I[2]MTC 2008 – IEEE International Instrumentation and Measurement Technology Conference
Victoria, Vancouver Island, British Columbia, Canada, May 12–15, 2008

An Adaptive Watermarking Algorithm for MP3 Compressed Audio Signals

Bingwei Chen[1], Jiying Zhao[1], Dali Wang[2]

[1]Multimedia Communications Research Laboratory, School of Information Technology and Engineering, University of Ottawa,
800 King Edward Ave., Ottawa, ON, Canada K1N 6N5
Phone: +1 613 5625800, Fax: +1 613 5625664, E–mail: {bchen, jyzhao}@site.uottawa.ca.

[2]Department of Physics, Computer Science and Engineering, Christopher Newport University
1 University Place, Newport News, Virginia 23606
Phone: +1 757 5947828, Fax: +1 757 5947919, E–mail: dwang@pcs.cnu.edu.

Abstract *– MP3 has been generating a significant popularity for distributing digital audio signals over the Internet. However, copyright issue has been raised because of illegal copy and redistribution. Digital watermarking is a technology that allows users to embed watermark into digital contents to identify the copyright holder, to prevent illegal copy, and to verify modification to the original content. This paper proposes a novel adaptive watermarking algorithm for MP3 compressed audio signals, based on human auditory system. In the proposed algorithm, watermark is embedded adaptively and transparently after Modified Discrete Cosine Transformation (MDCT) and before quantization. Gaussian distribution statistic analysis is introduced to make this watermarking algorithm adaptive. We tested our adaptive watermarking algorithm on various types of audio signals. The experimental results show that the new algorithm can survive most common signal manipulation including MP3 compression.*

Keywords *– Digital audio watermarking, MP3 compression, human auditory system, Gaussian distribution.*

I. INTRODUCTION

Audio signals are an important member of the multimedia family. MP3 has been generating a significant popularity for distributing digital audio signals over the Internet. Watermarking audio signals has been researched for copyright protection, content authentication, quality evaluation, etc. In the literature, quite a number of audio watermarking algorithms have been proposed. However, only a few of them directly embed watermark in the MP3 stream to provide superior performance against MP3 compression.

Wang et al. [1] proposed a watermarking algorithm to embed watermark during MDCT within a few coefficients during compression. The authors choose to embed watermark into low frequency coefficients. But if watermark, which is noise to the original audio signal, is embedded in low frequencies, it would impact the quality of the original audio the most.

Tachibana [2] proposed to use a two-dimensional pseudorandom array as watermarking identification key to embed and extract watermarks in MPEG-2 Advanced Audio Coding (AAC) bitstream. Detection is done by correlating the magnitudes of the frame contents with the pseudo-random array. According to the conclusion of the paper, the robustness of the method is weaker than that of uncompressed-domain watermarking and the acoustic quality has not been given [2].

Wu et al. [3] proposed a self-adaptive video watermarking algorithm. They introduced the statistical analysis for adaptive watermark embedding, but they cannot implement blind watermark detection.

In this paper, we will propose a new adaptive audio watermarking algorithm in MP3 compression domain, based on human auditory system.

II. PROPOSED ADAPTIVE AUDIO WATERMARKING ALGORITHM

The proposed watermarking algorithm works directly on MP3 frames, and uses psychoacoustic model and Gaussian distribution analysis to adaptively control the watermarking parameters.

A. *Watermark Embedding*

We use Gaussian distribution analysis on MP3 frames with energy consideration to adjust watermarking parameters. We implement the proposed adaptive watermarking algorithm during MP3 encoding. Figure 1 presents the diagram of the watermark embedding algorithm.

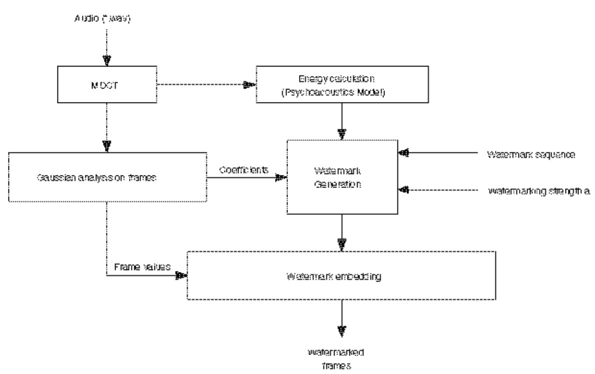

Fig. 1. Diagram of watermark embedding algorithm.

A.1 Watermarking position

We choose to embed watermark after Modified Discrete Cosine Transform (MDCT) and before quantization. During the MP3 encoding process, two-time compressions happen. One is the hybrid filterbank with psychoacoustic model II masking thresholding and the other is the quantization. Quantization in MP3 is a non-uniform quantization mapping amplitude values into finite number of bits. There are two-nested loop during the quantization. One is the inner iteration loop to control the quantization step size and the other one is the outer iteration loop to control the noise shaping factors for each scalefactor band.

