Digital Music Distribution and Audio Watermarking

Steve Czerwinski, Richard Fromm, Todd Hodes
Computer Science Division
University of California, Berkeley
{czerwin,rfromm,hodes}@cs.berkeley.edu

Abstract

Digital music distribution is a reality. We discuss the policy-level concerns of artists, consumers, and the recording industry, and describe the technical specifics of watermarking, which is either a possible approach to balancing these concerns or a failure waiting to happen – depending on whom you talk to.

1 Background

It is becoming apparent that the digital distribution of music is inevitable. Consumers clearly want it - just witness the huge proliferation of MP3 audio files over the Internet in recent years. As was noted in a recent article on the front page of the Washington Post, "millions of Americans are making a daily habit of an emerging Internet music technology that is threatening to upset the entire structure of the popular music business" [33]. The music industry, however, is greatly concerned by these developments. While they recognize certain potential advantages to digital distribution of music, they want to protect their intellectual property. Any system ultimately adopted needs to balance the concerns of the industry, artists, and consumers [25].

At a high level, there are two distinct approaches for dealing with digitally distributing music in a manner that provides for intellectual property protection: copy protection and copy detection. Copy protection involves using some technical means to prevent a user from making a copy of an audio file, unless she is explicitly authorized to do so. Copy detection attacks the problem from a different angle. Instead of using technical means to prevent unauthorized duplication, consumers are expected (and implicitly trusted) to follow the rules and not break the law. Owners of intellectual property then can use technical means to hunt out unauthorized copies of their property which are being illegally distributed.

It can be plausibly argued that the copy protection approach, applied to digital music, is doomed to fail. This is especially true if the music playing client is a general purpose PC. A PC is a general purpose device, and a sophisticated user can use it to do virtually anything she wants. Even if the music is encrypted, at *some* point the music has to be in the clear, and is therefore susceptible to unauthorized duplication. For instance, Liquid Audio and a2b are two different proprietary, secure systems for distribution of digital music. And they have both already been cracked once, by a program called a2b2wav [38]. This hack added a menu item to each of the player applications which allowed the user to save the current music track as a .wav file, a standard, unencrypted audio format [35] from which an unlimited number of copies could then be made and distributed for use with no restrictions. Even if the client is an embedded device, a user can always use a D/A to A/D path (playing the music and re-recording it) to get around any copy protection.

Arguments can also be made that copy protection isn't worth the trouble. It's a hassle for consumers, and it can get in the way of legitimate uses. For instance, during the 1980's, computer software was routinely distributed with increasingly elaborate copy protection schemes. This frustrated users and made it difficult to make backup copies. The software industry eventually gave in to consumer demands and for the most part stopped copy protecting software. Also, there is an economic theory that shared information goods amongst small groups of consumers can actually lead to *increased* profits for manufacturers [5]. Part of the rationale is that through small scale piracy, where friends share music with each other, people get exposed to music that they wouldn't otherwise experience. They are therefore more likely to buy music that they wouldn't have otherwise considered. Furthermore, if people purchase music with the expectation that they can share it with their friends, they are willing to pay more for that music than if a system imposed technical constraints upon them such that it was impossible for anybody else to listen to that music.

If we believe that copy protection is doomed to fail and/or not worth the trouble, then copy detection becomes an attractive alternative for protecting the intellectual property rights of digitally distributed music. Copy detection is unlikely to do anything about small scale piracy. If someone makes an unauthorized copy of an album and gives it to a friend, it is unlikely that anybody will ever find out about this (even if it is technically possible). But it can be argued that this is not a significant problem, and what the owners of intellectual property are really concerned about is one person purchasing an album, then making it available to everybody else in the world by putting it on the Internet. Their fear is that consumers would not bother paying for the album when they could instead download it for free. In this situation, the music is available for access by the owner of the intellectual property. If the owner can detect that his content is being made available in an unauthorized manner, she can use the existing legal system to combat this large scale piracy. It is this type of scenario for which copy detection is ideal.

Copy detection may be either content-based, or it can be accomplished through watermarking. FRAUD is an example of a content-based copy detection system. Psychoacoustic features, such as harmonicity and brightness, are extracted from a piece of music to form a signature. An automated program can then search the Internet, on behalf of a copyright owner, searching for music with matching signatures [41]. Another way of achieving copy detection is through watermarking. A watermark is a bitstream hidden in the digital audio; the watermark might contain copyright information. An automated program can again search the Internet, this time looking for music whose watermark matches that of the copyright owner.

2 Current Efforts

Before examining watermarking as a means to implement a secure system of digital music distribution, it is worthwhile to look at what are some existing and proposed systems to accomplish the same goal.

There are several efforts in the development of standards. A consortium of recording industry, computing, and electronics companies, led by the Recording Industry Association of America (RIAA), have formed the Secure Digital Music Initiative (SDMI). SDMI is planned to be an open architecture and specification for secure digital music distribution. The stated goals of SDMI are to mark music with rights management data in such a way that the data is permanently tied to music and is recognized and acted upon by all devices [40]. The group may be overly ambitious in their timetable. Reacting to the exploding popularity of MP3, and the introduction of portable consumer electronics devices that play (unprotected) MP3 files [29], the SDMI consortium would like to have SDMI-compliant devices available to consumers for the 1999 Christmas shopping season. To do so would require getting specs to consumer electronics manufacturers by June 30 [45]. Given the lack of progress by the consortium so far, meeting this deadline seems unlikely. And some labels may not be willing to wait; Universal Music, the largest record company, is planning to distribute music over the Internet by the end of 1999 with or without SDMI [37].

In contrast to SDMI, MPEG-4 is a standard whose inital version has already been agreed upon (for the most part) and released. The standard includes a system for Intellectual Property Management and Protection (IPMP), which supplements regular media with optional Intellectual Property Identification (IPI) information. The details of such information as well as its implementation are domain- and application-specific. The MPEG-4 standard simply specifies the IPMP interface [22].

There are a number of proprietary products which fall under the category of digital rights management (DRM) systems. DRM systems secure digital content and manage the use of the secured content in accordance with the rights and interests of all parties [23]. Several existing digital rights management systems include Intertrust [23], the Madison Project (from IBM) [1], Cryptolope (also from IBM) [20], and MagicGate and OpenMG (from Sony) [42]. Intertrust has announced plans to be incorporated into Microsoft's upcoming MS Audio 4.0. It will also be utilized to protect MP3 files in future versions of the Rio, a portable MP3 player from Diamond Multimedia. A partnership between the Universal Music Group and BMG for MP3 distribution plans to use it as well. The Madison Project will be used by RealNetworks for RealAudio files [1].

