

Audio Watermarking and Applications

Eric Metois, Ph.D.

ARIS Technologies, Inc. – September 1999

This discussion will attempt to outline some truths and common misconceptions about digital audio watermarking. It will survey the intrinsic obstacles that such technology is required to overcome, shedding light on its performance criteria, compromises and limitations. While doing so, it will also survey a few common types of applications, hopefully leading to a clear understanding as to the appropriateness of such technology and its expertise within multimedia content protection.

Introduction

Today's digital media have opened the door to an information marketplace where the true value of the product (digital content) is dissociated from any particular physical medium. While it enables a greater degree of flexibility in its distribution and a lower cost, the commerce of disembodied information raises serious copyright issues. Indeed, digital data can be duplicated and re-distributed at virtually no cost, potentially turning piracy into a simple "click and drag" process.

Cryptography has been clearly established as a technology of fundamental importance for securing digital transfers of data over unsecured channels. By providing for encryption and authentication of digital data, cryptography enables trustworthy point-to-point information exchange and transactions to be achieved. Yet, once the recipient validates and decrypts the data, the product can be subsequently stripped from any content identification, proof-of-ownership or other descriptive information, and any further duplication and re-distribution can leave the rights holders powerless and royalty-less. While such re-distributions may not represent a serious threat when the content consists of proprietary information that has a short life span, such piracy could have catastrophic implications for the entertainment industry, whose content has a very long life span.

Following the same line of thought, it is easy to see how digital sub-codes and other proprietary digital formats can fail in similar ways, since they are only volatile representations of a medium. The true value of the product (the content) can still be transferred effortlessly onto different formats and media. Any attempt to

secure the identities of a content's rights holders calls for a technology that enables some secure auxiliary information, or watermark, to travel with the content through any channel, format or medium where the content's value remains. A properly designed audio watermarking technology provides the means to do this in the context of audio content, while preserving the integrity of the original recording.

Watermarking Audio Content

Process

Unlike sub-codes, encryption or audio compression, an audio watermark should not rely on any specific format. In order to travel along with the content it protects, the watermark must be carried by the content itself. Embedding an audio watermark is an active modification of the audio waveform. Subsequent to this process, the watermarked audio content becomes a message carrier regardless of the format or medium it lives on.

In the context of standard audio processing and broadcast systems, the audio content may undergo various stages of band limiting. An audio watermark is expected to persist through such manipulations and therefore, it is imperative for the watermarking technology not to rely solely on portions of the audio spectrum that are perceptually less relevant. Of course, this should be done in a way that doesn't degrade the value of the content and the process of embedding a watermark should leave no perceivable audio artifact. In the end, an audio watermarking technology must face the ambitious challenge of opening an inaudible and reliable data channel within the most relevant part of the audio band.

Regardless of the actual approach that is chosen the disturbances that are introduced within the audio content at the time of watermarking will have to remain very small. The academic approach to this problem has typically been to consider such disturbances to be the message carriers (the watermark), while the original audio content would stand for a noisy contamination of this message. Set from this perspective, the previous challenge resembles the more familiar communication problem of opening a reliable data channel within an extremely low signal to noise ratio (SNR) environment. This view naturally points towards the concept of spreading the message across an appropriate domain (time, frequency, etc.). The standard low SNR communication problem also invokes the channel capacity theorem, which derives a theoretical maximum for the data capacity of a channel from its bandwidth and the expected SNR.

This academic approach may point the designer towards interesting choices and it certainly helps to develop an understanding of some compromises between transparency, robustness and data rate. However, any theory is only as good as its premises and the audio content that undergoes the watermarking process is hardly a noisy contamination of the communication channel. The very first realization is that over any reasonable time scale, the audio content will very rarely appear to be white. Most importantly however, the audio content is not the outcome of a random process that is added to the channel; it is perfectly known prior to the watermarking stage.

Performance Requirements

Transparency

Transparency is typically the first and foremost requirement from the perspective of content owners. Too much time, effort and money is spent on tedious production and mastering stages to compromise audio quality through a poorly designed audio watermarking system. In order to achieve such high audio quality standards most watermarking system will contain a psychoacoustic analysis stage of some sort. Psychoacoustic models are most popularly referred to in the context of perceptual audio compression. They aim to quantify the perceptual sensitivity of various elements of an audio. It is important to note that while some of

these models have gained a wide degree of acceptance, their expertise is often tied to the specific type of manipulation that follows the analysis. The psychoacoustic model that is used within an audio watermarking technique is designed to quantify the perceptual sensitivity of the material with respect to the kind of processing that is required in order to embed the watermark. The design and tuning of such models often requires iterative passes through audio quality tests.

