World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:1, No:12, 2007

Genetic Content-Based MP3 Audio Watermarking in MDCT Domain

N. Moghadam, and H. Sadeghi

***Abstract*—** In this paper a novel scheme for watermarking digital audio during its compression to MPEG-1 Layer III format is proposed. For this purpose we slightly modify some of the selected MDCT coefficients, which are used during MPEG audio compression procedure. Due to the possibility of modifying different MDCT coefficients, there will be different choices for embedding the watermark into audio data, considering robustness and transparency factors. Our proposed method uses a genetic algorithm to select the best coefficients to embed the watermark. This genetic selection is done according to the parameters that are extracted from the perceptual content of the audio to optimize the robustness and transparency of the watermark**.** On the other hand the watermark security is increased due to the random nature of the genetic selection. The information of the selected MDCT coefficients that carry the watermark bits, are saves in a database for future extraction of the watermark. The proposed method is suitable for online MP3 stores to pursue illegal copies of musical artworks. Experimental results show that the detection ratio of the watermarks at the bitrate of 128kbps remains above 90% while the inaudibility of the watermark is preserved.

***Keywords*—**Content-Based Audio Watermarking, Genetic Audio Watermarking.

I. Introduction

DIGITAL audio watermarking is the process of embedding inaudible information into audio data for the purposes of copyright protection, fingerprinting, carrying information and content manipulation detection of the musical artworks. *Robustness*, *Transparency, Data Payload* and *Security* are some of the important parameters of watermarking algorithms in watermarking literature from which *robustness* and *transparency* seem to be more significant. Robustness is covered by the survival of the watermark after modifying watermarked audio data and transparency is referred as the inaudibility of the embedded watermark and appearing insensible changes in audio data after embedding the watermark [1,2,3].

The wide use of compressed audio in MP3 format on the internet and the massive loss in the profits of the artists that is caused by illegally copying MP3 files, clarify the necessity of proposing a method for the protection of copyright for musical artworks in MP3 format. The design of MP3 file format and impossibility of modifying the content of MP3 files without reducing the quality, make us think of a method of watermarking that is carried out during the MP3 compression procedure. By this policy, the online MP3 stores can provide every online user with a unique version of their requested material that carries a watermark within itself. This watermark is embedded in the MP3 file during its compression, when the artwork is requested to be downloaded. In this method, the time-consuming compression and watermarking procedure can be handled in parallel with the streaming download process of generated file.

This paper is organized in six sections including the related works, MP3 compression fundamentals, the proposed watermarking method, experimental results and the conclusion.

II. Related Works

Many audio watermarking methods have been proposed so far. A group of them try to embed extra information into audio data by adding *echoes* to the signal. Extraction of the watermark in this method is carried out by cepstrum analysis of the watermarked signal [6,7]. In *spread spectrum* watermarking method, the watermark is generated based on a PN sequence and then it is weighted in the time (or frequency) domain. A correlation test can be useful for extracting the watermark [4,5,18]. *Patchwork* watermarking method is carried out by correspondingly adding and subtracting a constant value from two sets of the samples [8]. The main feature of the content-based watermarking schemes [8,9] is their dynamic behavior in accordance with the audio content. Different watermarking methods have been presented for MPEG audio too. Some of them try to embed extra information into MP3 frames by modifying the *scale factors* in the MP3 frames [11]. Scale factors are the multipliers that are used during requantizing the quantized MDCT coefficients in the decoding process. This method is simple but it affects the quality of the sound deeply. Some other methods like what is described in [12,13] try to embed extra bits in MP3 audio during compression by changing the termination condition of the *inner loop*, which is used during MP3 quantization process. These methods are sensitive to attacks and have a low degree of robustness (e.g. the watermark will be lost by decoding and re-encoding the same signal to MP3 format).

In this paper a novel MP3 watermarking scheme is proposed that optimizes the transparency and robustness of the watermark by a genetic algorithm. The proposed method also provides a higher degree of security due to the random nature

Manuscript received on May 20, 2005. This work was supported in part by the Iranian Telecommunication Research Center (ITRC).

of its behavior which is also dependent on the audio content.

III. MP3 Compression Fundamentals

MP3 is one of the most popular audio formats because of the preserved quality and high compression ratio. In this section we describe the MP3 compression method.

