Signal Processing 128 (2016) 222–242

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Review

Twenty years of digital audio watermarking—a comprehensive review

Guang Hua [a,*], Jiwu Huang [b], Yun Q. Shi [c], Jonathan Goh [d], Vrizlynn L.L. Thing [d]

[a] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798
[b] College of Information Engineering, Shenzhen University, Shenzhen 518060, China
[c] Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA
[d] Cyber Security & Intelligence Department, Institute for Infocomm Research, Singapore 138632

ARTICLE INFO

Article history:
Received 15 December 2015
Received in revised form
8 March 2016
Accepted 12 April 2016
Available online 13 April 2016

Keywords:
Audio watermarking
Robust watermark
Data hiding
Copyright protection
Media security

ABSTRACT

Digital audio watermarking is an important technique to secure and authenticate audio media. This paper provides a comprehensive review of the twenty years' research and development works for digital audio watermarking, based on an exhaustive literature survey and careful selections of representative solutions. We generally classify the existing designs into time domain and transform domain methods, and relate all the reviewed works using two generic watermark embedding equations in the two domains. The most important designing criteria, i.e., imperceptibility and robustness, are thoroughly reviewed. For imperceptibility, the existing measurement and control approaches are classified into heuristic and analytical types, followed by intensive analysis and discussions. Then, we investigate the robustness of the existing solutions against a wide range of critical attacks categorized into basic, desynchronization, and replacement attacks, respectively. This reveals current challenges in developing a global solution robust against all the attacks considered in this paper. Some remaining problems as well as research potentials for better system designs are also discussed. In addition, audio watermarking applications in terms of US patents and commercialized solutions are reviewed. This paper serves as a comprehensive tutorial for interested readers to gain a historical, technical, and also commercial view of digital audio watermarking.

Contents

* Corresponding author.
E-mail addresses: ghua@ntu.edu.sg, huaguang86@gmail.com (G. Hua), jwhuang@szu.edu.cn (J. Huang), yun-qing.shi@njit.edu (Y.Q. Shi), jonathan-goh@i2r.a-star.edu.sg (J. Goh), vriz@i2r.a-star.edu.sg (V.L.L. Thing).

http://dx.doi.org/10.1016/j.sigpro.2016.04.005
0165-1684/

1. Introduction

Digital audio watermarking is an important research branch of multimedia data hiding [1–5], which involves embedding the watermarks into host audio data and when necessary, performing watermark extraction for copyright protection, authentication, and other digital rights management (DRM) purposes. The original work on information hiding, i.e., dirty paper writing, was carried out from a communication theory perspective, which dated back to 1983 [6], while the first reported work on digital audio watermarking was seen in 1996 [1] and the first systematic work was presented in 1997 [7]. Therefore, there has been a history of about twenty years for digital audio watermarking. Within the twenty years, the advanced signal processing techniques have been efficiently utilized for this topic, and numerous solutions have been proposed alongside [6,8–36,7,37–70]. A generic digital audio watermarking system is depicted in Fig. 1, where the terms in solid rectangles specify the general phases of the signal manipulations in an audio watermarking system and the terms in dashed rectangles represent possible users who manipulate the data. Note that normal users may “attack” the watermarked signal unintentionally during the processes of lossy compression, equalization, or adding effects, etc. Thus, we also call such “processing attacks” as unintentional attacks while deliberate attacks aiming at destroying or removing the watermarks are referred to intentional attacks. The watermark extraction phase in Fig. 1 is also termed as watermark detection in many existing works. While watermark detection and extraction could refer to a similar signal processing purpose at the receiver end, they are actually slightly different tasks. Specifically, if the receiver end implements a threshold-based correlation and detection scheme, and the copyright is claimed by true positive detection results, then such a process is usually termed watermark detection. In this case, the original watermarks need to be available in order to calculate the corresponding correlation functions. However, it is also quite often that the receiver aims at restoring the original watermark sequence from watermarked copies without the knowledge of the original watermarks. In this situation, such a process is more precisely termed as watermark extraction. For simplicity, in this paper, the two terms are treated as interchangeable and used where appropriate, and the original watermarks are assumed to be known at the receiver end unless otherwise mentioned. Also note that some literature uses encoding and decoding to describe the embedding and extraction processes, which can also be considered as equivalent descriptions.

The effectiveness of an audio watermarking system is characterized by several performance criteria [2], i.e., *imperceptibility*, *robustness*, *security*, *capacity*, *computational complexity*, etc. First, imperceptibility characterizes the fidelity of watermarked audio data, indicating that the embedded watermarks should introduce perceptually indistinguishable changes to the host signal. Therefore, sometimes, fidelity, transparency, and inaudibility are used equivalently as imperceptibility. Second, robustness refers to the availability of successful watermark extraction when the watermarked signal has been attacked intentionally or unintentionally. It is the most complicated feature for an audio watermarking system because of the variety of attacks. Third, security means that the system should be designed in such a way that only authorized parties are able to extract the watermarks. Fourth, the amount of information that can be embedded into the given host data is called capacity. At last, the designed system is preferred to be computationally efficient.

For conciseness and efficiency, the comprehensive review reported in this paper mainly considers imperceptibility and robustness among the criteria, because they determine the key performance of most existing audio watermarking systems. In contrast, security is usually achieved via the use of random keys, which is widely incorporated in most of the existing solutions. However, we would like to note here that although not being in the scope of this paper, watermarking security is also an active research area and interested readers can refer to [71,72] and references therein for more information. For the criterion of embedding capacity, it is usually optional in those systems with typical objectives of successful extraction of the watermarks and declaring the ownership and copyright of the audio files. In this situation, watermarks are in the forms of random sequences (simply a mark, as can be seen from most of the reviewed works in this paper), and whether the watermarks correspond to meaningful information or how much it contains are less important. A different situation would be seen in another research area, i.e., steganography [73], where the hidden information (now it becomes a message rather than a mark) itself becomes important to establish covert communications in, e.g., military and health applications. Another important difference between watermarking and steganography is that at the extraction phase, the former has the knowledge of the watermarks and focuses mainly on matching it with the extracted version, while the latter does not have any clues of the hidden message. Lastly, based on current literature, none of the existing solutions suffer from severe computational issues, and the embedding and extraction of watermarks are usually performed “off-line” (there do exist several “on-line” applications, which are not the major focus of this review). Thus, we also exclude computational complexity. To further clarify the scope of this paper, we incorporate the classification of