Preface

Recent development of the Internet and the digital information revolution caused significant changes in the global society, ranging from the influence on the world economy to the way people nowadays communicate. Broadband communication networks and multimedia data available in a digital format (images, audio, and video) opened many challenges and opportunities for innovation. Versatile and simple-to-use software and decreasing prices of digital devices (e.g., digital photo cameras, camcorders, portable CD and mp3 players, CD and DVD recorders, laptops, etc.) have made it possible for consumers from all over the world to create, edit, and exchange multimedia data. Broadband Internet connections and almost an errorless transmission of data facilitate people to distribute large multimedia files and make identical digital copies of them.

Digital media files do not suffer from any quality loss due to multiple copying processes, such as analogue audio and VHS tapes. Furthermore, recording medium and distribution networks for analogue multimedia are more expensive. These advantages of digital media over the analogue ones transform to disadvantages with respect to the intellectual rights management because of a possibility for unlimited copying without a loss of fidelity cause a considerable financial loss for copyright holders (Cox, Miller, & Bloom, 2003; Wu & Liu, 2003; Yu, Kundur, & Lin, 2001). The ease of content modification and a perfect reproduction in digital domain have promoted the protection of intellectual ownership and the prevention of the unauthorized tampering of multimedia data to become an important technological and research issue (Kundur, 2001).

Therefore, the traditional methods for copyright protection of multimedia data are no longer sufficient. Hardware-based copy protection systems have already been easily circumvented for the analogue media. Hacking of digital media systems is even easier due to the availability of general multimedia processing platforms, for example, a personal computer. Simple protection mechanisms that were based on the information embedded into header bits of the digital file are useless because header information can easily be removed by a simple change of data format, which does not affect the fidelity of media.

Encryption of digital multimedia prevents access to the multimedia content to an individual without a proper decryption key. Therefore, content providers get paid for the delivery of perceivable multimedia, and each client that has paid the royalties must be able to decrypt a received file properly. Once the multimedia has been decrypted, it can be repeatedly copied and distributed without any obstacles. Modern software and broadband Internet provide the tools to perform it quickly and without much effort and deep technical knowledge.

Basic Definitions and Terms in Digital Watermarking

Digital watermarking has been proposed as a new, alternative method to enforce the intellectual property rights and protect digital media from tampering. It involves a process of embedding into a host signal a perceptually transparent digital signature, carrying a message about the **host signal** in order to "mark" its ownership. The digital signature is called the **digital watermark**. The digital watermark contains data that can be used in various applications, including digital rights management, broadcast monitoring, and tamper proofing. Although perceptually transparent, the existence of the watermark is indicated when watermarked media is passed through an appropriate watermark detector.

A watermark, which usually consists of a binary data sequence, is inserted into the host signal in the **watermark embedder**. Thus, a watermark embedder has two inputs; one is the watermark message (usually accompanied by a secret key) and the other is the host signal (e.g., image, video clip, audio sequence, etc.). The output of the watermark embedder is the **watermarked signal**, which cannot be perceptually discriminated from the host signal. The watermarked signal is then usually recorded or broadcasted and later presented to the **watermark detector**. The detector determines whether the watermark is present in the tested multimedia signal, and if so, what message is encoded in it. The research area of watermarking is closely related to the fields of information hiding (Anderson & Petitcolas, 1998; Johnson, Duric, & Jajodia, 2001) and steganography (Johnson & Jajodia, 1998; Katzenbeisser & Petitcolas, 1999).

Therefore, we can define **watermarking systems** as systems in which the hidden message is related to the host signal and **nonwatermarking** systems in which the message is unrelated to the host signal. On the other hand, systems for embedding messages into host signals can be divided into **steganographic systems**, in which the existence of the message is kept secret, and **nonsteganographic systems**, in which the presence of the embedded message does not have to be secret.

Audio watermarking initially started as a subdiscipline of digital signal processing, focusing mainly on convenient signal processing techniques to embed additional information to audio sequences. This included the investigation of a suitable transform domain for watermark embedding and schemes for the imperceptible modification of the host audio. Only recently watermarking has been placed to a stronger theoretical foundation, becoming a more mature discipline with a proper base in both communication modelling and information theory.

Digital Audio Watermarking and the Human Auditory System

Watermarking of audio signals is more challenging compared to the watermarking of images or video sequences due to wider dynamic range of the human auditory system (HAS) in comparison with the human visual system (HVS) (Bender, Gruhl, & Morimoto, 1996). The HAS perceives sounds over a range of power greater than 10^9:1 and a range of frequencies greater than 10^3:1. The sensitivity of the HAS to the additive white Gaussian noise (AWGN) is high as well; this noise in a sound file can be detected as low as 70 dB below ambient level. On the other hand, opposite to its large dynamic range, HAS contains a fairly small differential range, that is, loud sounds generally tend to mask out weaker sounds. Additionally, HAS is insensitive to a constant relative phase shift in a stationary audio signal and some spectral distortions interprets as natural, perceptually non-annoying ones (Bender et al., 1996).