We aim at a watermarking algorithm that performs particularly well for MP3 compression in terms of inaudibility and robustness. Watermark should be embedded with maximum-possible energy (robustness) by considering human auditory system [4]. Therefore, we decide to embed watermark after MDCT and before quantization.

A.2 Watermarking frame structure

Frames forms MP3 audio stream after the MDCT. According to the experimental results, if two (or more) original frames (32×18 matrix) are merged to a macro frame (32×36 matrix), the MDCT values in the macro frame follow the Gaussian distribution. One original frame alone can also be used to carry watermark, but the performance is not as good as when two frames are merged. After many experiments, it is decided that merging two frames together will generate good result while not taking too much watermarking space. We will use Gaussian analysis on frames with other parameters adjustments to make watermarking to be adaptive.

When an audio signal is stereo, for the simplicity of implementation, only one channel (right channel) is used even though there are two channels. So, actually the implementation is on mono audio (single channel) and the granule is a frame.

A.3 Energy calculation

The bigger the energy in a subband, the stronger watermark can be embedded inaudibly. In this paper, the MP3 audio energy is calculated by using the psychoacoustics model II defined in ISO/IEC 11172 - Part 3: Audio [5].

Table B in ISO/IEC 11172 - Part 3: Audio, is used to map the subband energy to corresponding MDCT values.

The proposed watermarking method uses subband energy to adaptively adjust watermarking strength.

A.4 Embedding steps

The following gives the detailed watermark embedding steps.

1. Generate 64-bit watermark sequence and expand the sequence to 192 bits by dictionary mapping. Each bit of the expanded sequence will be embedded into one macro frame.
2. Determine the first suitable watermarking frame. For example, the first non-zero frame can be the candidate.
3. Merge two neighboring frames (each frame with 32 subbands, and each subband with 18 frequency lines) to form one macro frame (32×36) for watermark embedding.
4. Generate the watermark pattern, (32×36) matrix, which is the same size as an macro frame. The content of the pattern is all "+1" or all "−1"to represent a bit of watermark "1" or a bit of watermark "0".
5. Generate Gaussian analysis values for all 32×36 MDCT values in the macro frame. That means, this step generates 1152 Gaussian analysis values.

   The following equations show the detail Gaussian distribution calculation on each macro watermarking frame:

$$f(X_{ij}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(X_{ij}-\overline{x})^2/(2\sigma^2)} \tag{1}$$

$$\overline{x} = \frac{1}{32\times 36}\sum_{i=1}^{32}\sum_{j=1}^{36} X_{ij} \tag{2}$$

$$\sigma = \sqrt{\frac{1}{32\times 36}\sum_{i=1}^{32}\sum_{j=1}^{36}(X_{ij}-\overline{x})^2} \tag{3}$$

   where, X_{ij} ($i = 1, 2, \cdots, 32$, $j = 1, 2, \cdots, 36$) represents an MDCT values in a macro watermarking frame.
6. Calculate energy for scalefactor bands in MDCT. See Section A.3.
7. Form the final watermark according to energy calculation and Gaussian analysis results. Equ. (4) through Equ. (5) illustrate the watermark embedding.

$$X'_{ij} = X_{ij} + \Delta X_{ij} \tag{4}$$

   where, X'_{ij} (i=1, 2, $\cdots$, 32 and j=1, 2, $\cdots$ 36) represents the watermarked MDCT values in a macro frame, and

$$\Delta X_{ij} = \alpha * \frac{log(energy_{ij})}{E} * f(X_{ij}) * W_{ij} \tag{5}$$

   where, E =100; $\alpha = 0.000001$ is set based on trial and error. $f(X_{ij})$ is the Gaussian analysis value for X_{ij} in a frame, and W_{ij} is watermark. If watermark bit "0" is embedded into a macro frame, the actual watermark W_{ij} is all "−1" representatively; If watermark bit "1" is embedded into a macro frame, the actual watermark W_{ij} is all "+1" representatively.

B. Watermark Extraction

Gaussian distribution analysis is used to detect the possibility of the watermark existence in a frame and make the blind watermark detection possible. When the analysis determines that the watermark is in the frame, we will try to extract watermark after alias detection and before IMDCT.

The following gives the detailed steps for the blind watermark extraction.

