watermarking of multimedia can also be viewed as a communication problem. The message information to be embedded is converted into a watermark signal, which is then sent through a channel to the receiver. The receiver must locate the watermark signal and attempt to recover the message through it. We refer this channel as watermarking channel. In this paper we assume the communication channel with AWGN, as watermark channel with additive noise as attack. This bitstream is transmitted over the noisy channel using coherent $\pi/4$-shifted quadriphase-shift key (QPSK) modulation [5]. The communication channel is modeled [6] with one transmitting antenna and one receiving antenna. Let the matrix for channel is given by

$$H = [\,h1\,] \tag{10}$$

for additive white Gaussian Noise $(AWGN)$ channel the path gain $h1 = 1$. Let W_t is the modulated watermarked symbols transmitted at time t through the antenna. The received signal R_t at time t is given by

$$R_t = H[W_t] + N_t \tag{11}$$

where N_t is the channel noise at time t. The noise is modeled as an independent samples of zero-mean complex Gaussian noise with variance $(1/(2\gamma))$ for both real and imaginary parts. We assume that the total energy of the signals at the transmitter is unity so that γ is the signal to noise ratio (SNR). The received signal power E_b/N_0, is computed from vector summation of arriving arrays. For $QPSK$ modulation scheme, bit error rate (BER) is derived from the values of E_b/N_0 using Monte Carlo Simulation. The relation ship between the received signal power E_b/N_0 and BER at time t is given by.

$$BER_{AWGN} = \frac{1}{N}\sum_{m=1}^{N} BER(mT_s) \tag{12}$$

Figure 3 shows the bit error rate for coherent $QPSK$ modulated signal which was transmitted over $AWGN$ channel.

V. WATERMARK EMBEDDING

The process of inserting a digital watermark into an audio file can be divided into four main processes. The original audio file in wave format is fed into the system, where it is subsequently framed, analyzed, and processed, to attach the watermark to the output signal.

STEP 1: The audio file is sampled at the rate of 44.1 kHz and portioned into frames of $N = 512$ samples. These frames are further weighted with a Hann window h(n),

$$h(n) = \frac{\sqrt{8/3}}{2}\left[1 - cos\left(2\pi\frac{n}{N}\right)\right], \tag{13}$$

where $n = 1, 2, ..., N-1$.

STEP 2: Subsequent to the framing of the unprocessed audio signal, we perform spectral analysis on the signal, consisting of a discrete Fourier transform (DFT), is given as

$$F_k = \sum_{n=0}^{N-1} s_n e^{i(2\pi nk/N)} \quad (14)$$

for $k = 0, 1, ..., N-1$. With a standard 16 bit CD quality audio file having a sampling rate, $Fs = 44,100$ samples per second.

STEP 3: Computation of threshold value of α using (2) and (9).

STEP 4: Let O denote the output of the watermarking system, which is the watermarked audio signal, H be the host signal and W be the watermark to be embedded, all of which are processed in the frequency domain [7]and are related by:

$$O' = H + \alpha W \quad (15)$$

where α is an appropriate scaling factor, which decides how much the amplitude of watermark is to be suppressed before adding it to the spectrum of host signal.

$$\begin{aligned} O(j) &= H(j), & if\,|O'(j)| > |Y(j)| \\ &= H(j) + \alpha W(j), & if|O'(j)| \leq |Y(j)| \end{aligned} \quad (16)$$

where $j = 0, 1, ..., N-1$. The noise is modeled as an independent samples of zero-mean complex Gaussian noise with variance $(1/(2\gamma))$ for both real and imaginary parts.

VI. WATERMARK DETECTION

In detection process the well known property of PN-Sequence is exploited, which is used as a watermark. PN-sequences are periodic noise like sequences which provide the author a unique code for identification. The threshold coefficient α is selected in a manner that, on every sample of original signal, there is a watermark sample floating on it. As a result, we are able to upload a watermark which is of same data length as of the host signal. Even with such good watermark holding capacity we find that there is low perceptual distortion in the watermarked audio. Figure 4 shows the original audio, watermarked audio and their superimposed picture.
The performance of watermarking scheme under various signal processing manipulations is discussed below.

A. Presence of Multiple Watermark

We generated about 400 normally distributed pseudo-random watermarks with mean 0 and variance 1. watermark detector [8], shows very low rate of false alarm as shown in Figure 5.

B. Presence of Additive White Gaussian Noise

The performance of watermark channel is evaluated in the presence of AWGN. A plot between BER and percentage watermark recovery is shown in Figure 6. As can be seen from the Table I more then 99 percent of watermark recovery is achieved for SNR value of $6dB$ and above. Also, in the given table the bit error rate corresponding to given SNR is shown.

