AN AUDIO WATERMARKING SCHEME ROBUST TO MPEG AUDIO COMPRESSION

*Won-Gyum Kim, *Jong Chan Lee and Won Don Lee*

Dept. of Computer Science, ChungNam Nat'l Univ., Daeduk Science Town, Taejon, Korea
*Dept. of Artificial Intelligence, ChungWoon Univ., Hongsung, ChungNam, Korea
{wgkim, jclee, wdlee}@brain.cs.chungnam.ac.kr

ABSTRACT

In recent years, digital watermarking has been introduced as a means of effectively protecting copyrights on the digitized media such as image, audio and multimedia data. In this paper, we present a technique for embedding digital "watermarks" into digital audio signals. The watermark must be imperceptible and should be robust to attacks and other types of distortions. In our method, the watermark is generated by the random sequence with a seed and is embedded into the subband coefficients directly for the robustness of MPEG audio compression. The seed is a watermark key known only the copyright owner.

1. INTRODUCTION

The proliferation of digitized media(audio, image, and video) is creating a pressing need for copyright enforcement schemes that protect copyright ownership. A method of copyright protection is the addition of a "watermark" to the digitized media. The watermark is a digital code embedded in the audio data and is inaudible, or preferably inaudible. A digital watermark is permanently embedded in the data: that is, it remains present within the original data after any distortion process. A watermark could be used to provide proof of authorship of a signal.

For digital watermarking of audio[1][2][3], a number of different characteristics of the watermarking process and watermark are desirable. These requirements are:

- **Inaudible** : The digital watermark embedded into the audio data should be inaudible to the human
- **Security** : Unauthorized removal and detection of the watermark must be impossible even if the basic scheme used for watermarking is known.
- **Robustness** : It should be impossible to manipulate the watermark by intentional or unintentional operations without degrading the perceived quality of the audio to the point of significantly reducing its commercial value. Such operations are, for example, filtering, re-sampling, compression, noise, cropping, A/D-D/A conversions, etc.
- **Constant bit-rate** : Watermarking in the bitstream domain should not increase the bit-rate.
- **Embedded directly in the data**, not in the header.
- **Multiple watermarks**

In this paper, we present a technique for embedding digital watermarks into audio signals. The watermark is generated by the pseudo-random sequence with a seed number and is embedded into the subband coefficients directly for the robustness of MPEG/audio compression. Unlike other methods[1][2], our watermark is embedded in the frequency domain: that is, it modifies each subband coefficients directly according to its value.

This paper is organized as follows. In Section 2, we discuss MPEG/audio compression and the polyphase filter bank. In Section 3, we describe how we embed and extract a watermark from the audio data. Experimental results are given in section 4 confirming the performance of the presented schemes. Finally we summarize our major findings and outline our future work.

2. BACKGROUND

2.1 MPEG/Audio Compression

The MPEG/audio compression algorithm is the first international standard for the digital compression of high-fidelity audio. MPEG/audio is a generic audio compression standard. Unlike vocal-track-model coders specially tuned for speech signals, the MPEG/audio coder gets its compression without making assumption about the nature of the audio source. Instead, the coder exploits the perceptual limitations of the human auditory system. Much of the compression results from the removal of perceptually irrelevant parts of the audio signal. Removal of such parts results in inaudible distortions, thus MPEG/audio can compress any signal meant to be heard by the human ear.

MPEG/audio offers a choice of three independent layers of compression. Layer Ⅰ is the simplest and is best suited for bit rates above 128 kbits/s per channel. Layer Ⅱ has an intermediate complexity and is targeted for bit rates around 128 kbits/s per channel. Layer Ⅲ is the most complex but offers the best audio quality, particularly for bit rates around 64 kbits/s per channel.

The key to MPEG/audio compression is quantization. Although quantization is lossy, this algorithm can give transparent, perceptually lossless, compression. Figure 1 shows block diagram of the MPEG/audio encoder and decoder. The input audio stream passes through a filter bank that divides the

This research is supported by Korea Science and Engineering Foundation(KOSEF) under grant 97-0100-06-01-3

input into multiple subbands of frequency. The input audio stream simultaneously passes through a psychoacoustic model that determines the ratio of the signal energy to the masking threshold for each subband. The bit or noise allocation block uses the signal-to-mask ratios to decide how to apportion the total number of code bits available for the quantization of the subband signals to minimize the audibility of the quantization noise. Finally, the last block takes the representation of the quantized subband samples and formats this data and side information into a coded bitstream. The decoder deciphers this bitstream, restores the quantized subband values, and reconstructs the audio signal from the subband values.