1. Find the first frame that contains watermark. According to the embedding algorithm, the first bit of watermark is embedded in the first non-zero frame. Therefore, the first non-zero frame should be the first detection candidate.
2. Merge two neighboring frames to one frame.
3. Apply Gaussian distribution analysis on all the MDCT values in the macro frame for watermark detection. Watermark W'_o is extracted based on Gaussian distribution analysis function D (X'_{ij}, σ) on watermarked macro frame.

$$W'_o = \sum_{i=1}^{32} \sum_{j=1}^{36} D(X'_{ij}, \sigma) \quad (6)$$

where, n=1152 is the number of the MDCT values in a macro frame, and

$$D(X'_{ij}, \sigma) = \left(\frac{1}{\sqrt{2\pi * \sigma}} \right) * e^{-\left[\frac{-\left[X'_{ij} - \sigma(X'_{ij})\right]}{n} \right]^{\frac{2}{2*\sigma^2}}} \quad (7)$$

where, the Gaussian distribution threshold $\sigma = 0.0005$ is determined based on experiments.
During detection, the sum W'_o of the Gaussian values is calculated in macro frame. If W'_o is greater than "+1", we determine that a watermark bit "1" was embedded into the original macro frame. If W'_o is less than "-1", we determine that a watermark bit "0" was embedded into the original macro frame. If W'_o is between $[-1, +1]$, we determine that no watermark was embedded into that macro frame.
In Figure 2, the distribution of detection values from an original macro frame with no watermark , whose Gaussian values is between $[-1, +1]$, is illustrated by a green line. The left curve is the result with a watermark bit 0 embedded and the right curve is the result with a watermark bit 1 embedded.

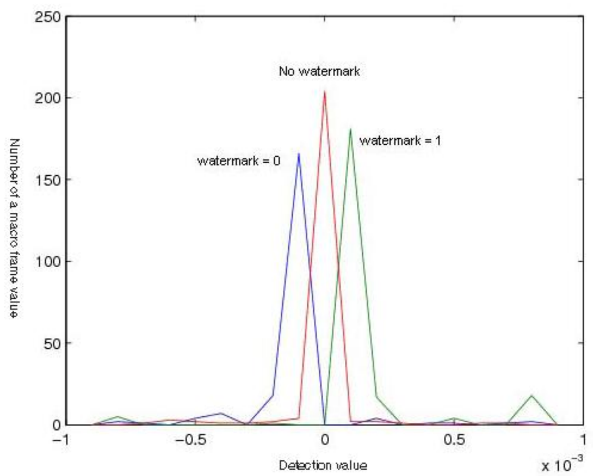

Fig. 2. Detection results for watermarked and non-watermarked macro frame.

4. Look up the dictionary table to extract watermark out.

III. EXPERIMENTAL RESULTS AND EVALUATIONS

In order to evaluate the performance of the proposed watermarking algorithm, the following experiments were considered. LAME 3.96.2 open source code [7] is used, which is an MP3 encoder and decoder based on ISO/IEC 11172 Standard - Part 3: Audio. The original LAME source code has been modified and we added the proposed watermarking algorithm in. The experiments were conducted on a personal computer with Intel Pentium core 2 duo 2.80 GHz CPU and 1 GB RAM. Music 1 and Music 2 are classic music, Music 3 and Music 4 are pop music, and Music 5 is a speech signal.

A. *Performance on different types of audio*

Watermark was embedded into the five different types of 16-bit signed mono audio signals sampled at 44.1 kHz. Table I gives the experimental results in terms of SNR (Signal-to-Noise Ratio), PSNR (Peak Signal-to-Noise ratio), and BER (Bit Error Rate).

TABLE I. 44.1-kHz 16 bit mono audio

Audio	SNR	PSNR	BER
Music1	32.8202	57.5957	0
Music2	31.0485	56.3534	0
Music3	28.9383	53.3593	0
Music4	29.9110	57.1185	0
Music5	27.8907	55.4103	0

B. *PESQ*

PESQ stands for Perceptual Evaluation of Speech Quality. MOS (Mean Opinion Score) is used to express the speech quality. MOS score is from -0.5 (worst) to 4.5 (best).

Table II and Table III show the PESQ MOS values for the watermarking algorithm with the proposed adaptive control and without the adaptive control. In Table II and Table III, all the BER values are 0. The embedding energies for both adaptive and non-adaptive algorithms are implemented as close as possible.

TABLE II. MOS comparison

	32 kbps		48 kbps		64 kbps	
	adaptive	non-adaptive	adaptive	non-adaptive	adaptive	non-adaptive
Music1	3.775	3.625	4.099	3.691	4.053	3.939
Music2	4.307	4.077	4.294	4.105	4.356	3.995
Music3	4.064	3.294	3.854	3.719	4.005	3.538
Music4	3.938	3.734	4.020	3.720	3.626	3.617
Music5	4.027	3.569	3.946	3.557	3.948	3.523

The experimental results show that the algorithm with the adaptive control gives better subjective quality.

TABLE III. MOS comparison (cont.)