Whereas digital rights management systems are independent of file format, an alternative option to secure music digital distribution is to define a proprietary file format that is designed to be secure [24]. Examples of such formats include Liquid Audio [26], a2b (from AT&T) [2], and ASFS (the AudioSoft file structure) [4]. The a2b system works as follows. A music file is first compressed, using a proprietary algorithm, and then encrypted. Each song has a unique key. A customer listens to songs with a freely available player program. When a song is purchased, the key for the song is transfered to the customer's player. Other proprietary systems incorporate similar features - some method of encryption, and decryption on a proprietary player. There are a few disadvantages to this approach. If the playback of a purchased song is tied to a single playback client, the user may have less control than she has with an ordinary CD (for instance, the ability to listen to the music both at home and work). Also, even though some of the proprietary music formats may be "better" than MP3 (for instance, AT&T claims that a2b achieves greater compression and superior audio quality than MP3, and that this has been backed by independent tests [2]), a variety of incompatible formats which are not interoperable could be difficult to use, if they each require a different player.

It is also possible to design such a secure system around a standard audio file format [24] [13]. A number of companies have secure systems built upon encrypted MP3 files, including Mjuice (from Audio Explosion) [28] and M-Trax (from MCY) [27].

It should be noted that the distinction made earlier between copy prevention and copy detection is somewhat blurred. Most of the systems listed above are mostly concerned with copy prevention, but it is possible to combine copy prevention and copy detection into a single system, and some of the systems do include watermarking. For instance, Liquid Audio, along with a group of record labels, artists, and MP3 product vendors, has formed the Genuine Music Coalition. They are developing a standard for watermarking technology, known as the Genuine Music Mark, which will be used for both Liquid Audio and MP3 files [12].

3 Watermarking Overview

3.1 Introduction

Copy detection systems, by definition, must examine all audio files published on the web to determine if they are fraudalent copies of music they are protecting. Examining these audio files is a daunting task, but even more so if the operation to detect copies is very time consuming. For example, a simplistic copy detection system could correlate each second of audio it is examining against all the digital audio it is protecting. This provides near perfect copy detection, at the cost of unrealistic CPU time for each comparision. Audio watermarking attempts to improve the performance of copy detection systems by allowing a more rapid determination of whether or not a given audio file is part of copyrighted song.

In the most general form, audio watermarking hides a user-specified bitstream in digital audio such that the addition of the bitstream (the watermark) is perceptually insignificant. Figure 1 shows the block diagrams for both embedding a watermark into an audio stream and how to recover an existing watermark. As this figure shows, to watermark digital audio, the original audio and the bit stream are inputted into the encoder along with a secret key. This secret key is known only to the watermarker; it is used to hide the bitstream so that no other party can locate the watermark in the digital audio. To recover the watermark, this secret key is used along with the watermarked audio. In some cases, recovery requires the original unwatermarked audio.

Once a watermark can be embedded in digital audio, it can be used in various ways. The most common use will be to put copyright ownership information into the bitstream. This way, a copyright detection system can quickly recover who owns a given piece of audio. A more elaborate copyright system could also encode the entire ownership history for the digital audio file, thus allowing for tracking. There are other uses besides copyright enforcement, such as annotating music by placing the artist and song's name into the digital audio, so that listeners can easily recover them. Finally, digital audio watermarking is also useful as a general steganography tool.

3.2 Watermarking Design Goals

Audio watermarking systems are design to meet three goals: maximizing the difficulty of removing the watermark without destroying the audio, minimizing the perceptual effect of the watermark, and maximizing the information which can be encoded per second of original audio. Designing a system that optimizes all of these goals is difficult because, for the most part, each of the goals work against the others.

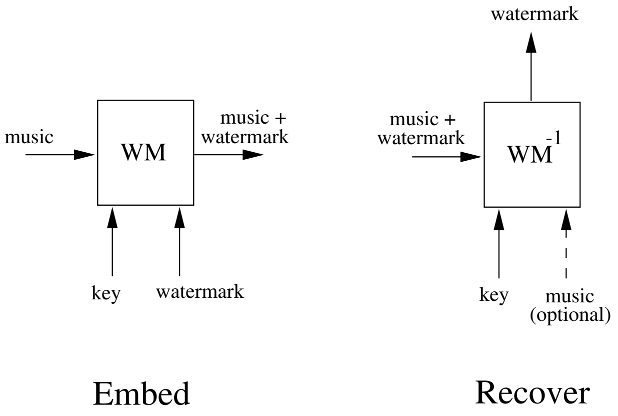

Figure 1: Watermarking block diagram

Attack	*Examples*
Digital Filtering	EQ, Time scaling, pitch-shifting, cropping, mixing, editing, resampling, add colored noise,...
Lossy encoding	MP3 (or anything based on psychoacoustic knowledge)
D/A to A/D	Audio-out to audio-in; play to speakers, record with mic
Key recovery	Correlate multiple watermarked streams

Table 1: Watermarking Attacks

In order to make a watermark difficult to remove without significantly degrading the audio, the bitstream must be placed in the perceptually "important" parts of the audio, or in other words, the portions of the audio that most affect human hearing. If it was placed in parts that did not affect human hearing, then the watermark could be removed by simply transforming the audio with a perceptual encoder that removes the portions that are inaudible.

However, placing the bitstream in the "important" parts goes against the goal of reducing the perceptual effect of the watermark. These "important" parts are the exact portions that affect human hearing, so modifying them leads to audible effects. To minimize these effects, the bitstream should be placed placed in less important places, but this of course, goes against the first goal.

The amount of information you can encode in the bitstream also works against the other goals. First, increasing the amount of information encoded makes it more likely listeners will be perceptually affected by watermark. Similarly, increasing the amount of information encoded also hampers the system's ability to hide the bitstream just by the mere fact that there is more to hide.

Watermarking systems attempt to balance these goals, so that each one is sufficiently fulfilled, while not hampering the others significantly.

3.3 Attacks

Probably the most difficult task of audio watermarking schemes is hiding the watermark such that it cannot be easily removed. This is because there are many diverse attacks that can be used to reduce the detectability of the watermark.

Table 1 lists the most common attacks used against watermarking schemes. The first three attacks are used to remove the watermark, while the fourth attack is used to recover the secret key used in the watermarking.

As Table 1 shows, the types of attacks are diverse. The first attack, described as digital filtering, actually covers many attacks that can be described by simple mathematical filters, such as time-scaling, pitch-shifting, cropping, mixing sources, and editing. To be robust against these attacks, the watermark must be redundantly placed in the audio, along with being resistant to the various transforms.

