Time-Scale Invariant Audio Watermarking Based on the Statistical Features in Time Domain

Shijun Xiang[1,2], Jiwu Huang[1,2], Rui Yang[1,2]

[1] School of Information Science and Technology,
Sun Yat-sen University, Guangzhou 510275, China
[2] Guangdong Key Laboratory of Information Security Technology,
Guangzhou 510275, China
isshjw@mail.sysu.edu.cn

Abstract. In audio watermarking, the robustness to desynchronization attacks such as TSM (Time-Scale Modification) operations, is still an open issue. In this paper, both mathematical proof and experimental testing show that the histogram shape (represented as the relative relation in the number of samples among three different histogram bins) and the audio mean are two robust features to the TSM attacks. Accordingly, a multi-bit robust audio watermarking algorithm based on the two statistical features is proposed by modifying the histogram. The audio histogram with equal-sized bins is extracted from a selected amplitude range referred to the audio mean, and then the relative relations in the number of samples among groups of three neighboring bins are designed to carry the watermark by reassigning the number of samples in the bins. The watermarked audio signal is perceptibly similar to the original one. Simulation results demonstrated that the hidden message is very robust to the TSM, cropping, and a variety of other distortions in Stirmark Benchmark for Audio.

1 Introduction

Audio watermarking [1][2] plays an important role in ownership protection. According to IFPI (International Federation of the Phonographic Industry) [3], STEP2000 [4] and SDMI (Secure Digital Music Initiative) [5], audio watermarking should be robust to temporal scaling of $\pm 10\%$ and be able to resist most common signal processing manipulations and attacks, such as random cropping, MP3 compression, resampling and etc.

Among the various problems to be solved in audio watermarking, the robustness against desynchronization distortions such as TSM and random cropping, is the most challenging one for previous watermarking schemes yet. Desynchronization attacks that cause displacement between encoder and decoder are difficult for a watermark to survive. In [6], the synchronization code was introduced aiming at conquering cropping attacks. However, the synchronization code is very vulnerable to TSM. For example, a small amount of scaling (i.e.,

J. Camenisch et al. (Eds.): IH 2006, LNCS 4437, pp. 93–108, 2007.

±1%) will be able to cause the watermark extraction failed. TSM is a common audio processing manipulation in a variety of software tools, such as CoolEditPro V2.1. Under TSM operations, even with the scaling amount of ±10%, the auditory quality of audio is still rather perfect since HAS (Human Auditory System) is not sensitive to TSM. This makes TSM to be a serious attack operation in audio watermarking. Generally, there are mainly two modes of TSM operations, *pitch-invariant* scaling and *resample* scaling. The pitch-invariant mode preserves audio pitch, while the resample mode keeps pitch and tempo neither by modifying playback speed.

Few algorithms can effectively resist the TSM. In the existing literature, several algorithms have been proposed aiming at solving this problem by using exhaustive search [7], synchronization pattern [6][8], invariant watermark [9][10], implicit synchronization [11], informed detection [12]. In [7], the authors applied the detection engine to search the watermark for resynchronization by performing multiple correlation tests. One possible problem for multiple correlation tests is the false alarm. In [9], the authors proposed a time-scale invariant watermarking embedding strategy by changing the length of the intervals between two successive peaks of the smoothed waveform. The watermark detection highly depends on the selection of the threshold. In [10], the authors presented an audio watermarking method by using music content analysis. The watermark is embedded into the edges of audio signals by using FFT (Fast Fourier Transform) technique. The watermark is robust to ±9% pitch-invariant TSM but vulnerable to resample stretching mode, which will change the edges in the signal. The watermarking methods based on the peak points may suffer from the attack of removing the peaks. In [12], side information is exploited to improve the searching of the watermark aiming at solving playback speed modifications. One weakness of this scheme is that the detection procedure is not blind. On the basis of [8], Tachibana [13] further improved the watermark performance against random time stretching from ±4% up to ±8% by using multiple pseudo-random arrays. This method is time consuming.

The above mentioned approaches share the problems that the watermark performance is difficult to satisfy the requirements of IFPI or STEP2000 (stretching of ±10%), and, the watermarking schemes usually focus on one type of desynchronization attacks, such as pitch-invariant TSM mode [10], resample TSM mode or playback speed modifications [12], and cropping [6]. In this paper, we propose a multi-bit audio watermarking algorithm based on the audio statistical features described by the histogram specification, concentrating on combating the desynchronization problem caused by time-scale modifications. Histogram-based watermarking strategy was first introduced for image watermarking in [14]. By using the robustness of image color histogram to rotations and geometric transformations, the authors in [15] proposed a general method for watermarking color histogram of image. The 1-bit watermarking scheme is very robust to image geometric distortions. The basic idea in our algorithm is that the TSM operations with the resample and pitch-variant stretching modes may be represented as an approximately temporal linear scaling operation in practice, verified by extensive