Unfortunately, testing for the audio quality of a process is a difficult and tedious task. Such evaluations must be primarily conducted with recognized recording industry professionals in their own facilities. The tests should employ master-quality audio recordings with which the listeners are familiar, as well as with recordings specifically selected by industry professionals (e.g. EBU, the MPEG audio group) to expose potential flaws in audio encoding processes. In order to gain an objective reading of such a subjective attribute, listening tests should follow rigorous guidelines such as the widely accepted A/B/X double-blind matching procedure or the ITU-R BS 1116 recommended double-blind, hidden-reference procedure. The purpose of these tests is to determine listeners' ability to distinguish original material from watermarked material with an appropriate statistical significance.

Persistence

An absolutely transparent audio watermarking system offers very little value if it is highly volatile and does not persist through a variety of audio manipulations that the watermarked material will undergo. There are two major types of persistence that need to be addressed while designing the watermarking process.

A wide variety of legitimate audio processing stages will be applied to the content during its distribution and the watermarks' persistence through these is usually referred to in terms of robustness or survivability. Such processing stages include lossy data rate compression, band limiting, additive noise, spectral equalization, D/A-A/D and sample-rate conversion, AM/FM/TV broadcast, matrix (surround) encoding, non-linear distortion, dynamic-range, group delay, linear speed change, pitch invariant time scaling, wow and flutter, echo, artificial reverberation, and multi-channel down mixing.

The second type of audio processing stages includes manipulations that may be designed explicitly to remove or obscure audio watermarks while preserving the high audio quality of the content that carries them. The difficulty of designing such removal stages is referred to as the tamper resistance of the watermarking technology.

In the context of either types of processing, the persistence of the audio watermarks should always be evaluated in relation with the amount of audio quality degradation they incur. Keeping in mind the fact that the content is the watermark carrier, it makes no doubt that a sufficient degradation of this content will eventually succeed in removing the watermarks. While nothing can insure that the watermark will survive through any arbitrary manipulation of the content, the watermark is expected to survive processing stages that don't degrade the audio content beyond its commercial potential.

Components and Features

Watermark Data

The Watermark Data typically refers to the information, or message, one wishes to embed in an audio stream. While an audio watermarking technology doesn't necessarily require a rigid structure for these messages, we found that in the context of most applications, their length is both fixed and known in advance. The nature of this data is obviously tied to the application. Watermarks can simply carry unique identifiers, which may be linked to large amounts of descriptive information through central databases, or the message they carry can be self-contained.

Raw and Effective Data Rate

As for most communication systems, an audio watermarking system will typically dedicate part of its raw data rate to various error protection and correction techniques. The amount and the nature of the chosen error correction scheme are obviously functions of the watermark's purpose. The choice is typically derived as a compromise between the desired Watermark Data Length, the desired degree of error protection, and the typical environment the watermark is expected to survive.

In any case, one should not expect effective data rates that are much larger than 50 bits per second or so. This should not come as a surprise, recalling the similarities between a transparent audio watermark and an extremely low SNR communication channel.

Watermark Data Length

In the light of the typical data rate one can hope for, it is obviously in the designer's advantage to define the shortest possible Watermark Data that answers his needs. A shorter watermark will be repeated more often within the audio content and it will therefore be available within smaller chunks of the watermarked material. Also, a watermark extractor will have a better chance at recovering a shorter (and more frequently repeated) watermark from heavily contaminated versions of the audio material by exploiting the redundancy of the watermark channel.

Fortunately, most applications do not require large Watermark Data to be embedded within audio. As an illustration, a standard ISRC code, which consists of a country code, an owner code, the year of recording and a serial number, can fit within 60 bits of data and yet identify uniquely each track of all distributed music album. Let's suppose now that one wishes for a watermark that will identify uniquely each transaction of individual songs, then a 100-bit-long watermark could handle over 1 trillion transactions (almost 200 times today's world population). From this figure, there doesn't seem to be any reason for a watermark to exceed 100 bits or so in length.