Auditory perception is based on the critical band analysis in the inner ear where a frequency-to-location transformation takes place along the basilar membrane. The power spectra of the received sounds are not represented on a linear frequency scale but on limited frequency bands called **critical bands** (Steinebach, Petitcolas, Raynal, Dittmann, Fontaine, Seibel, et al., 2001). The auditory system is usually modelled as a bandpass filterbank, consisting of strongly overlapping bandpass filters with bandwidths around 100 Hz for bands with a central frequency below 500 Hz and up to 5000 Hz for bands placed at high frequencies. If the highest frequency is limited to 24000 Hz, 26 critical bands have to be taken into account.

Two properties of the HAS dominantly used in watermarking algorithms are **frequency (simultaneous) masking** and **temporal masking** (Steinebach et al., 2001). The concept using the perceptual holes of the HAS is taken from wideband audio coding (e.g., MPEG Compression 1, Layer 3, usually called

MP3) (Noll, 1993). In the compression algorithms, the holes are used in order to decrease the amount of the bits needed to encode audio signal, without causing a perceptual distortion to the coded audio. On the other hand, in the information hiding scenarios, masking properties are used to embed additional bits into an existing bit stream, again without generating audible noise in the audio sequence used for data hiding.

The simplest visualization of the requirements of information hiding in digital audio is so called **magic triangle** (Johnson et al., 2001). Inaudibility, robustness to attacks, and the watermark data rate are in the corners of the magic triangle. This model is convenient for a visual representation of the required trade-offs between the capacity of the watermark data and the robustness to certain watermark attacks, while keeping the perceptual quality of the watermarked audio at an acceptable level. It is not possible to attain high robustness to signal modifications and high data rate of the embedded watermark at the same time. Therefore, if a high robustness is required from the watermarking algorithm, the bit rate of the embedded watermark will be low and vice versa; high bit rate watermarks are usually very fragile in the presence of signal modifications. However, there are some applications that do not require that the embedded watermark as a high robustness against signal modifications. In these applications, the embedded data are expected to have a high data rate and to be detected and decoded using a blind detection algorithm. While the robustness against intentional attacks is usually not required, signal processing modifications, like noise addition, should not affect the covert communications (Cox et al., 2003). To qualify as steganography applications, the algorithms have to attain statistical invisibility as well.

Contents of the Book

The book is organized as follows:

Chapter I gives a general overview of the audio watermarking fundamental definitions. Audio watermarking algorithms are characterized by five essential properties, namely perceptual transparency, watermark bit rate, robustness, blind/informed watermark detection, and security. Chapter I also reviews the most common signal processing manipulations that are frequently applied to the watermarked audio in order to prevent detection of the embedded watermark. Finally, several application areas for digital audio watermarking are presented and advantages of digital watermarking over standard technologies examined.

Chapter II is a comprehensive presentation of spread spectrum-based digital audio watermarking methods. The problem is viewed as the realization of a basic communications system where the host signal presents the available channel and the watermark presents the transmitted information that needs to be recovered in conditions that include noise distortion, standard compression, and deliberate attacks. Basic spread spectrum theory as it relates to audio watermarking is introduced followed by state-of-the-art improvements. Finally, the role of psychoacoustics in effective watermarking is emphasized and an enhanced psychoacoustic model based on the discrete wavelet packet transform (DWPT) is included.

In Chapter III, the use of parametric synthesis models in digital audio watermarking is described. It argues that, because human auditory perception is not a linear process, the optimal hiding of binary data in digital audio signals should consider parametric transforms that are generally nonlinear. To support this argument, an audio watermarking algorithm based on aligning frequencies of spectral peaks to grid points is presented as a case study; its robustness is evaluated and benefits are discussed. Toward the end, the future research directions are given, including watermark-aided sound source segregation and counter-measure against arithmetic collusive attacks.

Robust audio watermarking for copyright protection and digital rights management (DRM) in general has been an active research area during the last few years. Despite the achieved progress, there are still may open issues and challenges. Two state-of-the art approaches towards blind robust audio watermarking are presented in Chapter IV. Both utilize a correlation detection scheme and watermarks generated by chaotic systems.

In Chapter V, watermarking techniques are classified as one of three schemes: nonblind watermarking scheme, blind watermarking schemes with and without synchronization information. In this chapter, three audio watermarking techniques are described to illustrate the three different schemes. The time-frequency technique belongs to the nonblind watermarking scheme; the multiple-echo hiding technique and the peak-point extraction technique fall under the blind watermarking schemes with and without synchronization information, respectively.

Chapter VI introduces time-spread echo hiding as an advanced audio watermarking method based on echo hiding. Differences in the structure of echo kernels between the ordinary echo hiding and the time-spread echo hiding are schematically depicted to explain advantages of the new method. Several watermark-extracting methods are introduced for decoding process to raise the detection rate for various conditions. Performance of the method in robustness against several attacks is evaluated in terms of **delay** because of its statistical preciseness. The time-spread echo hiding exhibited fairly better performance than the ordinary echo hiding.