The lossy encoding attack is a little more sophiscated and harder to protect against. The idea behind lossy encoders, such as MP3, is to throw away any audio that is not important to human hearing, based on some psychoacoustic model. However, if the encoder can actually throw away the audio that is imperceptual to human listeners, then the watermark can most likely be removed, since, by definition, it should not affect human listeners significantly.

Converting digital audio to analog and then converting it back to digital (resampling) is another form of attack. Also, playing the audio out to speakers, and recording it through a mic is also difficult to handle. In both these cases, frequency-, amplitude-, and phase-dependent non-linearities can occur that destroy watermark data.

Finally, correlation attacks can be used to recover the keys used to watermark the original audio. If a watermarker uses the same key to encode multiple digital audio streams, then the result of these encodings can sometimes be correlated together to give insight into the location of the watermark and/or the key itself.

4 Perceptual Audio Coding

In this section, we describe how MPEG-1 level 3 and MPEG-2 level 3 (i.e., ".mp3" files) perceptual encoding works. This will provide necessary background to situate the various different watermarking techniques that need to be robust to such perceptual encoding.

MPEG-1/2 level 3 are desribed as "perceptual noise shaping" or "perceptual subband/transform coding" techniques. While the basic encoding technique for MPEG-1 and MPEG-2 level 3 audio is identical, MPEG-1 level 3 provides support for stereo sound coding (i.e., difference coding of two channels), while the MPEG-2 layer 3 version adds support for 5.1 sound spatialization, a few other "channelization" features, additional sampling frequencies, and additional support for very low bit-rates.

The main differences between MPEG-1/2 layer 3 and layers 1 or 2 include the use of the following:

- higher frequency resolution (18 times higher, specifically), which allows a Layer-III encoder to adapt the quantization noise to the masking threshold with more accuracy
- better alias reduction (aliasing is caused by overlap in the bands of the filterbank and because of overlapping of windows in the MDCT)
- entropy coding of data values (Huffman)
- non-uniform quantization for more consistent signal-to-noise ratios
- "noise allocation" rather than "bit allocation" – quantization noise due to adaptation of the quantization levels is measured and can therefore be accounted for and specifically allocated
- a bit reservoir to reduce artifacts during transients and other difficult-to-code segments
- more advanced joint-stereo coding.

The encoder analyzes the spectral components of the audio signal by calculating a filterbank (transform) and applies a psychoacoustic model to estimate the just noticeable noise-level. In its quantization and coding stage, the encoder tries to allocate the available number of data bits in a way to meet both the bitrate and masking requirements. The decoder is much less complex. Its only task is

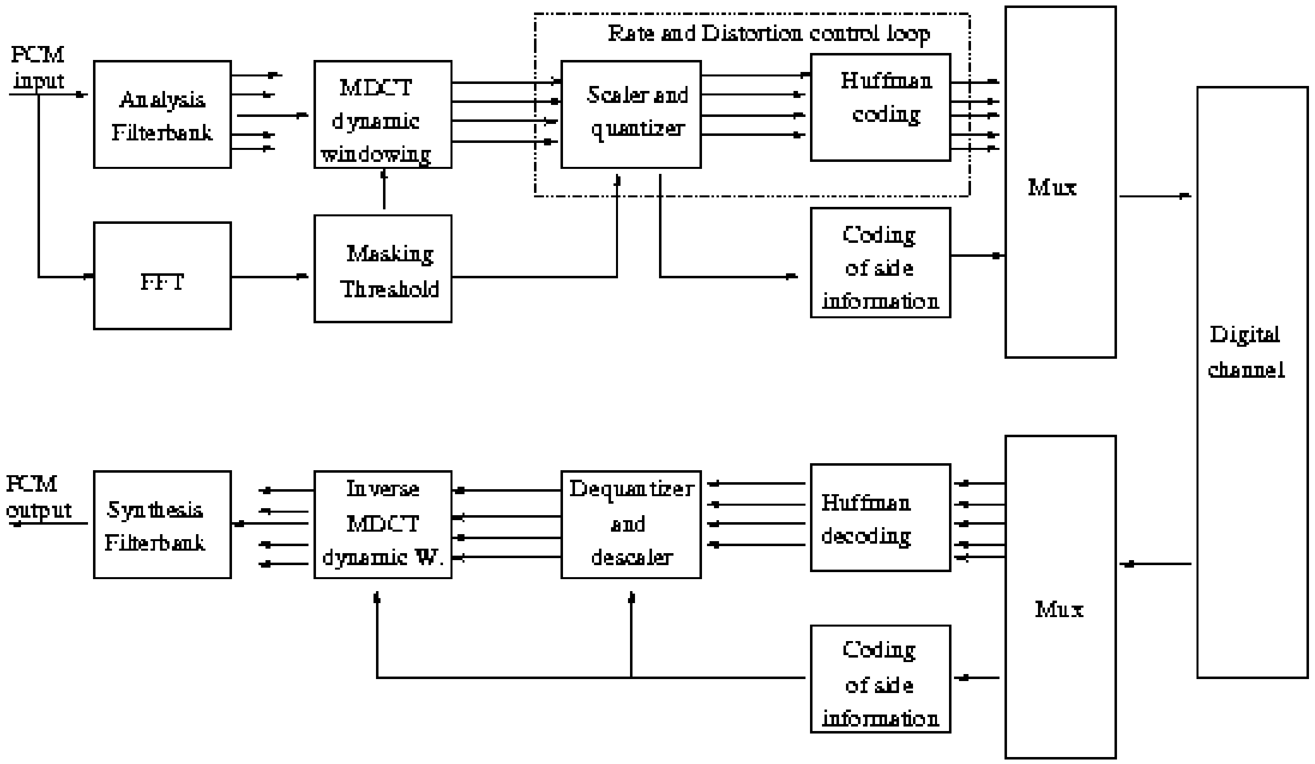

Figure 2: A block diagram of the MP3 encoding and decoding process. Figure courtesy D. Pan [32].

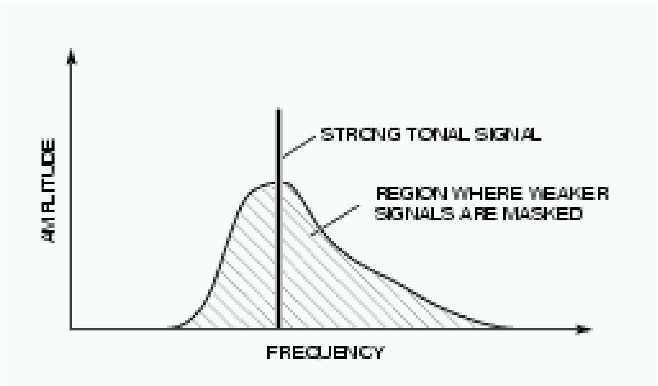

Figure 3: One form of psychoacoustic masking: adjacent frequency masking for signals occuring at the identical time. Figure courtesy of D. Pan [32].

to synthesize an audio signal out of the coded spectral components [32, 18, 15].