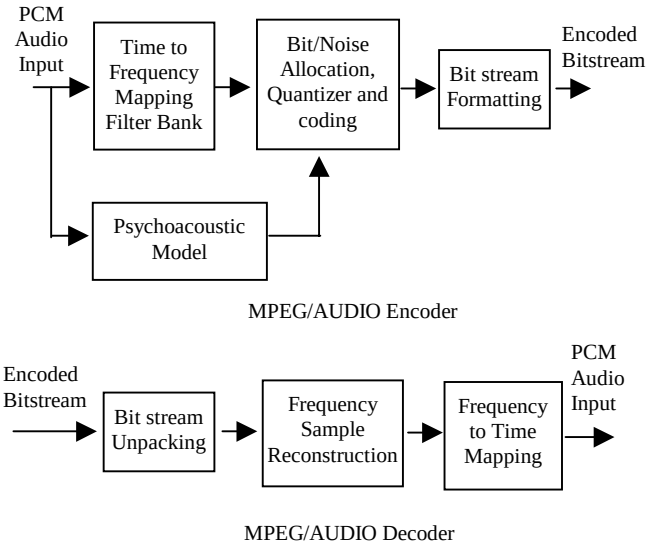

Figure 1. MPEG/Audio Compression/Decompression

2.2 The Polyphase Filter Bank

The filter bank divides the audio signal into 32 equal-width frequency subbands. There are two filterbanks used in the MPEG/audio algorithm, a polyphase filterbank and a hybrid polyphase/MDCT filterbank. Each provides a specific mapping in time and frequency. These filterbanks are critically sampled.

We define the filterbank outputs:

$$St[i] = \sum_{k=0}^{63} \sum_{j=0}^{7} M[i][k] * (C[k + 64j] * x[k + 64j])$$

where,

i : The subband index and ranges from 0 to 31.

$S_t[i]$: The filter output sample for subband *i* at time *t*, where *t* is an integer multiple of 32 audio sample intervals.

C[n] : One of 512 coefficients of the analysis window defined in the standard.

x[n] : An audio input sample read from a 512-sample buffer.

$M[i][k]$: $\cos\left[\frac{(2*i+1)*(k-16)*\pi}{64}\right]$,

The analysis matrix coefficients.

2.3 The Psychoacoustic Model

The MPEG/audio algorithm compresses audio data in large part by removing the acoustically irrelevant parts of the audio signal. That is, it takes advantage of the human auditory system's inability to hear quantization noise under conditions of auditory masking. This masking is a perceptual property of the human auditory system that occurs whenever the presence of a strong audio signal makes a temporal or spectral neighborhood of weaker audio signals imperceptible.

The psychoacoustic model analyzes the audio signal and computes the amount of noise masking available as a function of frequency. The masking ability of a given signal component depends on its frequency position and its loudness. The encoder uses this information to decide how best to represent the input audio signal with its limited number of code bits. The MPEG/audio standard provides two example implementations of the psychoacoustic model. Psychoacoustic model 1 is less complex than psychoacoustic model 2 and has more compromises to simplify the calculation. Each model works for any of the layers of compression. In this work, we use the psychoacoustic model 1 defined in MPEG/audio algorithm, for layer 1. Below is a general outline of the basic steps involved in the psychoacoustic calculations.

Convert audio to a frequency domain representation. : The psychoacoutic model should use a separate, independent, time-to-frequency mapping instead of the polyphase filter bank because it needs finer frequency resolution for an accurate calculation of the masking thresholds. The sychoacoustic model uses a Fourier transform for this mapping. Standard Hann weighting, applied to the audio data before Fourier transformation, conditions the data to reduce the edge effects of the transform window.

Separate spectral values into tonal and non-tonal components. : Tonal(sinusoidal) and non-tonal(noisy) components are identified because the masking abilities of the two types of signal are different.