A. Mp3 Encoding

Fig. 1 shows the block diagram of an MP3 encoder. The inputs of the encoder are 1152-sample frames of the signal each of which is divided into two 576-sample granules. The FFT and the psychoacoustic model are then used to detect the *masking threshold* for all frequencies. On the other hand, the input signal is transformed into 576 subband samples using the analysis filterbank and a *Modified Discrete Cosine Transform*. The masking thresholds are used to iteratively determine how many bits are needed in each critical band to code the samples so that the quantization noise is not audible. This phenomenon is carried out by two nested loops named the *inner loop* and the *outer loop*.

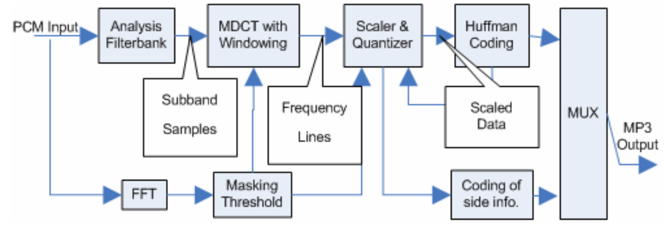

Fig. 1 An MP3 Encoder

The quantized samples are coded by the Huffman method and stored in the bit stream along with the scale factors and side information that form the MP3 bit stream [14-17].

B. MP3 Decoding

As shown in Fig. 2, after unpacking the MP3 frame, Huffman decoding and requantization, the MDCT coefficients are obtained and then used to create the PCM audio frames using a synthesis polyphase filterbank [14-17].

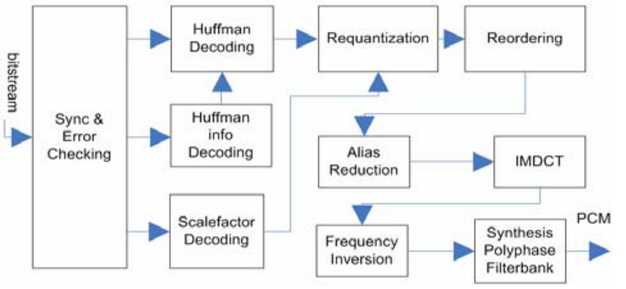

Fig. 2 An MP3 Decoder

IV. The Proposed Watermarking Method

Our watermarking procedure is integrated with the MP3 compression procedure and is implemented as a new MP3 compressor that embeds watermarks into the original audio file that it receives as its input in PCM WAV format.

A. Generating the Watermark Bit Sequence

For generating the watermark bit sequence we use a similar method that is described in [18]. In this procedure, the main watermark (e.g. the user ID of an online user) is passed to a hash function whose output number will be used as a seed for a random number generator. If we let L_w be the length of the watermark bit sequence then the random number generator is invoked for L_w times and each time a "0" or "1" is inserted into the bit sequence, according to the *sign* of the random number generator's output.

B. Embedding the Watermark Bits in MDCT Coefficients

After obtaining 576 MDCT coefficients during compressing each granule, a genetic algorithm is invoked, whose task is to select L coefficients to embed the first remained *L* bits of the watermark bit sequence into the related granule. If the number of the remained watermark bits is greater than L, the rest of them will be hidden in the next MDCT coefficients that will be calculated during compressing the rest of the signal. Embedding each bit in the MDCT coefficients is carried out in the same way similar to [19]. Each bit of the watermark is hidden in one MDCT coefficient. Let $Z(X)$ be the *number of the zeros after the decimal point on the MDCT coefficient* X ,(e.g. $Z(X=0.03456)=1$). In brief the new value, X', of each coefficient after embedding the bit , B, is calculated as follows:

$$X' = \begin{cases} X & ;if\ (B=1)\&(Z(X)\, is\, odd) \\ Sing(X)\times 5\times 10^{-Z(X)-2} & ;if\ (B=1)\&(Z(X)\, is\, even) \\ X & ;if\ (B=0)\&(Z(X)\, is\, even) \\ Sing(X)\times 5\times 10^{-Z(X)-2} & ;if\ (B=0)\&(Z(X)\, is\, odd) \end{cases} \qquad (1)$$

Experiments show that during the quantization phase in MP3 encoding, although the values of the MDCT coefficients are modified (e.g. very close-to-zero values will be replaced by zero), the number of their zeros after the decimal point is rarely changed. As the result, hiding extra bits can be done according to the odd and even state of the number of the zeros after the decimal point on the coefficients. Although one choice for adding one zero to the zeros of a coefficients is to divide its value by 10, but this is not the optimum choice due to the possibility of intense changes in the value of the coefficient during the quantization process. As the result, for increasing the number of the zeros of the coefficient X, we replace its value with $sign(X)\times 5\times 10^{-Z(x)-2}$. With this policy for the new value of X, the probability of being changed during quantization is reduced and the probability of losing the watermark bits is decreased. Experiments show that such modifications do not impose an intensive degradation to the audio quality even at lower bitrates.

