Robust and Efficient Digital Audio Watermarking Using Audio Content Analysis

Chung-Ping Wu, Po-Chyi Su and C.-C. Jay Kuo
Media Fair, Inc., 1055 Corporate Center Dr., Ste 580
Monterey Park, CA 91754
and
Department of Electrical Engineering-Systems
University of Southern California, Los Angeles, CA 90089-2564
E-mail:{chungpin,pochyisu,cckuo}@sipi.usc.edu

ABSTRACT

Digital audio watermarking embeds inaudible information into digital audio data for the purposes of copyright protection, ownership verification, covert communication, and/or auxiliary data carrying. In this paper, we first describe the desirable characteristics of digital audio watermarks. Previous work on audio watermarking, which has primarily focused on the inaudibility of the embedded watermark and its robustness against attacks such as compression and noise, is then reviewed. In this research, special attention is paid to the synchronization attack caused by casual audio editing or malicious random cropping, which is a low-cost yet effective attack to watermarking algorithms developed before. A digital audio watermarking scheme of low complexity is proposed in this research as an effective way to deter users from misusing or illegally distributing audio data. The proposed scheme is based on audio content analysis using the wavelet filterbank while the watermark is embedd in the Fourier transform domain. A blind watermark detection technique is developed to identify the embedded watermark under various types of attacks.

Keywords: digital watermark, blind watermark detection, audio content analysis, synchronization attack, human auditory system, malicious cropping attack,wavelet

1. INTRODUCTION

Digital audio watermarking, the embedding and detection of an imperceptible signal in digital audio data, has received increasing attention recently. Among various different uses of digital audio watermarking, copyright protection is the most highly demanded application. The fast growth of the Internet and the maturity of audio compression techniques enable the promising market of on-line music distribution. However, since the digital technology allows lossless data duplication, illegal copying and distribution would be much easier than before. This concern does make musical creators and distributors hesitant to step into this market quickly. Therefore, the proper content protection technology is the key to the emergence of this new market.

Encryption and watermarking are the two most important content protection techniques. Encryption protects the content from anyone without the proper decryption key. It is useful in protecting the audio data from being intercepted during transmission. However, after the intended receiver decrypts it with the correct key, audio data could be illegally distributed and misused. Watermarks, on the other hand, cannot be removed from audio data even by the intended receiver. The embedded watermark signal permanently remains in audio data after repeated reproduction and redistribution. Thus, this signal could be used to protect the copyright of audio content by playback prohibition, illegal copy source tracing and ownership establishment.

Other applications of digital audio watermarking include data hiding for covert communication, auxiliary data embedding for audio content labeling, and modification detection for authentication. Data hiding can also be used to complement encryption, i.e. enhancing communication security by concealing the existence of sensitive data transmission. Embedded auxiliary data can carry lyrics or descriptions of the carrying audio data, or serve as links to external databases. Disappearance of fragile watermark could indicate unauthorized modifications and be used for content integrity verification.

Different watermarking applications have different sets of requirements. Here, our discussion is focused on copyright protection because it has the most stringent requirement on the watermark's ability to survive intentional

attacks. This is considered as one of the most challenging issues of the watermarking technology today. Users benefit from embedded label data while hackers do not know the existence of hidden communication data. Thus, embedded watermarks in these two applications are generally not subject to malicious attacks.

This paper is organized as follows. The requirements for audio watermarking systems are described in Section 2. Previous work on audio watermarking is reviewed in Section 3. Our current work on salient point extraction and Fourier domain watermarking is presented in Section 4. Experimental results and their analysis are given in Section 5. Finally, concluding remarks are provided in Section 6.

2. REQUIREMENTS FOR AUDIO WATERMARKING SYSTEMS

In order for the embedded watermark to effectively protect the copyright of the digital audio data, it has been generally agreed[1–9] that a good watermarking scheme should satisfy the following properties:

1. The embedded watermark should not produce audible distortion to the sound quality of the original audio.
2. The computation required by watermark embedding and detection should be low. The complexity of watermark detection should be especially low to facilitate its integration into consumer electronic products.
3. Watermark detection should be done without referencing the original audio data. This property is known as blind detection.
4. The watermark should be undetectable without prior knowledge of the embedded watermark sequence. This property prevents attackers from reversing the embedding process to remove the watermark.
5. The embedded watermark should be robust against common signal processing attacks such as filtering, resampling and compression.
6. The watermark should survive malicious attacks such as random cropping and noise adding. However, severe attacks that produce annoying noise can be ignored for the survival test.

3. PREVIOUS WORK ON AUDIO WATERMARKING

A variety of audio watermarking methods with very different characteristics have been proposed. They will be reviewed in this section.