The coding process is illustrated in the block diagram in Figure 2, and consists of the following steps:

- Send audio signal into 32-band linear filterbank [43]. Since the bands are not aligned with the human critical bands (defined by the Bark scale), the lower-frequency bands contain large amounts of relevant information compared to other bands. These lower-frequency bands are allotted more bits in the "noise allocation" process to account for this.
- Send each individual band from the filterbank into a Modified Discrete Cosine Transform (MDCT) [36].
- As a side-chain processing step, send the audio signal into a Fast Fourier Transform. From the resulting spectrum, determine the following parameters:
  - the window length that should be used for the MDCT
    Longer windows optimize for spectral resolution (for periodic signals), while shorter windows optimize for temporal resolution (for transients).
  - masking due to sound adjacency
    The human ear *masks* certain sounds due to the presence of other high-amplitude sounds in the vicinity. This is a complex process that is frequency- and amplitude-dependent. Estimating such masking is done via extensive tables developed from careful psychoacoustic measurements. (Layer 3 tables are more complex than layer 1 or 2 tables.) Two major types of masking that occur are masking due to neighboring frequency components that are evident at the same time, as shown in Figure 3, and masking due to frequency components that occur either shortly before or shortly after each other in time.

- masking due to minimal audition threshold (aka absolute masking threshold or threshold in quiet)

  The minimal audition threshold of the ear is not linear. It is represented by carefully measured curves called the Fletcher-Munson equal-loudness curves [34]. It is not necessary to code sounds situated under this threshold, because they will not be perceived.

- masking due to tonal versus non-tonal components.

  The noise masking characteristics of each are different. The psychoacoustic model accounts for this because humans are very unforgiving of correllated noise effects (i.e. ones that cause tonal colorations), but are more forgiving of white (or slightly colored) random noise.

- From the side-chain FFT data and psychoacoustic model, an aggregate masking level for the signal is determined. From this and the filterband outputs, the signal-to-mask (SMR) ratio for each signal component is determined.

- The base sound pressure level of each component (SPL) and its SMR are used to iteratively (using a greedy algorithm) allocate bits to encode the components themselves. This is called "noise allocation" because SPL and SMR together are used to determine signal-to-noise ratios (SNRs) for each component. The iterative process assigns bits (and scale factors for those bits) in an attempt to minimize SNR.

- Determine from the signal properties which tables to use for Run-length encoding (RLE) and Huffman encoding. Perform the RLE and Huffman encode.

- Combine the payload (audio) data with the control data and any (optional) ancillary data, and apply an error-correcting code to the resulting bitstream.

5 Summary of Watermarking Mechanisms

We now summarize a few different approaches to watermarking perceptually-encoded digital audio. There are many approaches and variations available in the literature. This is a reasonably broad sampling rather than an exhaustive guide.

Before doing so, note that many of these techniques embed pseudo-noise (PN) sequences in the audio data. As a brief primer, PN-sequences are useful in this domain because of they exhibit nice correlation properties and are self-clocking. Self-clocking/reentrancy is required to maintain robustness to such attacks as cropping. They can be manipulated to contain particular characteristics such as frequency band-limiting [21], no DC-content (equal numbers of zeros and ones), and great length before repeating. Further description of PN-sequences are beyond the scope of this document.

5.1 Watermarking with Amplitude Modification

A basic approach to watermarking is to encode the information into the least-significant bits of the audio data. There are two basic classes of ways to do this: you can replace the lower order bits completely with a PN-sequence, or you can embed a PN-sequence into the existing low-order bitstream [6]. This technique works in the time domain by changing the amplitude of the audio data in a way that can be recovered given the PN-sequence. Varations on this approach include adaptively attenuating the amplitude of the embedded sequence to match the sound level of the current sound passage, and shaping the PN-sequence itself to match the underlying psychoacoustic masking characteristics to further bury the signal [8].

5.2 Watermarking with Dither

Dither [35] is a noise signal that is added to an input signal to provide better sampling of that signal. The use of dithering removes artifacts of quantization error, decorrelates the quantization error, and encodes signal amplitudes below the amplitude level of the minimum quantization increment. This embedding of "additional information" in the lower-order bit(s) of a sample occurs because the dither signal (added in front of an analog-to-digital converter) causes the sampler to make additional level transitions (i.e., the lower-order bit(s) flips back and forth), and these transitions embed a pulse-width modulated signal via the duty cycle. The result is that distortion (tonal, correlated noise) is almost completely minimized, at a cost of an increased noise floor. This is illustrated in Figure 4, where the top figure is a 1KHz signal sampled without dither (check out the periodic peaks), while the bottom figure has been sampled with dither (note the higher noise floor).

To implement dithering, a noise signal is added to the input signal with a particular probability distribution fuction (PDF). Common dither PDFs include Gaussian, triangular, and rectangular. To use dither for watermarking, the embedded watermark is used to modulate the dither signal and the host signal is quantized with an associated dithered quantizer. This is known as quantization index modulation (QIM) [9].

5.3 Watermarking with Echo

Watermarking via the use of echo [16] is based on the following idea: add a repeated version of a component of the audio signal with small enough offset, inital amplitude, and decay rate to make it imperceptible. Encode the watermarking bit by varing the offest (delay time) of the echoed signal.

The encoding implementation is illustrated in Figure 5-A and proceeds as follows: Partition the audio stream into a sequence of segments, as is illustrated in Figure 5-B. Then create two echo signal "kernels" with different delay times, as is illustrated in Figure 5-C. Echo creation is illustrated in Figure 5-D. To encode the watermark bitstream, fade between the two kernels using "mixer" signals as shown in Figure 5-E.

For watermark recovery ("decoding"), the autocorrelation of each signal segment's cepstrum is taken. This technique is called "cepstrum autocorrelation." The cepstrum autocorrelation produces a signal with two pronounced amplitude humps. The distance between the two humps determines whether a one or zero bit was encoded for that segment.

The echo-encoding technique works for certain types of signals because very-short-delay echos (on the order of 1 ms) are either imperceptible or heard only as "resonance," which may be acceptable for certain classes of audio signals (but *not at all* acceptable for others...). The fading between the kernels via the mixer signals is critically important to avoid discontinuities in the signal, which would cause aliasing and artifacts ("pops" or "clicks") at the discontinuity points.

