International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 - 5161

Available at http://inpressco.com/category/ijcet

Research Article

Digital Audio Watermarking techniques with Musical Audio Feature Classification

Faizur Rahman[Å*] and Nitin N. Mandaogade[Å]

[Å]Dept. of Electronics & Telecommunication, GHRCEAM, Amravati, Maharashtra, India

Accepted 05 Sept 2014, Available online 01 Oct 2014, **Vol.4, No.5 (Oct 2014)**

Abstract

Digital audio watermarking is an important and popular technique along the original musical audio content producers. Watermarks can be utilised as evidence for proving the ownership. Musical audio can be characterized by the common characteristics shared in low-level, mid-level and high-level features. These characteristics typically are timbre, temporal, rhythmic, pitch, harmony, genre, mood, instrument, and many more which can be extracted from the music. Automatic musical audio feature extraction is utilised for identifying genre and its respective utilization can be carried out in audio watermarking area so as to improve the performance. The paper describes the techniques developed till date for audio watermarking and audio feature extraction techniques in brief.

Keywords: *Digital Audio Watermarking, Musical Audio, Musical Features, Genre, Timbre.*

1. Introduction

Development in technological era has made the production, storage and distribution of digital multimedia data very easy, effective and efficient with increased reliability and flexibility. This technological development has advantage and dis-advantage both at the same time. Since it is very easy to manipulate digital information the illegal production and redistribution of digital media has taken a large share of the overall digital multimedia market. Thus it has threatened the owner regarding security and integrity of original media which can be easily duplicated and distributed using the developed technologies of internet, USB mass storage devices and many more. This creates the problem of protecting the intellectual copyrights of multimedia data. The copyrighted digital multimedia data is pirated without any consent of the owner (Al-Haj A *et al*, 2011).

Embedding a secondary image, video or audio data into the primary image, video or audio data provides a way to insert owners' or ownership information in it. The secondary data is termed as watermark and a process of embedding watermark is known as watermarking. Watermark as the name suggest is as transparent as water when watermark data is embedded in the primary data. The watermark can be extracted later to prove the ownership of the data under dispute. Watermarks must remain integral under transmission and any kind of intentional or un-intentional signal processing attacks. Lack of a watermark or degradation in it would lead to the decision that the primary data has been altered.

Watermarking performance can be judged on parameters like imperceptibility, robustness, efficiency, and embedding capacity. A watermarking technique must achieve high performance without degrading the primary signal.

Digital watermarking can be applied to image, audio or video and the watermark data can be an image, audio and text. Since, the HAS (human auditory system) is more sensitive than HVS (human visual system) and human ears can easily detect the presence of the watermark as low as one part in ten million (Khatri G. B *et al*, 2013); hence, the amount of work in the area of audio watermarking are less. This motivated us in audio watermarking and more specifically in musical audio watermarking.

To improve the performance of watermarking one must have the proper knowledge of genres in the music. The musical audio can be categorised on the basis of various low, middle and high level features of music and various algorithms have been developed for the same.

In section II of this paper, an overview on the techniques of digital audio watermarking has been collectively presented. Section III describes the musical audio features and its classification techniques; in section III an algorithm is proposed to include the effect of different genres of music on audio watermarking and vice-versa. Section IV concludes the paper on the overall topics presented.

2. Digital Audio Watermarking

The watermarking techniques for the audio are explored in less amount and the available techniques are very few. On the basis of survey done the audio watermarking can be majorly categorised into four groups viz. time domain, frequency domain, spread spectrum and patchwork, which are represented here to analysing the performances:

Time Domain Techniques

*Corresponding author **Faizur Rahman** is a ME Student and **Nitin N. Mandaogade** is working as HOD

It is the area where the least significant bit (LSB) substitution; echo hiding and quantization techniques are employed. In this parameters of signal samples like amplitude, masking of samples with lower amplitude by higher amplitude samples, samples LSB are varied to achieve embedding of watermark. It is very easy to implement the time-domain audio watermarking and requires few computing resources; however, it is less immune to signal processing attacks such as compression and filtering. It suffers from problems like low watermark data embedding capacity, easily detectable by the attacker, easy to decode watermark from primary audio.

