Literature Survey on Audio Watermarking

Adam Brickman
EE381K - Multidimensional Signal Processing
March 24, 2003

Abstract

Watermarking audio files has recently become the focus of much attention. This is primarily due to faster data transmission rates on the Internet, which has allowed the often illegal proliferation of digital audio files. Watermarking may give recording companies the ability to enforce copyright protection of their products. The difficulties in watermarking audio lie in both the desire to preserve file quality and the need for the watermark to remain intact after a number of possibly damaging file operations. This article discusses the concepts involved in audio watermarking, applications, previously proposed algorithms and poses a new possibility for a watermarking scheme.

1. Introduction

The MPEG-1 level 3 (".mp3") format has become one of the standard audio formats in existence today. Many mp3 files are made from "ripped" CDs, and audio piracy has become a real problem for the audio recording industry. Audio watermarking involves embedding a sequence of data as additional information into an audio file. It has numerous applications, most of which have not yet been fully exploited. It can also be realized in numerous ways. This goal of this paper is to present an overview of the applications, challenges, and various algorithms associated with audio watermarking. In addition, it presents a widely overlooked potential application and proposes possible methods of realization.

2. Background and Applications

2.1 Basic Theory of Audio Watermarking

Watermarking alters an original image I with data in the watermark W in a particular manner such

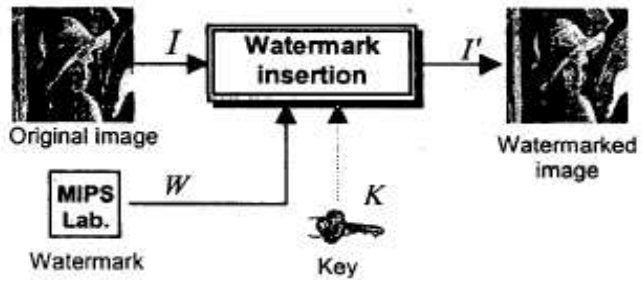

Figure 1 - Watermark Insertion [1]

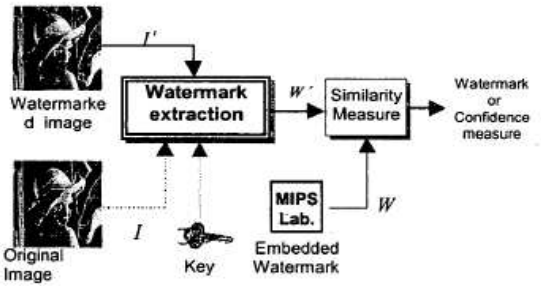

Figure 2 - Watermark extraction [1]

that the original image and watermark can be later recovered. From Fig. 1, we see that an original image is combined with a watermark and an optional key to produce the watermarked image. Fig. 2 depicts a "non-blind" form of watermark recovery, i.e. the original image is required in order to extract the embedded data. In the case of audio, we can conceptually replace the image I with an audio file A.

2.2 Measures of Evaluation

All audio watermarking schemes contain various parameters in common, in particular robustness, security, transparency, complexity, and capacity. Some of these parameters are mutually exclusive tradeoffs; that is, increasing the strength of one will decrease the strength of the other. Robustness describes the reliability of watermark detection after it has been through various signal processing operations [2]. Security reflects how difficult it is to remove a watermark. A scheme is truly secure if knowing the exact embedding algorithm does not help a user detect or extract the hidden data [3]. Transparency relates the human ability to hear the audio watermark. Usually, if not always, complete transparency (complete inaudibility) is desired. The complexity of an encoding scheme might be an important reason to choose one algorithm over another. For instance, a portable consumer device might not have the processing power to carry out an extremely complex scheme in a reasonable time or perhaps at all. Finally, capacity describes how many information bits can be reliably embedded.

There have been a few attempts to standardize watermark objectives and evaluation. The Recording Industry Association of America created the Secure Digital Music Initiative (SDMI) whose goals are to "develop open technology specifications for protected digital music distribution" [4].

However, as of May 18, 2001, the SDMI is "now on hiatus, and intends to re-assess technological advances at some later date." A limited number of benchmarks, such as the StirMark benchmark, have been created to allow registered users to test their watermarking schemes against a set of attacks (see below) and publish standardized, reliable results.

2.3 Audio Watermarking Attacks

Any operation that may decrease watermarking performance is called an "attack." [5] categorizes attacks into four main classes: removal, geometric, cryptographic, and protocol. Removal attacks remove the watermark without a necessary understanding of the watermarking scheme. Geometric attacks distort watermark detection through receiver desynchronization. Cryptographic attacks crack the watermarking scheme itself while protocol attacks exploit invertible watermarks to cause ownership ambiguity. [6] and [7] provide good examples of actual employed attacks.