A nice property of the technique is that it does not require the use of the original signal during recovery.

5.4 Watermarking with Phase Distortion

The phase distortion (or "phase coding") method works by substituting the phase of an inital audio segment with one of two reference phases. The reference phase determines whether a one or zero is encoded as watermarking data. In order to make this phase distortion minimally invasive, the phase of previous and subsequent segments are adjusted to preserve relative phase relationships. (Humans are very sensitive to relative phase differences but far less sensitive to absolute phase values.)

It has been claimed that phase coding is one of the most effective techniques in terms of SNR [7], but this is misleading because the

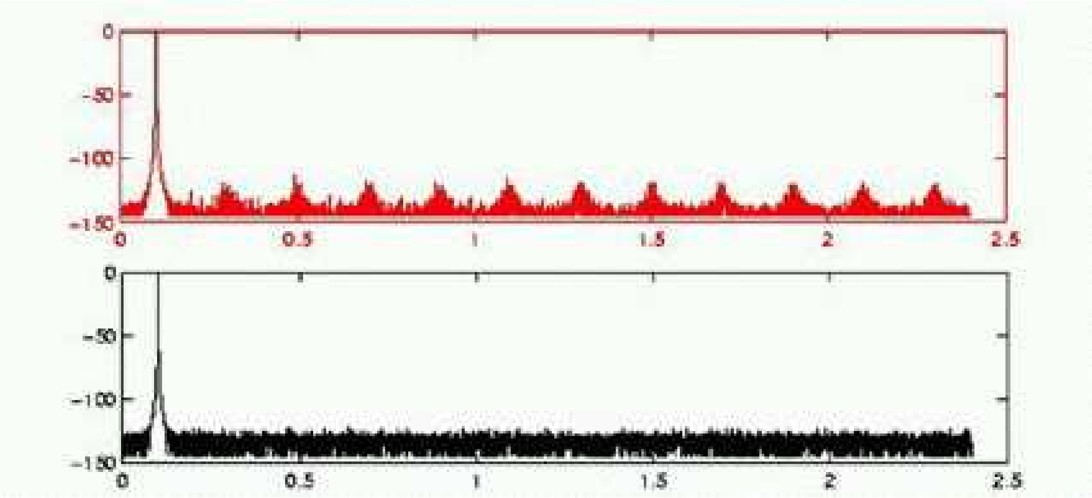

Figure 4: Comparision of dithered versus undithered sampling. Top: undithered 1KHz signal. Bottom: dithered 1KHz signal. Plots calculated by, and courtesy, of Jim Johnson, copyright jj@resarch.att.com 1998.

effects of the watermark are heard as artifacts ("warbling"), rather than noise.

In order to encode the watermarked signal, the following procedure is used: The sound (Figure 6-A) is broken into segments (Figure 6-B). For each segment, window the data and apply a Discrete Fourier Transform (DFT). This turns the time-domain segment into a set of frequencies, amplitudes, and phases (Figure 6-C). For each segment, compute the phase differences between them (Figure 6-D). Embed the codebit into the inital entry in the segment (Figure 6-E) by setting its phase to a particular value, and then update all subsequent segments using the phase differences calculated from before (Figure 6-F). Invert the DFT using the new phases but original frequencies and amplitudes (Figure 6-G), and concatenate them all to create the final ouput (Figure 6-H).

To recover the watermark, simply synchronize the sequence, perform phase detection on the signal, and compare the phases to the two reference phases at particular points.

In order to ensure smooth phase transitions, the data rate must be kept "low enough" to minimize phase dispersion – a break in the relationship of the phases between each of the frequency components that causes "beating." This constrains the data rate.

5.5 Spread Spectrum Watermarking

The use of spread spectrum techniques [10] to robustly encode data in a adversarial situation or over a channel with interference is borrowed from the communications community. The basic idea of spread spectrum is to spread the data across a large frequency band – in the case of audio, the entire audible spectrum. It is often done either through the use of frequency hopping (FHSS) – taking a narrow-band signal and hopping it around the wide band – or direct-sequence (DSSS), where a signal is modulated at low intensity across the entire bandwidth. For watermaking we report on this latter approach.

Watermarking with DSSS proceeds as follows. To encode, a "carrier" signal is modulated with the watermark and a PN-sequence (called the "chip" and acting as the key). The watermark bitstream is encoded as ones and negative ones (rather than one or zero) in order to affect bi-phase shift keying of the signal; additionally it should have error-correction (ECC) properties since noise will destroy portions. This process is illustrated in the top portion of Figure 7. This output is then attenuated and added to the original audio signal as what looks like random noise.

To recover the watermark, the audio signal is modulated with the chip signal again to remove it, and it is sent through a band-pass filter centered around the carrier frequency. The resulting carrier-frequency signal can have the watermark read off it by detecting its phase: in-phase implies the code bit is a one at that location, while out-of-phase implied the code bit is a negative one (i.e., zero). This process is illustrated in the bottom portion of Figure 7.

6 Balancing Concerns

6.1 Industry

As was stated earlier, for any secure digital music distribution system to become widely adopted and successful, it should balance the concerns of industry, the artists, and consumers.

From the viewpoint of the recording industry, the digital distribution of music opens the potential for new markets, such as purchasing individual songs, pay per listen, and music subscription services. But their chief concern is that any such system be secure, protect intellectual property rights, and allow them to continue to make profits from music distribution.

The recording industry also wants to maintain a certain level of control, on numerous fronts. The industry currently has a tremendous amount of power over artists, and there is the danger that their role might be diminished with digital music distrubution. For instance, if artists sold directly to consumers, it is possible that the traditional record labels would be much less relevant than today. The recording industry also wants to maintain control over technical standards. Despite the fact that the SDMI is a consortium of companies from numerious fields, including computers and consumer electronics, it is dominated by the RIAA. Quoting an (unnamed) consumer electronics industry executive, about the fees required to join the consortium, "They said, 'Pay us $50,000 and present your proposal, and we'll get back to you with something.' This isn't a democratic standards organization. This is the music industry acting out of fear." [45]

6.2 Artists

Artists are generally interesting in having their music being widely distributed, and reaching as wide an audience as possible. This has