LSB technique uses the simplest technique in which watermark data is embedded in the least significant bits of the audio sample values which ultimately helps in having easy data embedding and extracting algorithms. As the information contained in the LSB is less, it is replaced by the data of the watermark without producing the noticeable effect in the primary audio signal. The primary audio signal degrades as the number of watermark bits is increased. At the maximum 3 watermark bits per 16 bits of audio sample is allowed for imperceptibility. The noise becomes detectable by human auditory system if embedded sample rate goes above 3 bits per audio sample. Cvejic and Seppanen have tried to increase the capacity from 3–4bps without degrading the watermarked audio signal to noise ratio by using a three step technique. Minimum error replacement and error diffusion steps are used to minimise the degradation in SNR of the watermarked audio (Cvejic N., Seppanen T *et al*, 2012).

Echo-hiding watermarking embeds information into the original discrete audio signal by introducing a repeated version of an original sample of the audio signal with some delay and decay rate so as to make it undetectable (Bender W., Gruhl D., *et al*, 1996). Only binary information is embedded in the audio signal in the form of bits. Digital data is embedded by using four main parameters of echo: initial value, decay rate and different offset for 1 and 0. The offset is made small enough to make the presence of echo non-detectable for the human ear. Embedding is done by convolving the audio signal with the all 0 and all 1 kernel, then by using watermark data bits particular outputs are combined to form watermarked signal. Extraction is carried out by taking autocorrelation of cepstrum of the watermarked audio. Autocorrelation provides the power of signals at various shifts. For any particular shift in it one can easily determine the bit embedded. The watermark data embedding rate is given as 16 bps (bits per second), while it can vary in the range 2–64 bps with its dependence on the sampling rate and the signal type to be echoed.

By the technique of quantization, original sample of audio is replaced with the modified audio sample. The modified audio sample is defined as below,

$$y = \begin{cases} q(x,A) + \frac{A}{4} \ , \ if\ data\ bit\ is\ 1 \\ q(x,A) - \frac{A}{4} \ , \ if\ data\ bit\ is\ 0 \end{cases} \quad (1)$$

Where q (.) is quantization function and A is quantization step. The quantization function is given as,

$$q(x,A) = [x/A].A \quad (2)$$

Where [x/A] is rounded to nearest integer. This shows that in a single sample of audio signal one can embed only one bit of watermark. Thus a blind detection can be applied for watermark data extraction. Extraction can be done by following equation,

$$b = \begin{cases} 1 \ , \ if & 0 < y - q(x,A) < A/4 \\ 0 \ , \ if & -A/4 < y - q(x,A) < 0 \end{cases} \quad (3)$$

This is simple and easy to implement and is robust to noise as long as the noise margin is not above A/4. While the technique is easy but it has a very low watermark embedding capacity.

Frequency Domain Techniques

Transforms like discrete Fourier transform (DFT), the discrete cosine transform (DCT), and the discrete wavelet transform (DWT) comes in Frequency domain techniques of audio watermarking. It takes the advantage of masking of different tones of human auditory system (HAS) for an efficient watermarking. In this the watermark is either added or replaced in the magnitude or phase response after transforming the original audio into frequency domain by using any one among DFT, DCT or DWT. Then inverse transform is used to get the watermarked audio in time domain.

By using Discrete Fourier transform the signal's fundamental and harmonically related sinusoidal frequencies can be extracted. The human ears sensitivity declines after the peak sensitivity around 1 kHz. Magnitude response coefficients are replaced by the watermark data in the frequency range of 2.4 – 6.4 kHz (Tilki J. F., Beex A. A., *et al*, 1996). Also the human ears are insensitive to the absolute phase of the audio frequency; hence the phase difference between the phase signal coefficient and phase reference coefficient is used to modify the phase signal coefficient. Phase difference has to be added or subtracted when the watermark data bit is 1 or 0 respectively [6/8].