Early work on audio watermark embedding achieved inaudibility by placing watermark signals in perceptually insignificant regions. One popular choice was the higher frequency region,[10–12] where human sensitivity declines compared to its peak around 1 kHz. In some systems,[10,11] the watermark signal is high-pass filtered before being inserted into the original audio. In another system,[12] the Fourier transform magnitude coefficients over the frequency range from 2.4 kHz to 6.4 kHz are replaced with the watermark sequence. In these systems, inaudibility is further enhanced by only embedding watermarks in audio segments whose low frequency components have a higher energy value. The strong low frequency signals in the original audio could help to mask the embedded high frequency watermark signal.

Another human insensitive domain is the Fourier transform phase coefficients. Human ears are relatively insensitive to phase distortions, and especially lack the ability to perceive the absolute phase value. A scheme[1] proposed to substitute the phase of an initial audio segment with a reference phase that represents the watermark. The phase of subsequent segments is adjusted to preserve the relative phase between segments. In another system,[13] selected Fourier transform phase coefficients in higher frequencies are discarded and new values are assigned based on neighboring reference coefficients. The watermark is represented by the relative phase between selected coefficients and their neighbors. The problem with watermarking schemes that hide watermark signals in perceptually insignificant regions is that they are less robust to signal processing and malicious attacks. Compression algorithms do not preserve these regions well so that malicious hackers could implement stronger attacks in these regions without introducing annoying noise.

Another class of algorithms embed watermarks as echo signals of the original audio. The inaudibility of echo hiding is based on the theory that resonance is so common in our environment that human usually do not perceive it as noise. In these algorithms,[2,14] watermark signals are actually delayed and attenuated versions of the original

signal. The watermark sequence is represented by delay amounts which are retrieved by observing autocorrelation peaks in the time domain[14] or in the cepstrum domain.[2]

Recently, some researchers use a concept borrowed from spread spectrum communication and embed the watermark as pseudo-random noise in the time domain. It is guaranteed by spread spectrum theory that the embedded watermark is statistically undetectable by hackers. Since human ears have different sensitivity to additive noise in different frequency bands, all proposed work uses some filter to spectrally shape the pseudo-random (white) noise and achieve inaudibility. A simple band-pass filter was used in one work,[1] and a nonlinear filter was adopted in another.[4] In yet another system,[15] instead of filtering white noise, a scheme was developed to generate the band-limited pseudo-random watermark signal. The inaudibility of the embedded watermark could be further ensured by utilizing the masking effects of the human auditory system. One system[16,5] used MPEG-I Audio Psychoacoustic Model 1 to spectrally shape the watermark signal while another system[17] used the masking model from MPEG-II AAC. Watermark detection is done by calculating the correlation between the watermarked audio signal and the watermark signal. Armed with the spread spectrum communication theory, this type of watermarking usually survives pretty well under distortions and attacks. However, synchronization is difficult to implement, and its computational cost is high.

Another trend in digital audio watermarking is to combine watermark embedding with the compression or modulation process. The integration could minimize unfavorable mutual interference between watermarking and compression, especially preventing the watermark from being removed by compression. In one scheme,[18] watermark embedding is performed during vector quantization. The watermark is embedded by changing the selected code vector or changing the distortion weighting factor used in the searching process. The need of the original audio to extract the watermark greatly limits the applications of this scheme. Another algorithm[19] embeds watermark directly in the sigma delta modulation bitstream to eliminate the need of transforming it into PCM data, thereby keeping the computational cost low. This is important to the sigma delta modulation system, where hardware savings is the main goal. In another scheme,[6,20] watermarking is integrated with MPEG-II AAC compression. Watermark is embedded by modifying selected compression coefficients such as the scale factor.

4. PROPOSED ALGORITHM

Although the methods described in section 3 have their own features and properties, they share one common problem. That is, they are vulnerable to the synchronization attack in watermark detection. This problem could be resulted from casual audio editing such as cropping unwanted audio segments or intentional attacks such as randomly deleting or adding samples to watermarked audio data. This *random sample cropping attack* is very effective in interfering with the watermark detection process with respect to the algorithms mentioned above. This attack has a very low computational complexity. Besides, when done correctly, it would not introduce annoying noise to the underlying audio signals. One might argue that such a skillful attack could only be done by a few professionals and not by the majority of consumers. However, once a watermarking method is widely in use, it is almost certain that some professionals would produce and distribute attacking apparatuses so that a majority of common users would be able to perform the skillful attack. One method[5] was proposed to solve the synchronization problem, where an exhaustive search algorithm was used and the original audio signal was required. Consequently, its computational complexity is too high, and the need of original audio for watermark detection greatly limits its applications. Furthermore, it can only handle the casual editing attack, but not the random sample cropping attack.