C. The Optimum Genetic Selection

Each chromosome of the GA is a set of L unrepeated indexes of the chosen MDCT coefficients as follows:

$$I = \{N_k : 0 \le N_k < 576 \quad \text{and} \quad 0 \le k < L \\ \text{and} \quad \forall a,b : \quad \text{if} \ \ a \ne b \Rightarrow N_a \ne N_b\} \tag{2}$$

During each of the GA's iterations, the fitness values of the chromosomes are calculated in accordance with the *masking thresholds* that are extracted from the audio content by the psycho-acoustic model of the MP3 encoder as follows:

$$Fitness\ (I) = \alpha \cdot \lambda_T + \beta \cdot \lambda_R \tag{3}$$

Where α is the weight of transparency factor, λ_T, and β is the weight of robustness factor, λ_R. These factors are calculated as follows:

$$\lambda_T = \sum_{m=0}^{SFBc-1} (MT[m] - \sum_{n=SFBs[m]}^{SFBs[m+1]} (\frac{(X'[n] - X[n])^2}{SFB[m+1] - SFB[m]})) \tag{4}$$

$$\lambda_R = \sum_{k=0}^{L-1} \frac{-(Z(X'[I[k]]))^4 + 25}{L} \tag{5}$$

Table I describes the parameters in eq. (4) and (5). λ_T represents a criterion for the perceptual noise that is imposed to the signal because of changing the coefficients and is calculated in the similar way to MP3 quantization noise. λ_R represents a criterion for the robustness of watermark bits. Experiments show that the probability for variation of *Z(X)*, which is equivalent to the probability of losing the watermark bit, increases when *Z(X)* increases so that for *Z(X)>2* , X will be replaced by zero during quantization. As the result, the chromosomes with such *X* es should not gain a high fitness. Moreover, if the GA selects the coefficients with smaller number of zeros then the watermark will be more resistant against attacks, because the number of zeros in such coefficients is rarely changed by attacks. For each granule of the signal, the genetic selection is performed to find the best coefficients to embed the bits, considering transparency and robustness of the watermark in accordance with the content of the audio and the algorithm is done repeatedly till all the watermark bits are hidden into audio signal. The GA starts with a population of randomly generated individuals. Each GA iteration is carried out by *selection*, *crossover*, *mutation* and *convergence testing* [20]. These steps are done according to the population size P_{size}, crossover probability $P_{crossover}$, and mutation probability $P_{Mutation}$. If all the watermark bits are hidden before the compression process of the whole signal is over, the watermark bit sequence can be embedded in the rest of the signal for more times to increase robustness.

D. Saving the Selected MDCT Indexes for Future Extraction

It's clear that for the same audio signal, the watermarking algorithm may have different choices for embedding different user IDs into the MP3 version of that signal. Due to the content-based behavior of the genetic watermark embedding procedure, it is needed to save the sequence of the indexes of the selected MDCT coefficients for every watermarked MP3 file in a database for future extraction. This information can be considered as a *secondary key* for extracting the watermark.

TABLE I
PARAMETERS IN EQ. (4) AND EQ. (5)

Parameter	Description
SFBc	Total number of scale factor bands (for MPEG-I equals to 22).
MT[m]	The maximum masking threshold of the m_{th} scale factor band.
SFBs[m]	The index of the first MDCT coefficient, which is related to the m_{th} scale factor band. This array has 23 elements.
$X'[n]$	The new value for the n_{th} coefficient after embedding the corresponding watermark bit according to eq. (1).
$X[n]$	The original value for the n_{th} coefficient.
L	The number of bits that are going to be hidden in every granule of the signal, in other words the length of the chromosomes.
$I[k]$	The k_{th} gene in every chromosome that indicates the index of a selected coefficient.