2.4 Applications

Watermarking schemes are most commonly designed for copyright protection to resolve piracy disputes. These watermarks generally fall into two categories: proof of ownership and enforcement of usage policies [8a]. Proof of ownership watermarks may help determine rightful file ownership, perhaps as evidence in a court of law. Enforcement of usage watermarks could provide instructions or copyright information to consumer applications, which could refuse to duplicate or play music in violation of a usage policy.

"Fingerprint" watermarks provide information that allows one to track an audio clip's usage history. This could be especially useful to many record companies and advertisers as it could provide feedback about the popularity of a particular song or the number of times a commercial was played.

Audio watermarks could also be used to determine whether a file has been significantly altered. A number of operations can be performed on an audio file, some of which are most likely innocent (such as volume adjustment or equalization) whereas others are malicious and deliberate attempts to damage

watermark integrity. "Fragile" watermarks are designed to be easily broken when undesired operations are performed.

3. The Human Auditory System and MPEG Psychoacoustic Model

Inaudible watermarks are made possible by exploiting characteristics of the Human Auditory System (HAS). Audio watermarking is especially challenging as compared to image watermarking because the HAS is far more sensitive than the visual system. Perturbations in a sound file can be detected as low as one part in ten million [9]. Nonetheless, various "tricks" can be employed to cause inaudible distortion. The auditory system acts like a bandpass filterbank with strongly overlapping filters. The MPEG psychoacoustical model represents these filters as "critical bands," each with a particular sensitivity. If a signal is maintained below the threshold of sensitivity, then the watermark will be inaudible.

Above 2 kHz, the HAS focuses more on the temporal envelope of an audio signal than the actual structure [10][11]. Thus, small changes in the spectrum above 2 kHz are less likely to be noticed by a human listener. However, a major limitation to spread spectrum coding (discussed below) is that the MPEG model takes advantage of this and limits the upper frequency bit rate. Temporal masking involves two sounds that occur over a very short period of time (on the order of milliseconds). The ear essentially ignores the weaker sound even if it occurs before the stronger masking sound.

4. Spread Spectrum Audio Watermarking

The most popular method of audio watermarking employs spread spectrum techniques. This method is desirable for its statistically invisible properties and its strong robustness to noise. The basic idea is to spread the watermark data across the entire audible spectrum. A carrier signal is modulated with the watermark and a PN sequence with values of ±1 (called "chips"). The PN sequence is randomly

generated and appears as random noise in the frequency domain, hence the statistical invisibility. The resulting output sequence is attenuated and added back to the original audio signal, completing the embedding process. To detect the watermark, the audio signal is again modulated with the same chip sequence and sent through a bandpass filter centered around the carrier frequency. Detection is performed by a correlation detector or by statistical hypothesis testing [12]. The PN sequence generation and embedding schemes are of fairly low complexity. However, detection is significantly more involved.

A good schematic and implementation of the spread spectrum scheme is shown in Fig. 3. The watermark is multiplied by an m-sequence (a maximum length PN sequence), and the coefficients a(n) are determined by temporal analysis of the input audio sequence. The b(n) sequence allows the watermark to have additional power if the input sequence contains significant high frequencies, which are less noticeable by the HAS. Both a(n) and b(n) weight the watermark samples in order to boost the power of the watermark. Watermark extraction is achieved via segmenting the watermarked signal into blocks and measuring the cross-correlation with the m-sequence. A threshold decision block provides a majority opinion of three consecutive cross-correlation values to determine the value of the watermark bit [13]. Interestingly, frequency analysis is accomplished entirely in the time domain via counting of signal zero-crossings, which makes this algorithm extremely computationally efficient.

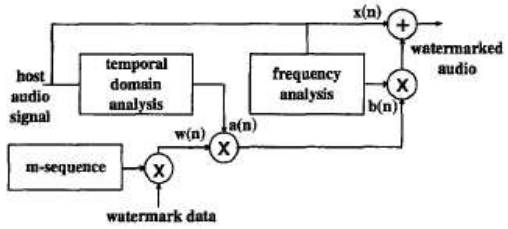

Figure 3 - A Realization of Spread Spectrum Encoding [13]

[12] uses a similar basic embedding approach. However, watermark inaudibility is achieved via noise shaping using a Hamming window rather than temporal exploitation of the HAS. Detection involves a correlation of the original watermark with the watermarked signal evaluated for all possible circular shifts of the watermarked signal. This provides robustness against synchronization attacks but at the cost of much greater computational complexity. The watermarking algorithm in [14] uses temporal

masking as in [12], but also uses the MPEG psychoacoustical frequency masking model. The authors claim that this watermarking technique embeds the maximum amount of information while remaining perceptually inaudible.