Apply a spreading function. : The masking ability of a given signal spreads across its surrounding critical band. The model determines the noise masking thresholds by first applying an empirically determined masking.

Set a lower bound for the threshold values : Determine absolute masking threshold, the threshold in quiet. This threshold is the lower bound on the audibility of sound.

Find the masking threshold for each subband : The model calculates the masking threshold with a higher frequency resolution than that provided by the polyphase filterbank

Calculate the signal-to-mask ratio : The sychoacoustic model computes the signal-to-mask ratio as the ratio of the signal energy within the subband to the minimum masking

threshold for that subband. The model passes this value to the bit(or noise) allocation section of the encoder.

2.4 Bit Allocation

The bit allocation process determines the number of code bits to be allocated to each subband based on the information from the psychoacoustic model. This process starts by computing the mask-to-noise ratio as given by the following equation :

$$MNR_{dB} = SNR_{dB} - SMR_{dB}$$

Where,

MNR_{dB} : Mask-to-noise ratio

SNR_{dB} : Signal-to-noise ratio

SMR_{dB} : Signal-to-mask ratio

from the psychoacoustic model

Once the bit allocation unit has mask-to-noise ratios for all the subbands, it searches for the subband with the lowest mask-to-noise ratio and allocates code bits to that subband. When a subband gets allocated more code bits, the bit allocation unit looks up the new estimate for the signal-to-noise ratio and re-computes the mask-to-noise ratio of that subband. The process repeats until no more code bits can be allocated.

3. WATERMARKING APPROACHES

3.1 Watermark embedding

The watermark embedding scheme proposed in this paper modifies coefficients directly in subband with the maximum number of bits allocated. The result is a slight modification of each coefficient in a way that does not produce any perceptual difference.

To generate the watermark, we first divide the audio signal into 32 equal-width frequency subbands using the polyphase filterbank algorithm and calculate the masking threshold of the signal using the MPEG/audio psychoacoustic model 1, as described above. The masking threshold is determined for audio segment of 384 samples weighted by a hanning window. The psychoacoustic model computes the signal-to-mask ratio as the ratio of the signal energy within the subband to the minimum masking threshold for that subband and determines the number of code bits to be allocated to each subband. The subband with the maximum number of code bits to be allocated is selected to embed the watermark. Then we generate watermark sequence in proportion to a subband coefficient value.

Let us denote

$$p_i, p_i \in \{-1,1\}, \quad 0 \le i < N$$

which is a pseudo-random sequence of watermark bits that has to be embedded into the selected subband. The sequence p_I is amplified with a locally adjustable factor *x(i)*, where *x(i)* is the watermark value in proportion to the subband coefficient value itself.

We define watermark signal *x(i)* as

$$x(i) = \alpha * p_i * v(i), \quad 0 \le i < N$$

where, α (≥ 0) is a adjustable scale factor, and *v(i)* is a subband coefficient. The scale factor α may be varied according to local properties of the audio signal and can be used to exploit temporal masking phenomena of the HAS such that the amplitude of the watermark is locally as large as possible without becoming audible.

The watermarked subband signal *v'(i)* is therefore :

$$v'(i) = v(i) + x(i) = v(i)(1 + \alpha * p_i), \quad 0 \le i < N$$

Due to the noisy nature of the pseudo-noise signal p_i, the watermark signal, *x(i)* is also a noise-like signal and thus difficult to detect and manipulate. Figure 2 shows block diagram of the overall embedding process of the watermark.