	80 kbps		128 kbps		320 kbps	
	adaptive	non-adaptive	adaptive	non-adaptive	adaptive	non-adaptive
Music1	4.293	3.927	4.293	3.951	4.287	3.933
Music2	4.333	4.000	4.351	4.151	4.352	4.019
Music3	4.162	3.533	4.149	3.877	4.135	3.522
Music4	3.965	3.676	4.187	3.524	4.162	3.706
Music5	3.873	3.565	3.970	3.451	3.977	3.570

C. MP3 compression

We tested our algorithm on 16-bit mono and stereo audio signals sampled at 32 kHz, 44.1-kHz, and 48-kHz. The test results are similar. Due to limited space, in this section, we will give only the BER performance results for 32 kHz mono audio. See Table IV.

TABLE IV. BER under MP3 compressions on 32 kHz 16-bit mono audio

Compression	Embedding bit rates (kbps)						
bit rates (kbps)	32	48	64	80	128	160	320
32	0	33	33	34	36	31	27
48	0	0	0	0	0	0	0
64	0	0	0	0	0	0	0
80	0	0	0	0	0	0	0
128	0	0	0	0	0	0	0
160	0	0	0	0	0	0	0
320	0	0	0	0	0	0	0

D. Low-pass filtering

Table V showed the performance against the low-pass filtering attack. The cutoff frequency was set to 20 kHz.

TABLE V. BER against low-pass filtering (20 kHz)

Audio	SNR	PSNR	BER
Music1	26.6670	50.0690	0
Music2	20.3884	47.9038	0
Music3	15.9261	37.3472	5
Music4	22.4124	48.3351	0
Music5	15.9261	37.3472	1

The BERs for Music 3 and 5 are not 0%. Music 3 is a very fast and strong pop music. Music 5 is a speech audio, whose energy is concentrated in the middle frequencies, and it does not show much rhythm.

E. Gaussian noise

We set Gaussian noise to be 10 dB and then added this period of noise on watermarked audio in order to simulate the transmission through a communication channel with noise and also to simulate the attack with a noise added to watermarked audio intending to blur the watermark. The BERs for Music 1, 2, 4 are 0. And the BERs for Music 3 and 5 are respectively 4 and 1.

F. Performance on various compression rates

To demonstrate that the proposed algorithm works with various MP3 compression rates, we tested our algorithm with the most common used compression rates, such as 32 kbps, 48 kbps, 64 kbps, 80 kbps, 128 kbps and 320 kbps.

Figure 3 summarizes the results of this experimental test. It shows the watermarking algorithm's performance is stable under most common MP3 compression bit rates. Note BER curve is not visible in the figure, because all the values are 0.

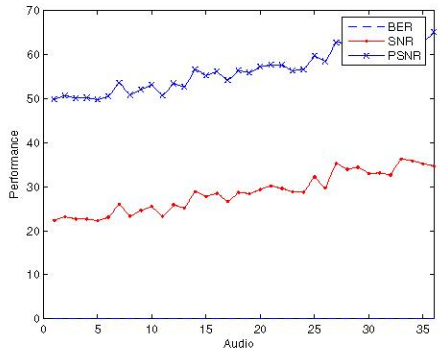

Fig. 3. Performance on various compression bit rates.

IV. CONCLUSIONS

In this paper, an adaptive digital audio watermarking algorithm for MP3 compression has been proposed. Watermark is embedded directly in MP3 domain after MDCT and before quantization in the process of MP3 encoder. Human auditory system is used. Gaussian distribution analysis on frames and original audio energy of subbands are used for adaptive control. Watermark detection is blind. The experimental results show that this new watermarking algorithm is robust against MP3 compression and can survive most common attacks. The results are preliminary, and will be improved.

[1] C.T. Wang, T.S. Chen, and W.H. Chao, "A new audio watermarking based on modified discrete cosine transform of MPEG/audio layer III", Proceedings of the IEEE International Conference on Networking Sensing and Control, pp. 984-989, 2004.

[2] R. Tachibana, "Two-dimensional audio watermark for MPEG AAC audio", Proceedings of SPIE, Vol. 5306, Security, Steganography, and Watermarking of Multimedia Contents VI, pp. 139-150, 2004.

[3] G.M. Wu, W.T. Zhuang, F. Wu, and Y.H. Pan, "Self-adaptive MPEG video watermarking based on block perceptual features", Proceedings of the Third International Conference on Machine Learning and Cybernetics, pp. 3871 - 3876, 2004

[4] L. Boney, A.H. Tewfik, and K.N. Hamdy, "Digital watermarks for audio signals", Proceedings of IEEE International Conference on Multimedia Computing and Systems, pp. 473-480, 1996

[5] International Standard ISO/IEC 11172, Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5M bit/s - Part 3: Audio 1993.

[6] I.J. Cox, M.L. Miller, J.Bloom, "Digital watermarking", Morgan Kaufmann, IBSN: 1-55860-714-5, 2002.

[7] http://lame.sourceforge.net/