[15] increases robustness to detector desynchronization attacks (a major problem with spread spectrum coding) via temporal beat detection in the host audio file. Although repeated chip coding can help alleviate synchronization problems [16], it facilitates watermark estimation attacks. Realizing this, the authors instead employed beat detection as a means of synchronization, which delivered up to a fourfold reduction of code redundancy.

The problem of spread spectrum watermarking has been related to a communications channel problem. The optimal attack strategy is the solution of a particular rate-distortion problem, and the optimal hiding strategy is the solution to a channel coding problem [17][18]. The channel capacity is defined as the maximum mutual information between an input X (the watermarked data) and output Y (2):

$$C_{chan} = \max_{p(x)} I(X:Y) = \max[h(Y) - h(Y|X)] \qquad (2)$$

The maximum is taken over all possible distributions p(x), and the term h(X|Y) represents information loss due to channel noise, which is essentially due to the combination of the original audio and signal processing procedures.

5. Other Techniques

Audio watermarking algorithms have also been accomplished in the frequency domain. [10] presents a form of covert audio watermarking using phase modulation for proof of ownership applications. The key to maintaining phase shift inaudibility is to keep the absolute phase shift small. After segmenting the original audio signal using overlapping windows and taking the FFT, the watermark is inserted into every other segment, which is windowed and added to the non-watermarked blocks in an

overlap fashion. The phase modulation itself is obtained by a phase window function (3). Each bit of the watermark a_i to be embedded in a particular segment k is multiplied by shifted versions of this phase window function, and then summed together to perform the overlapping (4). This is performed for all the audio segments.

$$\phi(b) = \sin^2(\pi(b+1)/2) \qquad (3)$$

$$\phi_k(b) = \sum_{i=1}^{I} a_i \phi(b-i), 0 \le b \le I \qquad (4)$$

Reversal of this procedure produces a very noisy retrieved phase. It is modeled by a hidden Markov model. Instead of making a decision for each message bit, [10] proposes to consider the entire observation sequence to determine the best result using the Viterbi search algorithm. For N blocks of M samples each, encoding complexity is approximately N•O(M•log(M)). Again, we find that decoding complexity is significantly higher.

"Echo hiding" relies heavily on the temporal limitations of the HAS (as described in Section III). It embeds binary data by echoing the host audio with one of two delay kernels which have different offsets in time [19] (Fig. 4). Additional parameters include initial kernel amplitude and the decay rate. As the offset (delay) between the original and the echo decreases, the two signals become indistinguishable to the human ear.

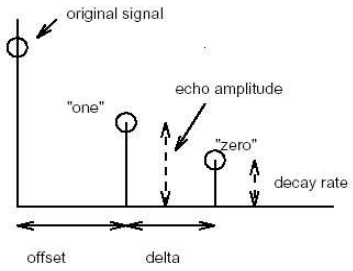

Figure 4 - Echo Hiding System Kernels [19]

Decoding commonly involves detection of the echo spacing by using the distance between two pronounced autocepstrum peaks. Rather than using a single system kernel as mentioned above, [20] uses multiple echoes to achieve high encoded bit rates. Decoding was achieved with a half-blind scheme in which the positions of the watermarked segments are transmitted to the decoder. This helps lower algorithm complexity. The number of computed correlation peaks determines the watermark bit values.

[21] applies negative echoes to interesting effect (the resultant sound was described as “sharper”).

Briefly, another watermarking method involves changing the least significant bit of audio data. This method is seldom used, however, because of its almost total lack of robustness to additive noise and HAS sensitivity. This method does otherwise yield a high bit capacity with low complexity.

6. Conclusion and Project Plans

Spread spectrum watermarking is robust to noise attacks, and detection does not require the original audio. However, the detector is susceptible to desynchronization errors. Echo hiding is another very successful and popular method of watermarking. Although easily detectable, it is very robust to many attacks, including MPEG compression. All audio watermarking takes advantage of the limitations of the HAS.

One could strongly argue that the premise of watermarking for copyright protection/detection is fundamentally flawed. In an environment where security is either fully intact or fully breached, it is probably not possible to design a perfect algorithm that can “do it all.” Somebody, somewhere will figure out a way to defeat the watermark. When a user buys a CD from a music store, are they going to be required to disclose their personal information? What would happen to the used CD market?