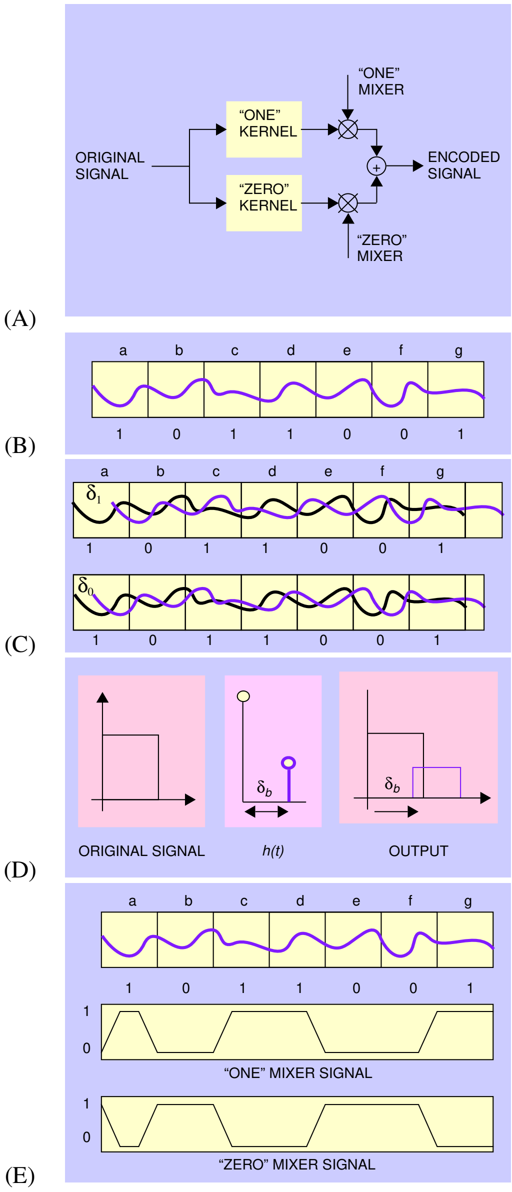

Figure 5: The echo watermarking technique. Figures courtesy of Bender, et.al. [7].

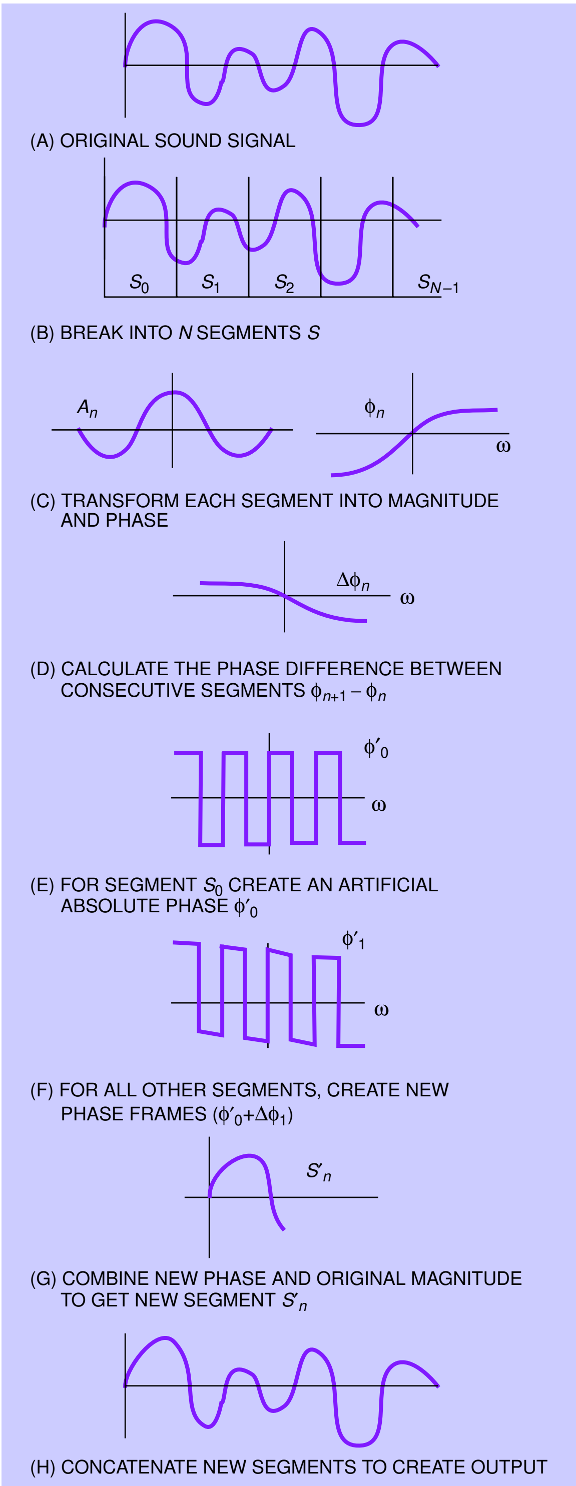

Figure 6: The phase distortion watermarking technique. Figure courtesy of Bender, et.al. [7].

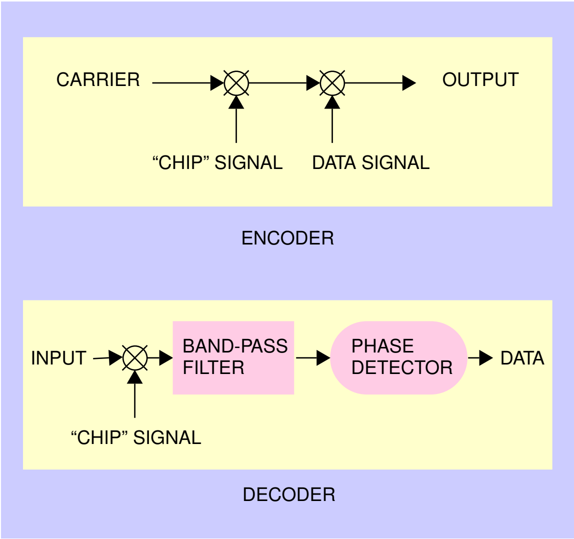

Figure 7: The direct-sequence spread spectrum watermarking technique. Figure courtesy of Bender, et.al. [7].

traditionally been facilitated by record labels. But digital distribution offers artists the possibility to directly reach consumers, and this has the possibility of shifting the balance of power in the music industry. The truth is that artists currently receive a small percentage of the music purchase price as royalties, and digital distribution gives artists the potential for greater control and profits.

There can be downsides of completely eliminating the middlemen, however. [The Artist formerly known as] Prince tried directly selling to his fans a few years ago and was largely unsuccessful.[1] Record labels provide many services for artists, including promotion, obtaining radio airplay, etc., that artists may not be willing or able to do on their own. One plausible scenario for the future is that record labels don't completely disappear, but today's traditional labels are augmented or replaced with newer labels that either concentrate or solely rely upon digital distribution. For instance, Alanis Morissette and Tori Amos recently signed deals [17] with mp3.com [30], a leading provider of (legal, unprotected) MP3 files over the Internet.