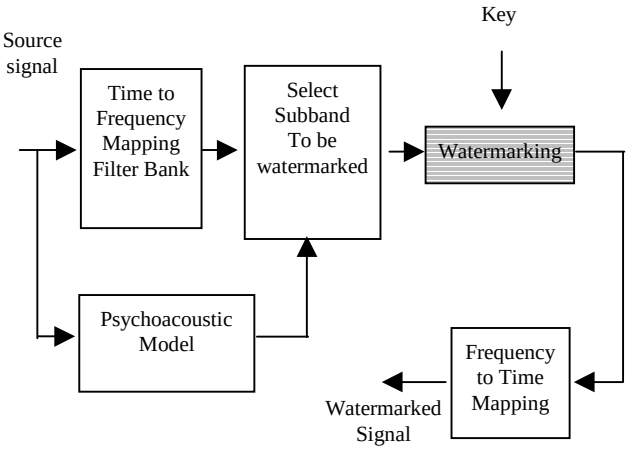

Figure 2. Embedding process of the watermark

3.2 Watermark retrieval

In order to recover the embedded bit, an original audio signal is needed. At first, we devide the source audio signal into 32 equal-width frequency subbands and calculate the masking threshold of the signal using the MPEG/audio algorithm as is done in embedding process. Then we select subbands to be watermarked and generate reference watermark sequence s_i, using the owner's key. The owner performs subband decoding on the marked signal at the same time, and retrieves the watermark by taking the sign of the difference between watermarks.

After retrieving the watermark, the user can compare the results with the referenced watermark subjectively. A similarity measure of the extracted and the referenced watermarks can be defined as :

$$Correlation \quad Value(CV) = \frac{\sum_{i=0}^{N-1} \overline{X} \sum_{i=0}^{N-1} X}{\sum_{i=0}^{N-1} X \sum_{i=0}^{N-1} X}$$

where, X is the total number of embedded bits, and $\overline{X}$ is the total number of matched bits between extracted watermarks and referenced watermarks.

If the wrong pseudo-noise sequence is used, or if it is not in synchronization with the pseudo-noise sequence used for embedding, the recovered watermark is random.

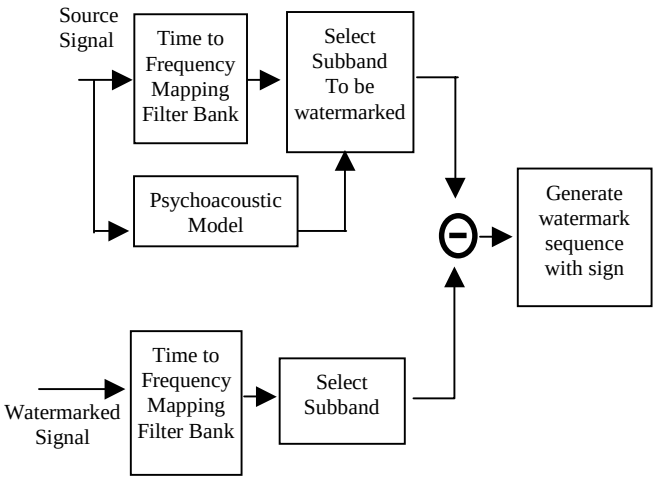

Figure 3. Detection process of the watermark

4. EXPERIMENTAL RESULTS

We have performed detection tests on a 16-bits signed mono audio signal of length 384*1395 sampled at 44100Hz. The scale factor α is set to 0.1. The correlation value(CV) lies between 0 and 1. If the wrong pseudo-noise sequence or invalid key is used, the CV is set to around 0.5. Experimentally, a watermark threshold may be set above 0.75, in order to decide whether a certain watermark exists in the signal. In our test, 1000 keys have been used to detect a watermark in a watermark signal. Figure 4 shows the result of test for watermarked signal without any kind of attack.

The robustness of the watermark has been tested using layers Ⅰ of the MPEG/audio algorithm. In figure 5, watermarking detection after decompression indicates a slight decrease of the correlation value in the watermarked signal. Since the correlation value is above the threshold, we have succeeded in watermark detection in this experiment.

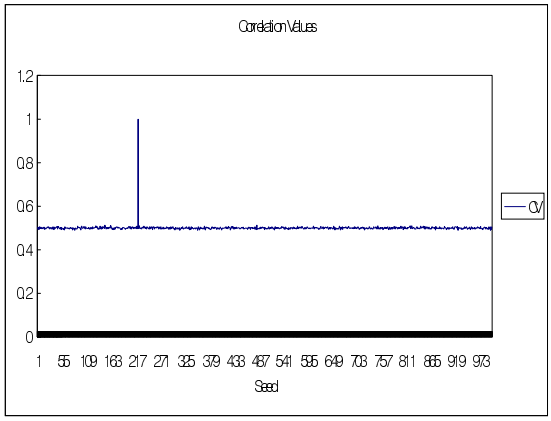

Figure 4. Detection values in a watermarked signal using various seeds.(Key=222, CV=0.99)