E. Watermark Extraction

The following hypothesis test is used to detect the watermark:

H_0 : `The MP3 file does not contain the user ID`

H_1 : `The MP3 file contains the user ID,` Uid

The extraction process is handled in the similar way to MP3 decoding procedure. After re-generating the bit sequence of the watermark according to Uid and fetching the related secondary watermarking keys from the database, the ordinary MP3 decoding process is carried out. The extracted MDCT coefficients, whose indexes are in correspondence with the indexes of the bit-carrying coefficients, are then tested. For every bit-carrying extracted coefficient, X_x, the hidden bit is zero if $Z(X_x)$ is even and vice versa. The detection ratio is defined as follows:

$$DR = \frac{\sum_{k=0}^{L_w-1} (\overline{W'_k \oplus W''_k})}{L_w} \tag{6}$$

Where W'_k is the k_{th} extracted watermark bit, W''_k is the k_{th} bit of the re-generated watermark bit sequence, L_w is the length of the watermark bit sequence and $\oplus$ is the binary EX-OR operator that returns 0, if its both operands are equal and vice versa. The result of comparing the detection ratio with a selected threshold, τ, can lead us to the existence of the watermark within the MP3 file.

V. Experimental Results

The proposed watermarking method has been implemented using 8Hz source code [21] as the embedding/compression engine and the MPGLib source code [22] as the extraction engine. Fig. 3 illustrates the detection ratio at different bitrates with typical parameters ($P_{Mutation} = 0.1$, $P_{Crossover} = 0.7$, $L = 10$, $L_w = 100$, $\alpha = 20$, $\beta = 1$, $P_{size} = 200$).

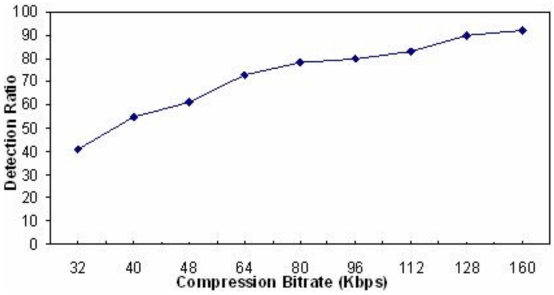

Fig. 3 Detection ratio at different bitrates

Fig. 4 illustrates the number of the times that each MDCT coefficient has been selected by the GA during watermarking the first 10 granules of a sample signal with the same typical parameters at 128kbps.

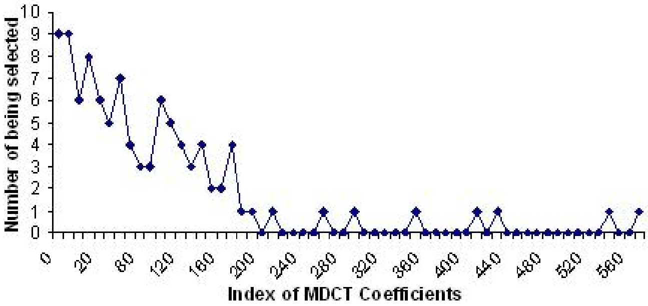

Fig. 4 The number of times that each MDCT coefficient has been selected

As can be seen in Fig. 4, the first MDCT coefficients have been selected for more times because of their higher absolute value (consequently less zeros after the decimal point) and their lower effect on the audio quality.

The length of the chromosomes, L, is very effective on the detection ratio. As shown in Fig. 5, experiments show that for L>5 the detection ration decreases because finding better coefficients for the GA gets harder.

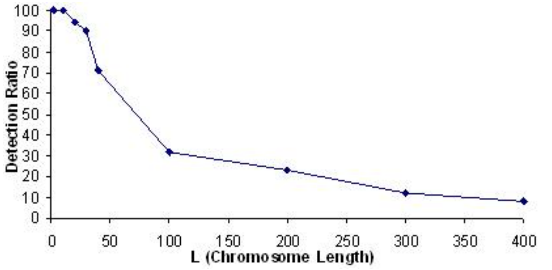

Fig. 5 A greater chromosome length decreases the detection ratio

Experiments show that for L<5 the best balance between transparency, robustness and data payload is obtained while the algorithm can embed a 16000-bit watermark into a signal that is shorter than 120 seconds. It's while the watermarked MP3 can hardly be detected from the original MP3 version of the signal, even by the music professionals.