In this research, we propose a low complexity solution to the synchronization problem caused by both casual and malicious attacks. The solution is composed of a salient point extraction technique and a Fourier transform domain watermark embedding procedure. Salient point extraction through audio content analysis is done during both watermark embedding and detection processes so that synchronization is regained at each salient point. The extraction algorithm is designed such that salient points remain stable after distortion. The Fourier transform domain watermark embedding and detection is adopted since the frequency domain information is less effected by sample cropping in the time domain.

One common characteristic among most existing audio watermarking algorithms is that their watermark is embedded throughout the entire audio signal. However, this may not be the most efficient way to embed and detect watermarks. For a skilled attacker, different amount of attack could be applied to different segments of the audio signal to avoid introducing annoying noise. For example, randomly cropping (deleting) one sample out of every 100 samples in high energy tonal segments of audio signals would produce noticeable noise, but the effect of doing so in

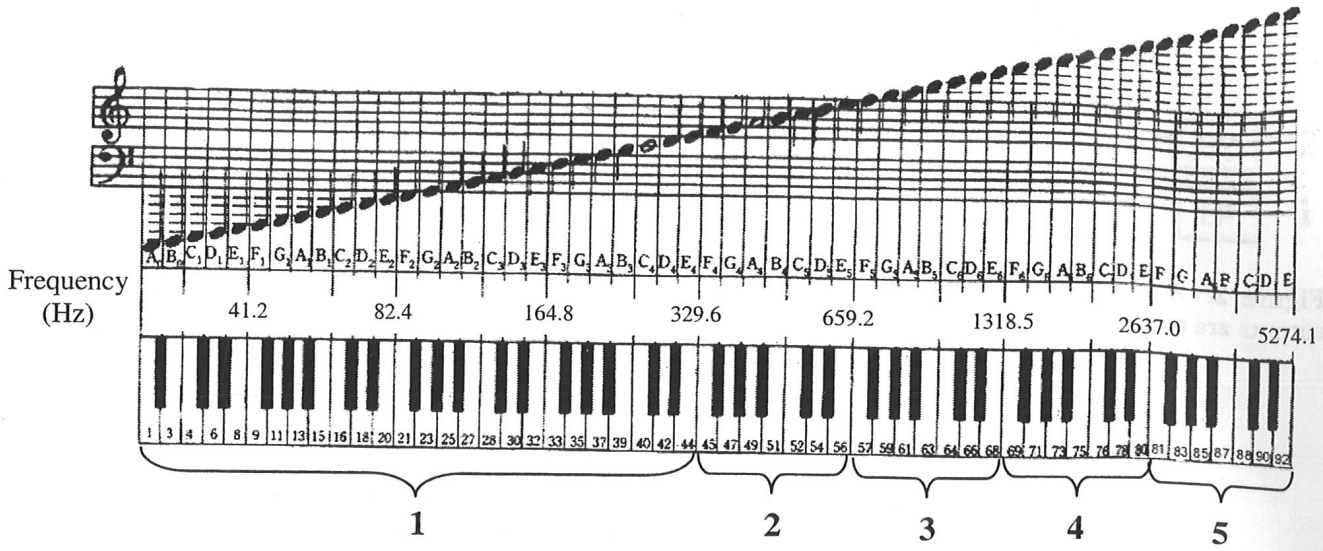


Figure 1. Illustration of the correspondence between music notes and frequency values, and the 5-subband partition adopted in this work

low energy segments would be inaudible. Thus, watermarks embedded in highly-attackable areas will face heavier attack and are more likely to be destroyed. The second major contribution of this work is the introduction of "attack-sensitive regions" via audio content analysis. If the watermark is only embedded in attack-sensitive regions where little attack could be applied, the computational complexity of both watermark embedding and detection could be reduced.

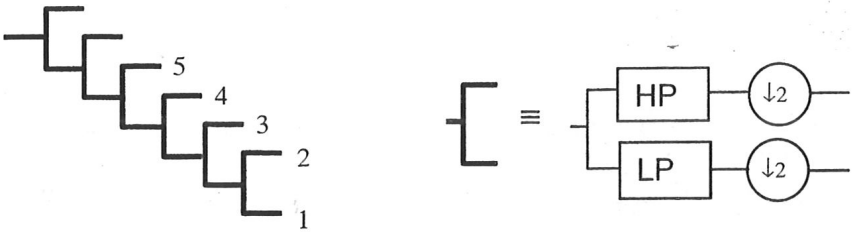
By combining techniques of salient point extraction, attack-sensitive region identification, and Fourier transform domain watermark embedding and detection, we propose a complete audio watermark embedding and detection system for copyright protection. This system satisfies all desired properties of watermark design described earlier. Furthermore, it has a very low computational complexity, and it is robust to casual and intentional synchronization attacks. Although we incorporate the concept of salient point extraction and attack-sensitive regions into our own watermark embedding method here, it is our belief that other watermark embedding algorithms will benefit from the same concepts as well.