The real valued coefficients are available when we use discrete cosine transform. Properties of DCT such as high compaction of signal energy in transform domain, highly decorrelated coefficients are used to embed data in the transform domain.

Discrete wavelet transform requires fewer calculations as compared to DFT and DCT to obtain the coefficients and it decomposes the signal in time and frequency at the same time. Several advantages of applying DWT on audio signal are given by wu and huang such as 1) It is able to localize the audio in time-frequency both with multi-resolution property, 2) variable decomposition levels are available, 3) less number of operations than DFT and DCT (Wu S., Huang J.., *et al*, 2005). If there are N samples in the audio then number of operations in DFT, DCT and DWT are $O(N \bullet log_2(N))$, $O(N \bullet log_2(N))$ and $O(L \bullet N)$ respectively, where L is the length of wavelet filter. A data payload capacity of 172 bps is achieved by embedding the self-synchronised watermark data in the wavelet domain without degrading the SNR too much (Wu S., Huang J.., *et al*, 2005).

Spread Spectrum Technique

This technique involves embedding of watermark in the original audio signal by spreading it over the bandwidth of audio signal (Bender W., Gruhl D., *et al*, 1996). It utilizes DSSS (Direct Sequence Spread Spectrum) technique of spread spectrum technology. The encoding of watermark in the audio is shown in Figure 1.

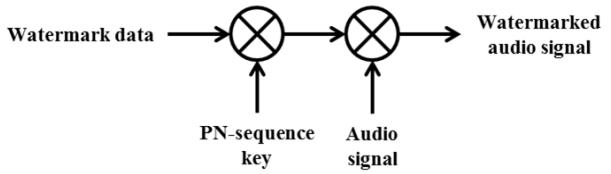

Figure 1 Direct Sequence Spread Spectrum Encoding

DSSS PN-sequence uses to spread the watermark data in the whole frequency range of audio and then added to the audio signal by proper attenuation, so that the watermark data is treated as additive random noise. For extracting the watermarked data same sequence is again used and by performing correlation between watermarked audio and PN-sequence it is extracted.

Patchwork Technique

This technique was first proposed for image watermarking and it is a pseudorandom statistical approach (Bender W., Gruhl D., *et al*, 1996). It involves a method in which it selects two subsets (patch) of the primary signal and to embed the watermark, the sample values of these two subsets are moved in opposite directions by a constant value d, which defines the watermark strength and watermark bit. The imperceptibility of watermark in primary signal depends on value of d. By taking the difference of the means of these two subsets and making decision based on the obtained value, the decoding is done. The assumption in this method is that the difference of the means of the two patches is zero for the original primary signal and is nonzero for the watermarked primary signal. This technique can extract the watermark without the original primary signal by using these two subsets (patch).

Yeo and Kim have proposed a modification on the patchwork technique (Yeo I. and Kim H., *et al*, 2003). They utilised the two subsets taken from DCT domain of original audio signal to embed the watermark data bit in. The technique becomes more robust against signal processing attacks such as down-sampling, equalization, compression, filtering by the use of transform domain for embedding watermark.

3. Musical Audio Features Classification

Audio is sound within the acoustic range available to humans. An audio frequency (AF) is an electrical alternating current within the 20 to 20,000 hertz (cycles per second) range that can be used to produce acoustic sound. Audio classification is shown in Figure 2.

Music can be divided into many categories based on styles, rhythm, and even cultural background. The styles are what we call the genres. The boundary of music genres is ambiguous and one song may belong to several genres with different weighting.