Watermark Layer

It is often very desirable for the watermarking technology to allow the same content to carry multiple watermarks simultaneously; each one living on a different Watermark Layer as to ensure they don't interfere. Multiple layers may be chosen to carry the same or totally independent Watermark Data. Such Watermark Layers may also be embedded within the same content at various stages of the distribution process if necessary.

Even within the context of a single watermarking core technology, the properties of the watermark data channel are typically controlled by a large number of configuration parameters. The optimal configuration will be a direct function of the watermarks' purpose and in some case, a combination of configurations, leading to

multiple Watermark Layers, are the best way to cover the specific requirements of a given application.

Public and Private Layers

In addition to the nature of the Watermark Data, one can further draw distinctions between various purposes of watermarking any material. One of these distinctions may be whether the watermarking entity wants the content of the watermark to be publicly accessible or not. One concern may be that an undetectable watermark will tend to be less prone to tampering than a publicly readable watermark. On the other hand, if the watermarking entity is the only party that can detect and retrieve its watermark, then it becomes solely responsible for regulating any violation of its propriety.

A Private Watermark Layer requires an appropriate secret key on the decoder side. The idea is not limited to the scrambling the Watermark Data - the presence of the watermark would still be detectable - but rather to hide the Watermark itself within the protected audio material.

Applications

An audio watermarking technology offers means to open digital communication channels within audio content. Given this technology, the question remains as to the nature of the data that is carried by the watermarks. As we recall, the typically limited data rate of such channel suggests that the transmitted data be the shortest possible. This leads towards rather rigid data structures that will be a direct function of the watermarks' purpose. Additionally, we recall that the choice of a (set of) Watermark Layer(s) will typically be driven by the application's expected environment.

Content Identification

Many concerns ranging from proof of ownership and copyrights to usage monitoring and royalty tracking call for Watermark Data that identifies the content that carries it. Such a watermark would typically be embedded within the content at the last stage of production, insuring that all further copies are marked with the same identification tag. Whether these tags are based on codes that are currently in use (IWRC, ISWC) or not, they should be registered and

administrated appropriately in order to avoid risks of collusion and to identify clearly who the proper rights holders are.

Monitoring the usage of a song (either for royalty purposes or others) will also typically rely on the deployment of monitoring stations. These stations should be able to extract such content identification tags from all watermarked content, regardless of their sources. Therefore, the appropriate Watermark Layer(s) should not be private.

Serial or Transactional Identification

Audio watermarks can also serve to trace a specific chain of distribution. For instance, a content distributor may chose to serialize copies of the same content in order to identify the source (i.e. the distributor) and/or the recipient of a specific copy through a transaction identifier as Watermark Data. While such a watermark may not actively prevent the recipient from making pirated copies, it will serve as a great forensic tool that can trace all subsequent pirated copies back to the original recipient. Such serial or transactional Watermark Data will serve as a piracy deterrent.

There is no incentive for all content distributors to agree on a unique structure for these watermarks. In fact, many distributors may wish to retain complete control over this watermarking stage, which will take place within their distribution scheme. Private Watermark Layers are especially well suited for this kind of application.

Usage Control

When the deterring power of a forensic tool is not sufficient, audio watermarking can also serve within more active security solutions. The idea is simply to use the watermarks' persistent data channel as a means to transmit usage rules to compliant components and devices. These compliant components may include players, recorders and any other tool that is designed to manage or manipulate audio content.

Usage Control Watermarks should be extracted by a potentially large number of compliant components. As a result, the Layers that carry them cannot be private. The embedding of such watermarks can take place at various stages within the chain of distribution. The appropriate

time and circumstances for this embedding stage will be functions of the policies that are enforced within the embedded usage rules.

Others

Finally, the three previous types of watermarks do not bind the extent of audio watermarking applications. Many other remain, ranging from consumer awareness and marketing to toy applications and ratings.

Conclusion

Audio watermarking is a persistent data communication channel within an audio stream. It should survive through various format changes and manipulations (either legitimate or not) of the audio material, as long as the content retains some commercial potential. Additionally, it should do so without introducing any perceivable audio artifacts.

Its naturally limited data rate suggests that this communication channel should be used in conjunction with carefully designed Watermark Data. This data is preferably small, as shorter watermarks will be repeated more frequently throughout the content, enhancing the channel's redundancy.

The purpose of a watermarking stage will typically dictate the nature of this Watermark Data while the appropriate Watermark Layer will typically take into account the environment of the application and its associated threats.