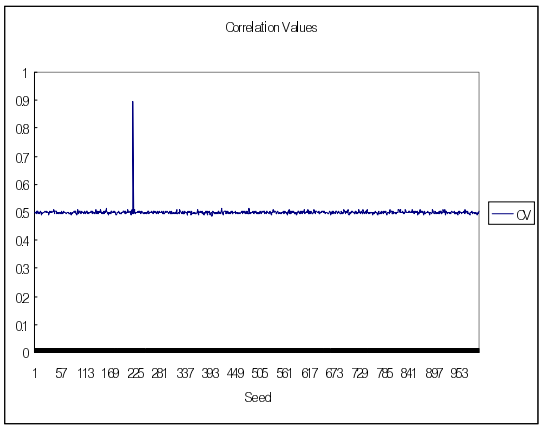

Figure 5. Detection values in a watermarked signal using various seeds after MPEG audio compression. (Key=222, CV=0.89)

5. SUMMARY

Our proposed watermark algorithm is to insert and retrieve watermark directly in the frequency domain. To make it robust to the MPEG/audio compression attack, we divide audio signal into 32 equal-width frequency subbands using the polyphase filtebank algorithm which is used in the MPEG/audio algorithm. The number of bits to be allocation for each subband is calculated by the signal-to-mask information of the psychoacoustic model. And finally, the watermark is inserted to be most bit-allocated subband. The embedded watermark is the bitstream of {1, -1} generated by the pseudo-random sequence and the seed number is the key which would be acknowledged by the copyright owner only. The owner performs subband decoding on the source signal while decoding on the watermarked signal at the same time, and retrieves the watermark by taking the sign of the difference between the coefficients. Though similar study[1] on MPEG/audio algorithm has been achieved recently, our algorithm is different from those

as follows. Tewfik's algorithm repeats mapping to the time-domain after generating watermark sequence using masking threshold of psychoacoustic model. However, in our algorithm, the embedding of the watermark is achieved directly on the coefficient of the frequency domain. That can abbreviate embedding process and give the strong robustness to the other kind of attack like filtering by inserting watermark directly to the frequency domain.

6. REFERENCES

[1] L. Boney, H. Tewfik, K. N. Hamdy. "Digital Watermarks for Audio Signals", *Proc. of EUSIPCO'96*, Trieste, Italy, 1996.

[2] V. Basia and I. Pitas, "Robust Audio Watermarking in the time-domain", *Proc. of EUSIPCO'98*, September 8-11, Rhodes, Greece, 1998.

[3] D. Gruhl, A. Lu, W. Bender, "Echo Hiding", *Info Hiding 96*, pp.295-315

[4] I. J. Cox, J. Kilian, T. Leighton, T. Shamoon. "Secure spread spectrum watermarking for Multimedia", *Tech. Report, NEC research Institute*, 95-10, p.1.

[5] Scott Craver, Nasir Memon, Boon-Lock Yeo and Minerva Yeung. "Can invisible watermarks resolve rightful ownerships?", *Proc. of the IS&T/SPIE Conference on Storage and Retrieval for Image and Video Databases V*, San Jose, CA, USA, FEB. 13-14, 1997, vol.3022, pp.310-321.

[6] E. Koch and J. Zhao. "Toward robust and hidden image copyright labeling", in *Proc. of 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, June 1995.

[7] A.G. Bors and I. Pitas, "Image watermarking using DCT domain constraints", in *Proc. IEEE Int. Conference on Multimedia Computing and Systems*, pp.86-90, 1994.

[8] W. Bender, D. Gruhl and N. Morimoto, "Techniques for data hiding", *in Proc. of SPIE*, volume 2420, page 40, FEB. 1995

[9] F. Hartung and B. Girod. "Digital watermarking of raw and compressed video". *In N. Ohta, editor, Digital Compression Technologies and Systems. For Video communications, volume 2952 of SPIE Processings Series*, pp205-213, Oct., 1996.

[10] ISO/IEC 13818-2, "Generic Coding of Moving Pictures and Associated Audio", Recommendation H.262 (MPEG-2), 1995, International Standard.

[11] D. Pan, "A Tutorial on MPEG/audio Compression", *IEEE Multimedia Summer* 1995