Conclusion

The presented watermarking method can guarantee a detection ration of 90% at the bitrate of 128kbps, while preserving the quality of the original sound. The proposed method can be useful for online MP3 stores, if the phenomenon of compression and watermarking are done in parallel with the process of downloading the requested MP3 files. Future studies can be done to reduce the dependency of the proposed algorithm on the secondary extraction keys.

References

[1] Ingemar J. Cox, Matthew L. Miller, Jeffrey A. Bloom, "Digital Watermarking" , Morgan Kaufmann Publications, 2002

[2] Thomas Page, "Digital Watermarking As A Form Of Copyright Protection", Computer Law & Security Report Elsevier, November-December 1998

[3] Christine I. Podilchuk and Edward J. Delp, "Digital Watermaking, Algorithms and Applications", IEEE SIGNAL PROCESSING MAGAZINE, p 33- 46, July 2001

[4] Darko Kirovski and Henrique S. Malvar, Fellow, "Spread-Spectrum Watermarking of Audio Signals", IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 51, NO. 4, APRIL 2003

[5] I.J. Cox, J. Kilian and T. Leighton: "Secure spread spectrum watermarking for multimedia", IEEE Trans. Image Processing 6:1673–1687,1997

[6] D. Gruhl, A. Lu and W. Bender: "Echo hiding". Proc. of Information Hiding Workshop, University of Cambridge, pp. 295–315, 1996

[7] H. Oh, H.W. Kim, J.W. Seok, J.W. Hong and D.H. Youn: "Transparent and robust audio watermarking with a new echo embedding technique". Proc. of IEEE Int. Conf. on Multimedia and Expo, Tokyo, Japan, 2001

[8] Yeo, I.-K., and Kim, H.J., "Modofied patchwork algorithm: A novel audio watermarking Scheme", IEEE Transactions on Speech and Audio Processing, vol. 11, (to appear), 2003

[9] Chung-Ping Wu, Po-Chyi Su and C.-C. Jay Kuo, "Rubust and Efficeint Digial Audio Watermarking Using Audio Content Analysis", IS&T/SPIE 12th International Symposium on Electronic Imaging, (San Jose, CA, USA), Jan. 2000, Proceedings of SPIE Vol. 3971, pp.382-392

[10] Changsheng Xu, David Dagan Feng, "Robust and Efficient Content-Based Digital Audio Watermarking", International Conference of Multimedia Systems, 2000

[11] Nahrstedt, L. Qiao, in Dittmann et al., "Non-Invertible Watermarking Methods for MPEG Video and Audio", '6th ACM International Multimedia Conference (ACM Multimedia'98)' ACM, Bristol, England, Sept. 1998, pp. 93–98

[12] http://www.petitcolas.net/fabien/steganography/mp3stego

[13] http://web.uvic.ca/~jinxing/wmp3/CSC%20586B%20Final%20Project%20Report.doc

[14] Pan D, "A Tutorial On Mpeg Audio Compression". IEEE Multimedia 2(2): p 60–74, 1995.

[15] ISO/IEC JTC1/SC29/WG11 (MPEG), International Standard ISO/IEC 11172-3, "Coding Of Moving Pictures And Associated Audio For Digital Storage Media At Up To About 1.5 Mbits/S", 1992

[16] ISO/IEC JTC1/SC29/WG11 (MPEG), International Standard ISO/IEC 13818-3, "Generic Coding of Moving Pictures and Associated Audio: Audio", 1994

[17] P. Noll, "MPEG Digital Audio Coding," IEEE Signal Processing Magazine, pp.59-81, Sep. 1997.

[18] Dimitrios Simitopoulos , Sotirios A. Tsaftaris , Nikolaos V. Boulgouris ,Alexia Briassouli and Michael G. Strintzis, "Fast Watermarking Of Mpeg-1/2 Streams Using Compressed-Domain Perceptual Embedding And A Generalized Correlator Detector", EURASIP Journal on Applied Signal Processing, 2004

[19] Ching-Te Wang, Tung-Shou Chen, Wen-Hung Chao, "A new audio watermarking based on Modified Descrete Cosine Transform of MPEG audio layer III", Proceedings of the 2004 IEEE international conference on networking, sensing and control, Taipei, Taiwan, March 21-23, 2004

[20] D.E. Goldberg, "Genetic Algprithms In Search, Optimization And Machine Learning", Addision Wisely, 1989

[21] http://www.8HZ.com/MP3

[22] http://www.mpg123.de