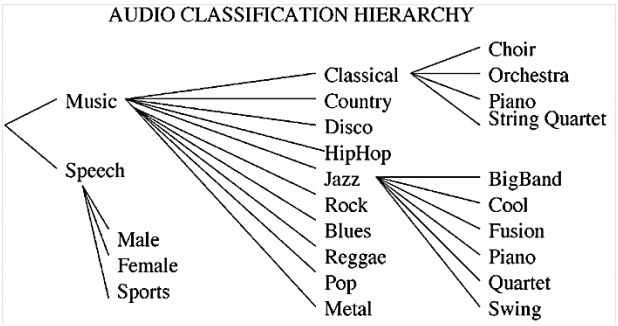

Figure 2 Audio Classification Hierarchy (Tzanetakis G., *et al*, 2002)

The key components of a classification system are feature extraction and classifier learning (Fu Z., Lu G.., *et al*, 2011). Feature extraction addresses the problem of how to represent the examples to be classified in terms of feature vectors or pairwise similarities. The purpose of classifier learning is to find a mapping from the feature space to the output labels so as to minimize the prediction error. We focus on music classification based on audio signals unless otherwise stated.

From the perspective of music understanding, we can divide audio features into two levels, low-level and mid-level features, as illustrated in the bottom two rows of Figure 3 along with top-level labels.

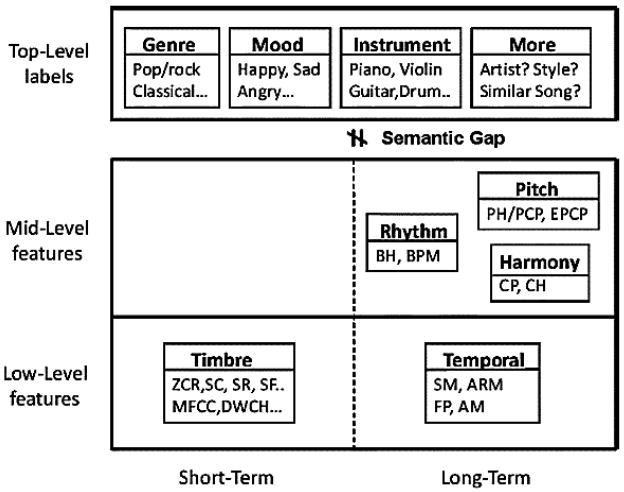

Figure 3 Characterizations of Audio Features (Fu Z., Lu G., *et al*, 2011)

Low-level features have two classes of timbre and temporal features. The tonal quality of sound that is related to different instrumentation is captured in timbre features, whereas temporal features capturethe variation and evolution of timbre over time. Using various signal processing techniques like Fourier transform, analysis such as spectral or cepstral, etc. low-level features can be obtained directly. Many other features are there in each class of low-level feature, as shown in Figure 3 by the

abbreviations below the name of the class in the box. Due to the fact that low-level features are easy to obtain and having good performance they are widely used. However, they are not closely related to the intrinsic properties of music as perceived by human listeners.

Based on features, namely rhythm, pitch, and harmony mid-level features provide a closer relationship. These features are usually extracted on top of low-level ones. At the top level, semantic labels provide information on how humans understand and interpret music, like genre, mood, style, etc.

Based on time frames or durations, audio features can also be categorized into short-term features and long-term features, as illustrated by columns of Figure 3. Short-term features like timbre features usually capture the characteristics of the audio signal in frames with 10–100 ms duration, whereas long-term features like temporal and rhythm features capture the long-term effect and interaction of the signal and are normally extracted from local windows with longer durations. Hence, the main difference here is the length of local windows used for feature extraction.

4. Proposed Algorithm

A model has been suggested here for embedding and extracting the watermark data in and out of musical audio signal which is proposed on the techniques presented in the previous section.

