Signal Processing 123 (2016) 157–166

Contents lists available at ScienceDirect

Signal Processing

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

Authentication and recovery algorithm for speech signal based on digital watermarking

Zhenghui Liu [a,b], Fan Zhang [b], Jing Wang [c], Hongxia Wang [d], Jiwu Huang [a,*]

[a] College of Information Engineering, Shenzhen University, Shenzhen 518060, China
[b] College of Computer and Information Technology, Xinyang Normal University, Xinyang 464000, China
[c] College of Mathematics and Information Science, Xinyang Normal University, Xinyang 464000, China
[d] School of Information Science and Technology, Southwest Jiaotong University, Chengdu, Sichuan 610031, China

ARTICLE INFO

Article history:
Received 28 June 2015
Received in revised form
25 September 2015
Accepted 27 October 2015
Available online 30 November 2015

Keywords:
Content authentication
Tamper detection
Tamper recovery
Digital watermarking

ABSTRACT

A content authentication and tamper recovery scheme for digital speech signal is proposed. In this paper, a new compression method for speech signal based on discrete cosine transform is discussed, and the compressed signals obtained are used to tamper recovery. One block-based large capacity embedding method is explored, which is used for embedding the compressed signals. For the scheme proposed, watermark is generated by frame number and compressed signal. If watermarked speech is attacked, the attacked frames can be located by frame number, and reconstructed by using the compressed signal. Theoretical analysis and experimental results demonstrate that the scheme not only improves the security of watermark system, but also can locate the attacked frames precisely and reconstruct the attacked frames.

1. Introduction

For the development of high-speed networks and the explosive increase of digital audio devices, digital audio signals used are increasingly growing, and provide convenience for people's life. While, with the increasingly rich of production tools, digital audio signals are easily be edited and attacked. The meaning of attacked signal is different from the original one to be expressed. If audiences or users consider the attacked signal is the original one, and act according to the instructions of the attacked signal, it may cause serious consequences. So, it is necessary to verify the authenticity of speech signal firstly. At the same time, after the attacked signals being detected, the recovery of the attacked content can minimize the user's loss.

As to digital speech signal, there are a lot of achievements in the field of the research on speech enhancement [1–5] and speaker recognition [6–9]. As to content authentication, the method based on digital watermark provides a solution to verify the authenticity of speech signal. Digital watermarking technique has achieved an outstanding progress in the past few years. The mainly results are focused on robust audio watermark schemes [10–13], authentication and recovery schemes for digital images[14–18]. However, the authentication and recovery schemes for digital speech are rarely [19–21]. In [19], authors proposed a speech content authentication algorithm based on Bessel-Fourier moments. The scheme has the ability of tamper location for maliciously attacks, and has stronger robustness to common signal processing operations. In [20], the author proposed an authentication scheme for compressed audio recordings, using detection of multiple compression and encoder's identification. The compressed digital audio recordings are authenticated by evaluation of statistical features extracted from MDCT coefficients and other parameters obtained from

[*] Corresponding author.
E-mail address: jwhuang@szu.edu.cn (J. Huang).

http://dx.doi.org/10.1016/j.sigpro.2015.10.023
0165-1684/© 2015 Elsevier B.V. All rights reserved.

compressed audio files, which are used for training selected machine learning algorithms. The scheme enhances the robustness and the effectiveness. While the method need a large number of training data, which is inconvenienced in application. In [21], for compressed speech signal, authors proposed an authentication scheme based on compression technique and codebook-excited linear prediction. Watermark bits are generated by the features extracted during compression process based on codebook-excite linear prediction. As to the speech signals compressed based on other speech codecs, the scheme is ineffective. In addition, the embedding method is based on lest significant bits (LSBs), which is fragile to signal processing operations. For the scheme, it regards the signal processing operations as hostile attack.

Although most of the present authentication schemes have the ability of tamper location, they cannot reconstruct the attacked content. Once speech signal is attacked, tamper recovery will be very important since it is related to the critical business of data owners.