TABLE I
PERFORMANCE OF WATERMARK DATA IN THE PRESENCE OF AWGN

E_b/N_o	BER	% Watermark Recovery
1	7.4120 $\times 10^{-2}$	19.366
2	2.4559 $\times 10^{-2}$	57.813
3	5.0851 $\times 10^{-3}$	91.797
4	3.9063 $\times 10^{-4}$	98.633
5	6.9754 $\times 10^{-6}$	99.219
≥ 6	6.9754 $\times 10^{-6}$	99.414

C. Robustness to mp3 Compression

A watermarked audio file was converted from wav to mp3 format. The sampling rate of uncompressed data was 44.1 kHz, 16 bit per sample. In compression scheme the sampling rate of 44,100 Hz was first converted into 12,000 Hz and then the size was further reduced to 20 kbps. The watermark was successfully retrieved after this compression/decompression process.

VII. CONCLUSIONS

The proposed method introduces the watermark in the spectral domain by exploiting the amplitude spreading effect of DFT components of the audio signal. The computations done for evaluating the frequency and temporal masking effects [9] is not required in this method. Our watermark is imperceptibly embedded into the audio signal and is easy to detect by the author. Due to the correlation properties of PN-sequence, the algorithm has a very low rate of false alarm. The results show that our watermarking scheme is robust to various attacks like presence of multiple watermarks, addition of white gaussian noise and MP3 compression.

REFERENCES

[1] De, A. Multimedia Watermarking in Enhancing Digital Security. *ITU-T/ITU-D Workshop*, Aug.2001,Banglore.
[2] Schroeder, M. R., B. S. Atal, and J. L. Hall. Optimizing digital speech coders by exploiting properties of the human ear. *Journal Acoust. Soc. America*, vol. 66, no. 6, pp. 1647–1652, December 1979.
[3] Su, J. K., F. Hartung, and B. Girod. A channel model for a watermark attack. *In security and watermarking of multimedia contents*, , SPIE, Vol. 3657, pp. 159-170, Jan 1999.
[4] Zwicker, E. and H. Fastl. Psychoacoustics: Facts and Models. Springer-Verlag, Berlin, Germany, 1990.
[5] Haykin, S. Communication Systems. 3rd Edition. John Wiley and sons, 1994.
[6] Milleth, J. K., K. Giridhar, and D. Jalihal. Performance enhancement of space-time trellis codes when encountering AWGN and Ricean channels. *In IEEE Trans. on vehicular technology*, Vol. 54, No. 4, July 2005.
[7] Boney, L., Ahmed H. Tewfic, Khaled N. Hamdy. Digital watermarks for audio signals. *In proceedings of multimedia computing and systems*, pp.473-480,1996.
[8] Lee, S. K. and Y. S. Ho Digital audio watermarking in Cepstral Domain. *IEEE Transactions on Consumer Electronics*, Vol.46, No.3,August 2000, pp.744-750.
[9] Painter, E. M and A. S. Spanias. A review of algorithms for perceptual coding of digital audio signals. *13th International Conference on Digital Signal Processing Proceedings*, DSP-97, vol.1,July 1997,pp.179-208.

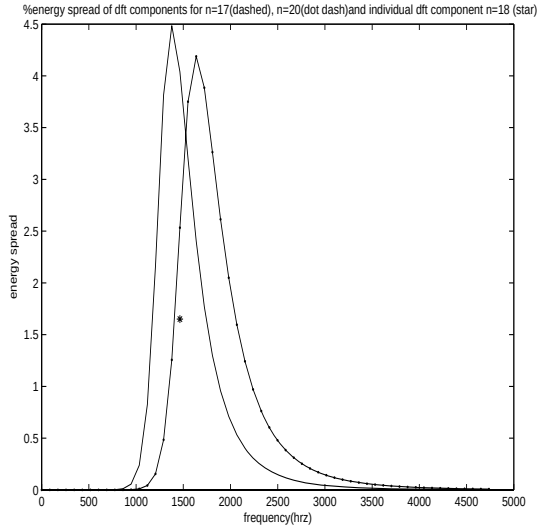

Fig. 1. Amplitude spread of DFT components for k=17 and 20, star represents individual DFT component at k=19