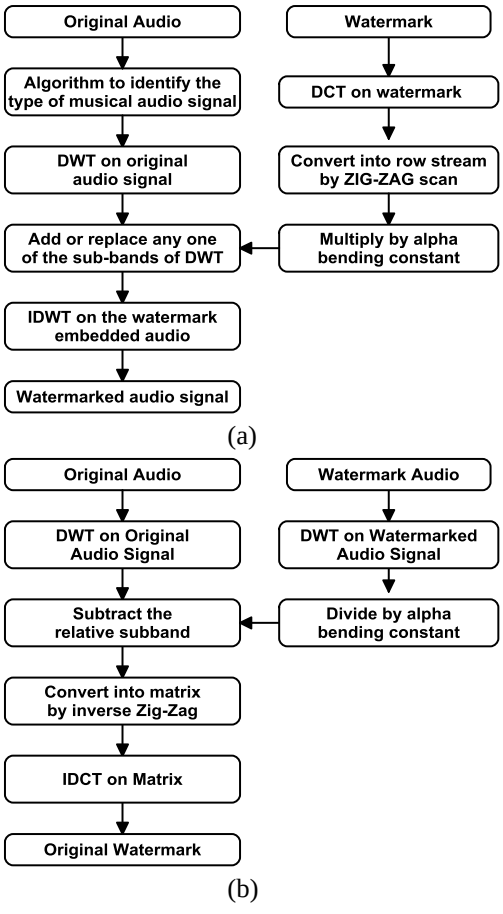

Figure 4 (a) Embedding Model (b) Extracting Model

Transform domain is secure and reliable than time domain and hence discrete wavelet transform has been proposed for speedily and efficiently transforming audio in time-frequency domain, while using discrete cosine transform to decorrelate and compress watermark image.

Transformed watermark image coefficients must be normalised and multiplied with an attenuation constant before embedding. Attenuating the coefficients helps to keep noise level low in the audio signal. Since watermark image is compressed using discrete cosine transform less number of transform coefficients are used for embedding and this improves the signal to noise ratio and also the watermark data embedding capacity.

In this new algorithm we will try out combinations of the previous algorithms to overcome the drawbacks such as low watermark embedding capacity, security to the most of the attacks and also as a more security the watermark can be scrambled before embedding. Furthermore, this system embeds the secondary data according to the category of music based on features of it.

Conclusions

Various algorithms of digital audio watermarking and the features of music for classifying the genre of music have been presented in the previous sections. The algorithm has been proposed such that it overcomes the limitation of normal digital audio watermarking by properly categorising the music genre, so that the performance can be increased. The proposed algorithm will overcome previous techniques and at the same time high performance with less computational cost will be achieved.

References

Al-Haj A., Mohammad A., Bata L. (2011), DWT based Audio Watermarking, *The International Arab Journal of Information Technology*, vol. 8, no. 3, pp. 326 – 333,.

Khatri G. B. and Chaudhari D. S. (March–April 2013), Digital Audio Watermarking Applications and Techniques, *International Journal of Electronics and Communication Engineering & Technology, IJECET/IAEME*, vol. 4, issue 2, pp. 109–115.

Cvejic N., Seppanen T. (2002), Increasing the Capacity of LSB based Audio Steganography, in *Proceedings of the IEEE International Workshop on Multimedia Signal Processing*, pp. 336–338.

Bender W., Gruhl D., Morimoto N., Lu A., Techniques for Data Hiding, *IBM Systems Journal*, vol. 35, no. 3 and 4, pp. 313–336, 1996.

Tilki J. F., Beex A. A. (1996), Encoding a Hidden Digital Signature onto an Audio Signal using Psychoacoustic Masking, *7th International Conference on Signal Processing Applications & Technology*, Boston MA, pp. 476–480.

Tilki J. F., Beex A. A. (1997), Encoding a Hidden Auxiliary Channel onto a Digital AudioSignal using Psychoacoustic Masking, *IEEE Southeastcon*, Blacksburg, VA, pp. 331–333.

Wu S., Huang J. (2005), Efficiently Self-Synchronized Audio Watermarking for Assured Audio Data Transmission, *IEEE Transactions on Broadcasting*, vol. 51, no. 1, pp. 69–7

Yeo I. and Kim H. (2003), Modified Patchwork Algorithm: A Novel Audio Watermarking scheme, *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 4, pp. 381–386

Tzanetakis G. and Cook P. (July 2002), Musical Genre Classification of Audio Signals, *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5.

Fu Z., Lu G., Ting K. M., and Zhang D. (April 2011), A Survey of Audio-Based Music Classification and Annotation, *IEEE Transactions on Multimedia*, vol. 13, no. 2.