In Chapter VII, detailed explanations would be given on the role of echo hiding playing in audio watermarking, in terms of background, functions, and applications. Additionally, a section is dedicated to discuss the various approaches proposed in the past to solve the flaws of echo hiding. Lastly, the proposed analysis-by-synthesis echo watermarking scheme based on interlaced kernels is introduced. Comparisons in audio quality and robustness performance are also looked at the proposed and conventional echo watermarking schemes.

The patchwork algorithm is investigated in detail in Chapter VIII. The performance of this algorithm in terms of imperceptibility and robustness has been shown to be superior. Robustness of the patchwork algorithm against the curve-fitting attack and blind multiple-embedding attack is presented as well. Toward the end, robustness against jitter attack which is a natural enemy of this watermarking algorithm is studied.

In Chapter IX, an overview of our time-frequency (TF) based audio watermarking methods is presented. TF techniques provide flexible means to analyze nonstationary audio signals. The joint TF domain for watermark representation is given and pattern recognition schemes for watermark detection were employed. In this chapter; two watermarking methods are introduced: embedding nonlinear and linear TF signatures as watermarking signatures. Robustness of the proposed methods against common signal manipulations is also studied.

In Chapter X, a brief overview of audio watermarking with the focus on a novel audio watermarking scheme is given. The scheme is based on watermarks with a “semantic” meaning and offers a method for MPEG Audio Layer 3 files which does not need the original watermark for proving copyright ownership of an audio file. The main feature of this scheme is that it offers enhanced capability of proving the ownership of the audio file not by simply detecting the bit pattern that comprises the watermark itself, but by showing that the legal owner knows a hard to compute property of the watermark bit sequence (“semantic” meaning).

The main objective Chapter XI is to provide an overview of existing speech watermarking technology and to demonstrate the importance of speech processing concepts for the design and evaluation of watermarking algorithms. This chapter describes the factors to be considered while designing speech watermarking algorithms, including the choice of the domain and speech features for watermarking, watermarked signal fidelity, watermark robustness, data payload, security, and applications. The state-of-the-art robust and fragile speech watermarking algorithms are presented and their advantages and disadvantages discussed.

Chapter XII discusses the robustness of audio watermarking algorithms against digital-to-analogue and analogue-to-digital conversions. It provides an overview on distortions caused by converting the signal in various scenarios. The aim is to show the complexity of influences that need to be taken into account. Additionally, experimental results of our author's own audio watermarking algorithm are presented, with respect to analogue recordings using a microphone which proves that a high robustness in this area can be achieved. At the end, strategies against downmixing and playback speed changes of the watermarked audio signal are presented.

Methods for evaluating the quality of watermarked objects are detailed in Chapter XIII. A presentation of subjective evaluation standards used in testing the transparency of marked audio tracks is given, as well as the evaluation of marked items with intermediate quality. Since subjective listening tests are expensive and dependent on many not easily controllable parameters, objective quality measurement methods are discussed. The process of testing the quality taking into account the methods discussed is presented as well, with particular emphasis to a detailed description of the test setup, item selection and the practical limitations.

Digital watermarking studies have always been driven by the improvement of robustness. Most of articles of this field deal with this criterion, presenting more and more impressive experimental assessments. On the contrary, security received little attention in the watermarking community. Chapter XIV presents a comprehensive overview of this recent concept. The authors list the typical applications which require a secure watermarking technique. For each context, a threat analysis is purposed. Using this presentation, the authors illustrate all the certainties the community has on the subject, the intuitions and future trends.

REFERENCES

Anderson, R., & Petitcolas, F. (1998). On the limits of steganography. *IEEE Journal on Selected Areas in Communications, 16*(4), 474–481.

Bender, W., Gruhl, D., & Morimoto, N. (1996). Techniques for data hiding. *IBM Systems Journal, 35*(3), 313–336.

Cox, I., Miller, M., & Bloom, J. (2003). *Digital watermarking*. San Francisco: Morgan Kaufmann.

Johnson, N., Duric, Z., & Jajodia, S. (2001). *Information hiding: Steganography and watermarking-attacks and countermeasures*. Boston: Kluwer Academic Publishers.

Johnson, N., & Jajodia, S. (1998). Steganalysis: The investigation of hidden information. In *Proceedings of the IEEE Information Technology Conference* (pp. 113–116). Syracuse, New York.

Katzenbeisser, S., & Petitcolas, F. (1999). *Information hiding techniques for steganography and digital watermarking*. Norwood, MA: Artech House.

Kundur, D. (2001). Watermarking with diversity: Insights and implications. *IEEE Multimedia, 8*(4), 46–52.

Noll, P. (1993). Wideband speech and audio coding. *IEEE Communications Magazine, 31*(11), 34–44.

Steinebach, M., Petitcolas, F., Raynal, F., Dittmann, J., Fontaine, C., Seibel, S. et al. (2001). Stirmark benchmark: Audio watermarking attacks. In *Proceedings of the International Conference on Information Technology: Coding and Computing* (pp. 49–54). Las Vegas, Nevada.

Wu, M., & Liu, B. (2003). *Multimedia data hiding*. New York: Springer Verlag.

Yu, H., Kundur, D., & Lin, C. (2001). Spies, thieves, and lies: The battle for multimedia in the digital era. *IEEE Multimedia, 8*(3), 8–12.