The RIAA claims to represent the views of the artists, and the concerns that they have voiced about the protection of intellectual property are reasoned as being on behalf of the artists. A counter view to this, however, is that the RIAA largely represents their own interests and is concerned about protecting their existing power. The recording industry is scared of MP3s, but it is not clear that all artists feel the same way. For example, Public Enemy recently was forced by their label (PolyGram) to remove MP3 files from its web site [46]. Similar situations have occurred with other artists. And it is not clear that small, independent labels will ever go along with a secure digital music distribution system; they may be perfectly satisfied with unprotected MP3s.

Artists are also concerned with protecting the integrity of their music. A watermark, even if perceptually insignificant, does alter the original content of the audio. It is possible that artists will not be willing to go along with any such system on the grounds that the music should be left as is. And the artists' hired mastering/mixdown engineers/producers will be loathe to degrade the quality of their work, because quality is exactly what their reputations are based upon.

[1] It could be argued that Prince's failure was due to other reasons. This was before the explosive growth of consumer use of the Internet. The music was purchased using a toll-free 800 number, and it was delivered via regular mail. This is quite different than purchasing and delivering music over the Internet.

6.3 Consumers

Consumers are not directly concerned with security; they want music, its delivery, and music devices to be convenient, low cost, and easy to use. Consumer electronics manufacturers also share these same concerns, since happier consumers are likely to lead to higher profits. There are some tradeoffs between these desires and security. A music playback device with security features has to have *some* additional functionality compared to an unsecure device, and that is likely to make such a device more expensive. Furthermore, there is a definite danger that adding such features will in some ways be inconvenient and/or hard to use for consumers; it is difficult to imagine any scenario in which security increases convenience or makes devices easier to use [45]. It is the belief of Michale Robertson, president of MP3.com, that "users will never go for any security-laden system." [31]

Consumers are not going to want to sacrifice any functionality that they currently enjoy with respect to music. They will want interoperability; any CD can be played on any CD player. They will want to be able to loan music to friends; the purchase of a CD is not tied to any particular person or any particular player - it can easily be shared. They will want to be able to resell music that they no longer want in an analogous manner to that of the thriving used CD market.

Any secure digital music system that attempts to prevent illegal use should not interfere with legitimate use. There is the danger that an RIAA-approved system, like SDMI, may not meet this criteria. The RIAA wants a prohibition on microphone inputs on SDMI-compliant devices, they don't want such devices to be able to play unprotected MP3s, and they want restrictions on user's access to software for encoding with SDMI [45]. All of these run counter to the interests of consumers.

A previous example of a copyright-protection system for digital audio was the Serial Code Management System (SCMS) for Digital Audio Tape (DAT) recorders, mandated by the Audio Home Recording Act of 1992 [3]. Two bits within the data stream signify copy permissions: either no restrictions at all (00), allow one generation copy (11), or do not allow copies (10) [11]. But this indeed interfered with legitimate uses of consumer equipment. For instance, any recording through the analog inputs on a non-professional tape deck encoded with SCMS 11. Therefore, amateur musicians recording themselves, or anybody legally taping a live recording, ended up with restricted recordings. It was realitively easy to defeat this system, and people did sell devices to strip SCMS from a data stream. It is interesting to note, however, that the recently enacted Digital Millenium Copyright Act would likely make it illegal to distribute such a device, even if it was used for a legitimate purpose [39].

Industry has a history of asserting their interests in ways that run counter to the consumers' interests. Universal Studios and Walt Disney Productions sued Sony (maker of the Betamax video cassette recorder), arguing that recording television programs and movies is copyright infringement. (The case went all the way to the United States Supreme Court, which ruled in favor of Sony.) Sony was later sued by the music industry in a (failed) attempt to prevent sales of DAT recorders and blank cassettes [19]. This is all happening once again. The RIAA has sued Diamond Multimedia [44], the manufacturer of the Rio portable MP3 music player [14], claiming that it is a recording device and therefore is covered by the Audio Home Recording Act [3], which requires manufacturers of digital recording devices to pay royalties to the music industry and implement copy protection technology in their products.

Consumers are concerned about audio quality, and they may be resistant to accept a system that they perceive as harming audio quality by altering the digital data (such as with the inclusion of a watermark), whether or not such differences are noticeable to the user. The degree to which this matters will vary from consumer to consumer. Audiophiles may never accept any lossy compressed music format (and some won't accept any digitized format, prefering LPs to CDs), but the vast majority of the market consists of non-audiophiles. For people who are willing to accept possible quality degradation through MP3 encoding (and whose equipment and/or untrained ears may not be good enough to tell the difference between the compressed file and the original), slight perterbations in the audio through watermarking are not likely to make much of a difference.

Consumers may be concerned about their privacy. A digital distribution system, as well as watermarks which have the potential of tying identification and purchase information to a particular piece of music, increases the amount of information available about consumers and their buying habits.

Any digital distribution system (whether or not it is secure) will force consumers to sacrifice the artwork and the proposition of owning the physical media that is provided by the current system. This may not be amenable to some consumers, most notably collectors.

7 Conclusion

In this paper, we have presented an overview of digital audio watermarking schemes and how they relate to digital audio distribution. We have discussed many of the current commerical and academic efforts to solve the problem of protecting copyrighted material. We have also presented several digital audio watermarking algorithms that may be used to aid in copy detection. The current algorithms are diverse, employing techniques such as echo hiding and spread spectrum in order to hide a bitstream in an existing digital audio stream without perceptually affecting it. We also discussed many of the concerns that are affecting the key players in developing a digital audio distribution scheme. Regardless of whether watermarking balances the concerns of industry, artists, and consumers, there are clearly open issues in the domain.

Hiding copyright ownership information into digital audio such that it cannot be removed seems technically viable if you believe the technical papers in the field. But, as is analogous to cryptographic systems, you can only prove what attacks these schemes are resistent to, not that they are completely secure (i.e., resistant to *future* attacks). There has not been sufficent investigation into *negative* results (i.e., not enough "cryptanalysis" of audio watermarking schemes) to determine whether any proposed scheme is actually viable.

In order for watermarking to be successful, all forms of copyrighted digital audio must contain the watermark. This is impossible though: all current music exists in non-watermarked form.

In summary, one point is clear: during the next decade, significant changes will occur in the creation and distribution of digital audio. Whether the scheme will include watermarking or not is a good question, and only time will tell.

References

[1] ANDERSON, L. Music giants fight a corporate war online. CNN interactive, April 14, 1999. `http://cnn.com/TECH/computing/9904/14/musicwar.idg/`.

[2] AT&T. a2b music Technology. `http://www.a2bmusic.com/technology.asp`.