4.1. Audio Content Analysis for Watermarking

In our system, audio content analysis is performed for the purposes of salient point extraction and attack-sensitive region identification. Salient points in an audio signal allow watermark detection to resynchronize at these locations. Synchronization by salient points has far less complexity than exhaustive search and makes blind watermark detection possible. It should be noted that we do not insert salient points, but extract them from the raw audio via content analysis. This approach has two advantages over explicitly embedding synchronization signals. One is that our content analysis approach does not introduce any distortion to the original audio signal since we do not add anything to it. The other is that the explicitly added synchronization signal is more likely to be taken out by attackers.

A good salient point extraction method should produce approximately the same set of salient points from audio signals before and after attacks such as audio compression, low-pass filtering and noise adding. To achieve this, we extract salient points based on audio features that are sensitive to human ears. In this way, if an attacker wants to destroy these salient points, he/she would have to alter these features and produce noticeable distortions. We choose the energy variation as the main feature for salient point extraction because the associated computational cost is low and alterations in this feature would be audible.

The basic scheme is to extract salient points as locations where the audio signal energy is climbing fast to a peak value. While this approach works well for simple music pieces with few instruments, it has two problems with more



(a) (b)

Figure 2. A 6-level dyadic wavelet decomposition, where each branch in (a) represents the structure in (b) and outputs are numbered corresponding to subbands in Fig.1

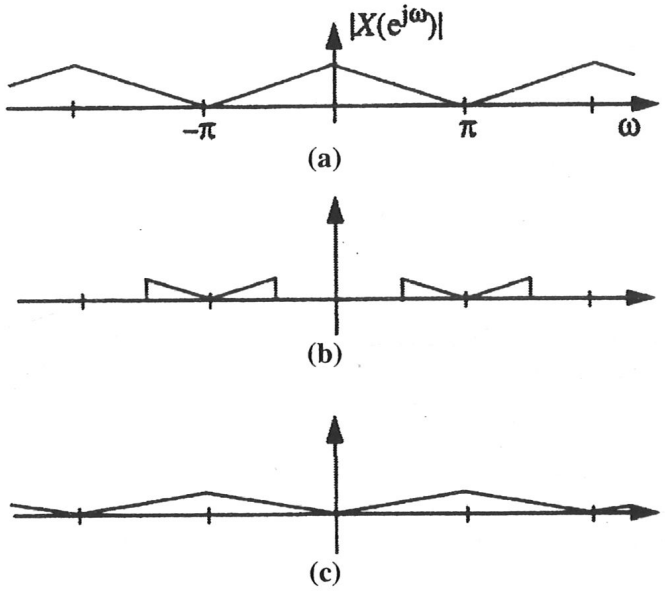


Figure 3. The effect of frequency inversion due to downsampling after high-pass filtering: (a) the spectrum before filtering, (b) the spectrum after high-pass filtering, (c) the spectrum after downsampling, where the highest frequency in (a) is now mapped to the lowest frequency.

complex music pieces. The first problem is that the overall energy variation becomes ambiguous for complicated music where many instruments are played together. Thus, the stability of salient points decreases. The other problem is that optimal threshold values are different for music pieces with different complexity. While a high threshold value is suitable for music with sharp energy variation, the application of the same value to complex music would yield very few salient points.

Therefore, it is beneficial to parse complex music into several simpler ones so that stability of salient points could be improved and the same threshold could be applied to all music pieces. Complex music is usually composed of instruments whose fundamental frequencies occupy different frequency ranges in order to form harmony. Figure 1 illustrates the correspondence between music notes and frequency values. It also shows the partition in our design, which consists of 5 frequency ranges. Note that the frequency width of each octave is not the same. The frequency intervals in Figure 1 correspond to outputs of a 6-level dyadic wavelet decomposition under a sampling rate of 44.1kHz as shown in Figure 2.

In order to prevent the frequency inversion effect[21] due to the application of downsampling to the output of high-pass filtering as shown in Figure 3, we modify the dyadic wavelet decomposition of Figure 2 into Figure 4 by

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.