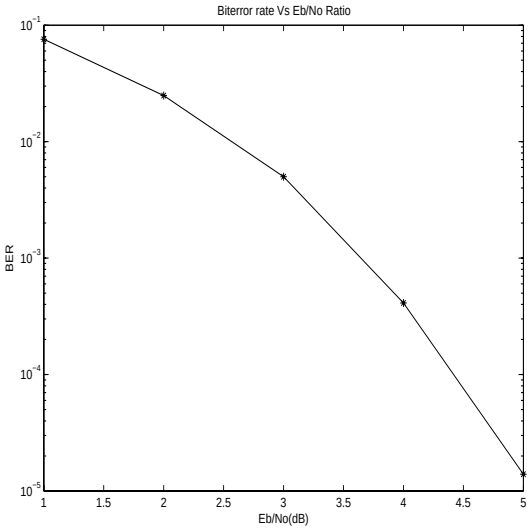

Fig. 3. bit error rate vs Eb/No

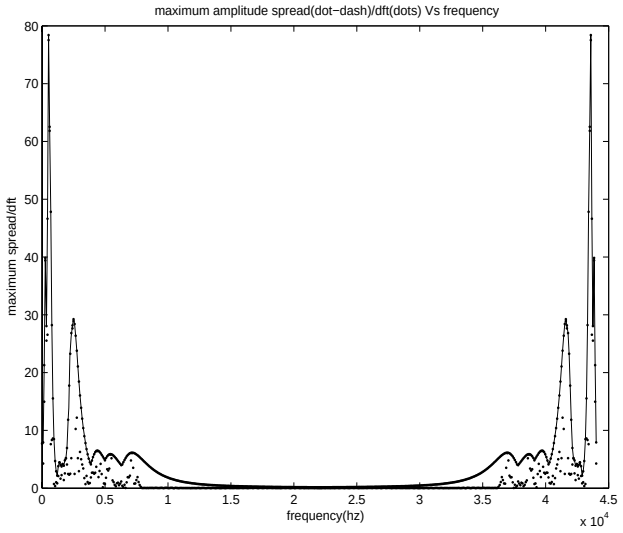

Fig. 2. Maximum amplitude spread for frame $N = 512$

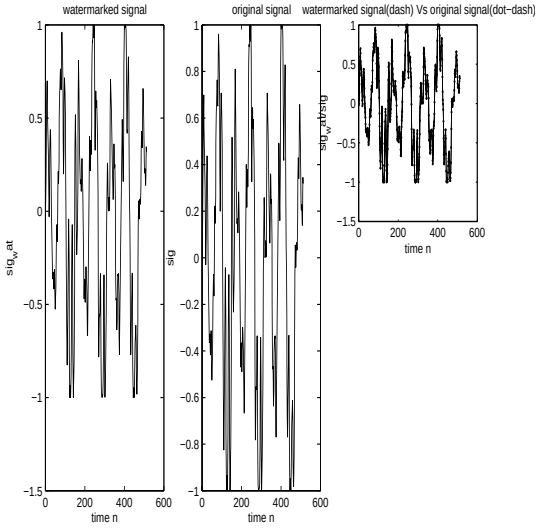

Fig. 4. watermarked signal(red),original signal(yellow) and watermarked imposed on original signal for frame $N = 512$

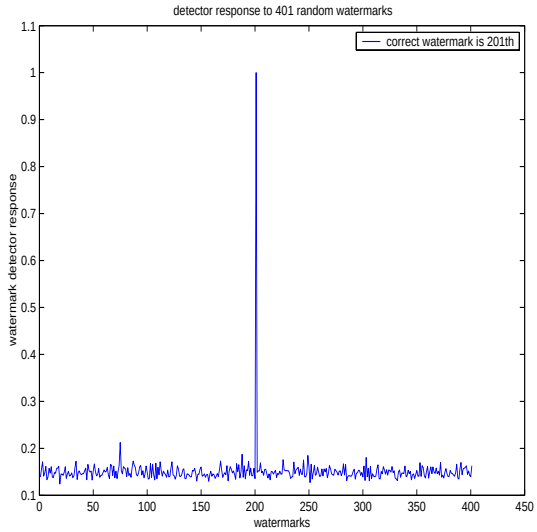

Fig. 5. Detector response to 401 randomly generated watermarks,showing peak at the location of correct watermark $k = 201$

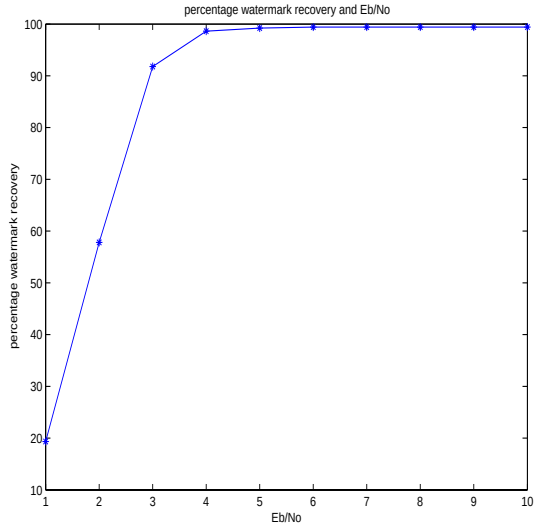

Fig. 6. Percentage watermark recovery vs Eb/No