[3] The Audio Home Recording Act of 1992. Copyright Act of 1976, As amended. Chapter 10. Digital Audio Recording Devices and Media. `http://www.hrrc.org/audio.html`.

[4] AUDIOSOFT. `http://www.audiosoft.com`.

[5] BAKOS, Y., BRYNJOLFSSON, E., AND LICHTMAN, D. Shared Information Goods. Journal of Law and Economics, December 1998.

[6] BASSIA, P., AND PITAS, I. Robust Audio Watermarking in the Time Domain. *Proceedings of ICASSP* (1999).

[7] BENDER, W., ET AL. Techniques for Data Hiding. *IBM Systems Journal* (1996), 313–36.

[8] BONEY, L., TEWFIK, A., AND HAMDY, K. Digital Watermarks for Audio Signals. *International Conference on Multimedia Computing and Systems* (1996), 473–80.

[9] CHEN, B., AND WORNELL, G. W. Dither Modulation: A New Approach to Digital Watermarking and Information Embedding. *Proc. of SPIE: Security and Watermarking of Multimedia Contents 3657* (Jan. 1999), 342–353.

[10] COX, I., ET AL. Secure Spread Spectrum Watermarking for Images, Audio, and Video. *International Conference on Image Processing 3* (1996), 234–246.

[11] DAT-heads Frequently Asked Questions, September 2, 1992. `http://www.eklektix.com/dat-heads/FAQ`.

[12] DEUTSCH, R. W. Liquefying MP3. Wired News, January 23, 1999. `http://www.wired.com/news/news/culture/story/17498.html`.

[13] DEUTSCH, R. W. More Tales from Encryption. Wired News, January 18, 1999. `http://www.wired.com/news/news/culture/mpthree/story/17340.html`.

[14] DIAMOND MULTIMEDIA. Rio PMP300. `http://www.diamondmm.com/products/current/rio.cfm`.

[15] FRAUNHOFER-GESELLSCHAFT. Frequently Asked Questions about MPEG Audio Layer-3. `http://www.iis.fhg.de/amm/techinf/layer3/layer3faq/index.html`.

[16] GRUHL, D., LU, A., AND BENDER, W. Echo Hiding. *Information Hiding First International Workshop Proceedings* (1996), 295–315.

[17] HARING, B. Morissette, Amos plug into MP3. USA Today, April 26, 1999. `http://www.usatoday.com/life/music/lmds538.htm`.

[18] HERRE, BRANDENBURG, ET AL. Second Generation ISO/MPEG Audio Layer-3 Coding. *98th AES* (Feb. 1995).

[19] HOME RECORDING RIGHTS COALITION. A Chronology of Events in Home Recording Rights. `http://www.hrrc.org/history.html`.

[20] IBM. IBM Cryptolope Technology - Executive Summary. `http://www.software.ibm.com/security/cryptolope/about.html`.

[21] IKEDA, M., TAKEDA, K., AND ITAKURA, F. Audio Data Hiding By the Use of Band-Limited Random Sequences. *Proceedings of ICASSP* (Mar. 1999).

[22] INTERNATIONAL ORGANISATION FOR STANDARDISATION. Overview of the MPEG-4 Standard. Section 2.8. Content-related IPR identification and protection, March 1999. `http://drogo.cselt.it/mpeg/standards/mpeg-4/mpeg-4.htm#E10E14`.

[23] INTERTRUST. `http://www.intertrust.com/`.

[24] JONES, C. Crypto Creeps into MP3 Domain. Wired News, January 18, 1999. `http://www.wired.com/news/news/culture/mpthree/story/17390.html`.

[25] LACY, J., SNYDER, J., AND MAHER, D. Music on the Internet and the Intellectual Property Protection Problem. *Proceedings of ISIE* (July 1997). http://www.a2bmusic.com/docs/musicipp.doc.

[26] LIQUID AUDIO. http://www.liquidaudio.com/.

[27] MCY MUSIC. http://www.mcy.com/.

[28] MJUICE. http://www.mjuice.com.

[29] MP3 Hardware Guide. http://www.mp3.com/hardware/.

[30] MP3.COM. http://www.mp3.com.

[31] NICKELL, J. Liquid and Iomega Make Music. Wired News, November 16, 1998. http://www.wired.com/news/news/business/story/16294.html.

[32] PAN, D. Digital Audio Compression. *Digital Technical Journal 5*, 2 (Spring 1993).

[33] PEGORARO, R. Net Has Music Giants Singing Wary Tune. *Washington Post* (April 29, 1999), A1.

[34] PIERCE, J. *The Science of Musical Sound*, Revised ed. W.H. Freeman and Company, 1983.

[35] POHLMANN, K. *Principles of Digital Audio*, Third ed. McGraw Hill, 1995.

[36] PRINCEN, J., JOHNSON, A., AND BRADLEY, A. Subband / Transform Coding Using Filter Bank Designs Based on Time Domain Alias Cencellation. *Proceedings of ICASSP* (1987), 2161–2164.

[37] RICHTEL, M. Record Label Will Distribute Music Online. *New York Times* (May 5, 1999), B1.

[38] ROBERTSON, M. Can Music Be Secure? http://www.mp3.com/news/115.html.

[39] SAMUELSON, P. Intellectual Property And The Digital Economy: Why The Anti-Circumvention Regulations Need To Be Revised. *Electronic Commerce Conference* (March 1999). http://www.sims.berkeley.edu/~pam/dmca/.

[40] SHERMAN, C. Presentation to SDMI Organizing Plenary, February 26, 1999. https://www.sdmi.org/dscgi/ds.py/GetRepr/File-38/html.

[41] SHIVAKUMAR, N. Large Scale Copy Detection. http://www-db.stanford.edu/~shiva/SCAM/scamInfo.html.

[42] Sony Develops Copyright Protection Solutions For Digital Music Content. Sony Corporation of America, February 25, 1999. http://www.sony.com/SCA/press/feb_25_99.html.

[43] SPORER, T., BRANDENBURG, K., AND EDLER, B. The Use of Multirate Filter Banks for Coding of High Quality Digital Audio. *6th European Signal Processing Conference (EUSIPCO) 1* (June 1992), 211–214.

[44] STAMPER, C. Diamond Countersues, Defends Rio. Wired News, December 2, 1998. http://www.wired.com/news/news/business/story/16586.html.

[45] STRAUSS, N. Pirate-Proof Music on Web? So Far, That Does Not Compute. *New York Times* (April 24, 1999), B1, B14.

[46] SULLIVAN, J. Public Enemy's Chuck D on MP3. Wired News, December 5, 1998. http://www.wired.com/news/news/culture/story/16597.html.