The design of tamper recovery scheme is much more difficult. To sum up, the reasons are as follows: (1) Tamper location is the first step for tamper recovery schemes. However, the tamper location method existed based on synchronization codes cannot locate the attacked signals precisely and has some shortcomings [12,22,23]. For the schemes, on the one hand, the synchronization codes embedded are vulnerable to be attacked [19], which cannot be detected. On the other hand, the signal between two neighboring synchronization codes is regarded as watermarked signal. It does not verify the authenticity of the watermarked signal. That is to say, if the watermarked signal is subjected to attack, it will not be detected. (2) It is hard to generate the signal used to tamper recovery and easily to be embedded. For the recovery schemes based on digital watermark, the signals used to tamper recovery are embedded into speech. Great number of embedding reduces the speech quality and intelligibility seriously. So, the signals used to tamper recovery should be as small as possible. (3) It is hard to embed the generated signal used to tamper recovery, for lacking the corresponding large capacity embedding method.

Considering the background, a tamper recovery scheme based on digital watermark is proposed in order to solve the problems above. In this paper, a compression and reconstruction method for digital speech signals based on discrete cosine transform (DCT) is discussed, and one block-based large capacity embedding method is explored. Frame number and compressed signal are as watermark and embedded. Frame number is used to locate attacked frame precisely, and compressed signal is used to reconstruct the attacked signal. Theoretical analysis and experimental evaluation results demonstrate that the scheme proposed improves the security and the accuracy of tamper location, and has a certain ability of tamper recovery for speech signals.

The organization of this paper is as follows. Section 2 introduces the compression and reconstruction method for speech signal. Section 3 describes the block-based large capacity embedding method, the tamper location and recovery scheme. Section 4 analyzes the performance of the algorithm theoretically, in which abilities of the scheme are compared with other schemes. In Section 5, experiment results are presented, which demonstrate the effectiveness of the proposed scheme. Finally, the conclusion is summarized in Section 6.

2. Signal compression, reconstruction and scrambling

The signal used to reconstruct the attacked content is essential to tamper recovery schemes. In this paper, the original speech signal is compressed to generate the signal used to tamper recovery. And the compressed signal is regarded as watermark and embedded. As we known, watermark embedded will affect the quality of speech signal. So, the original speech should be compressed as small as possible under the condition of high recovery quality. In the following, the compression and reconstruction methods for digital speech signal based on DCT are discussed. Denote $A = \{a_l, 1 \le l \le L\}$ as the original speech signal, and F is the sampling frequency of A.

2.1. Signal compression

The process of signal compression is shown in Fig. 1, and the details are as follows.

Step 1: Divide the original signal A into P non-overlapping frames. The i-th frame is denoted by A_i, and the length of A_i is L/P.
Step 2: Resample the signal A using the sampling frequency F', $F' < F$. The signal obtained is denoted by A', and the length of A' is $L' = L \cdot F'/F$.
Step 3: Similar to the first step, A' is cut into P frames. A'_i is denoted as the i-th frame, and the length of each frame is L'/P.
Step 4: DCT is performed on the i-th frame A'_i, and the coefficient is denoted by $D_i = \{d_{i,j}, 1 \le j \le L'/P\}$. $C_i = \{c_{i,j} | c_{i,j} = d_{i,j} 1 \le j \le M\}$, the anterior M coefficients, is regarded as the compressed signal of A_i, $M \ll L'/P$.

2.2. Signal reconstruction

Based on the compression method and the compressed signal obtained above, the reconstruction steps are listed as follows.

Step 1: Generate $D'_i = \left\{ d'_{i,j} 1 \le j \le L'/P \right\}$ by using the compressed signal $C_i = \{c_{i,j}, 1 \le j \le M\}$, where $d'_{i,j} = c_{i,j}$, $1 \le j \le M$; $d'_{i,j} = 0$, $M+1 \le j \le L'/P$.
Step 2: Inverse DCT is performed on D'_i. And resample the signal obtained using the sampling frequency F, to reconstruct the signal A'_i.

In the following, performance of the reconstruction method is tested. The original speech signal is selected and shown in Fig. 2(a), and the reconstructed one is shown in

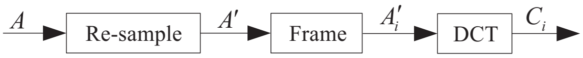

Fig. 1. The process of signal compression.