Instead, watermarks could be used to supplement and enrich digital audio. I propose to create a sufficiently high capacity watermarking scheme to be applied toward embedding text into an mp3 file. A potential application is the ability to embed song lyrics into a file.

My project has several advantages over previously proposed realizations. Because stereo files have two channels, I can effectively double the embedding bit rate and total bit capacity. It may also be possible to exploit redundancy of the stereo signal to provide additional watermark robustness. There is little need for statistical invisibility or defense against malicious attacks because there is little incentive for such attacks. Instead, the focus will be robustness against mp3 compression and signal processing

operations and a high data capacity. It seems desirable to use echo embedding as my watermark technique mainly because spread spectrum watermarking competes with the MPEG psychoacoustical model and appears overused in current literature. It is also attractive for its adjustable parameters. The results and complexity of my algorithm will be provided in my final report.

REFERENCES

[1] S. Lee and S. Jung, "A Survey of Watermarking Techniques Applied to Multimedia," *IEEE International Symposium on Industrial Electronics*, vol. 1, pp. 272-277, 2001.
[2] J. Dittman *et al.*, "Media-independent Watermarking Classification and the Need for Combining Digital Video and Audio Watermarking for Media Authentication," *International Conference on Information Technology: Coding and Computing*, pp. 62-67, 2000.
[3] M.D. Swanson *et al.*, "Current State of the Art, Challenges and Future Directions for Audio Watermarking," *IEEE International Conference on Multimedia Computing and Systems*, vol. 1, pp. 19-24, July1999.
[4] Secure Digital Music Initiative, "SDMI - Home," http://www.sdmi.org/, Sep. 2002.
[5] S. Voloshynovskiy *et al.*, "Attacks on Digital Watermarks: Classification, Estimation-Based Attacks, and Benchmarks," *IEEE Communications Magazine*, vol. 39(8), pp. 118-126, Aug. 2001.
[6] F.A.P. Petitcolas *et al.,* "StirMark Benchmark: Audio Watermarking Attacks," *Information Technology: Coding and Computing*, pp. 49-54, Apr. 2001.
[7] M. Wu *et al.*, "Analysis of Attacks on SDMI Audio Watermarks," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1369-1372, 2001.
[8] S.A. Craver *et al*., "What can we reasonably expect from watermarks?" *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp. 223-226, 2001.
[9] C. Xu *et al.*, "Digital Audio Watermarking and its Applications in Multimedia Database," *Proceedings of the Fifth International Symposium on Signal Processing and Its Applications*, vol. 1, pp. 91-94, 1999.
[10] S. Kuo *et al*., "Covert Audio Watermarking Using Perceptually Tuned Signal Independent Multiband Phase Modulation," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1753-1756, May 2002.
[11] P. Noll, "MPEG Digital Audio Coding," *IEEE Signal Processing Magazine*, vol. 14(5), pp. 59-81, Sep. 1997.
[12] P. Bassia *et al.*, "Robust Audio Watermarking in the Time Domain," *IEEE Transactions on Multimedia*, vol. 3(2), pp. 232-241, June 2001.
[13] N. Cvejic *et al.*, "Audio Watermarking Using m-Sequences and Temporal Masking," *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp. 227-230, 2001.
[14] L. Boney *et al.*, "Digital Watermarks for Audio Signals," *Proceedings of the Third IEEE International Conference on Multimedia Computing and Systems*, pp. 473-480, June 1996.
[15] D. Kirovski and H. Attias, "Audio Watermark Robustness to Desynchronization via Beat Detection,"*Information Hiding Workshop, 2002.*
[16] D. Kirovski and H. Malvar, "Robust Spread Spectrum Audio Watermarking," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1345-8, 2001.
[17] P. Moulin and J.A. O'Sullivan, "Information-Theoretic Analysis of Watermarking," *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 3630-3633, 2000.
[18] A. Ambroze *et al.*, "Turbo Code Protection of Video Watermark Channel," *IEE Proceedings on Vision, Image and Signal Processing*, vol. 148(1), pp. 54-58, 2001.
[19] D. Gruhl *et al.*, "Echo Hiding," *Info Hiding 96*, pp. 295-315, 1996.
[20] S. Foo *et al.*, "An Adaptive Audio Watermarking System," *Proceedings of Electrical and Electronic Technology*, vol. 2, pp. 509-513, 2001.
[21] H.O. Oh *et al.*, "New Echo Embedding Technique for Robust and Imperceptible Audio Watermarking," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1341-1344, 2001.