Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 931–938
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Speech authentication system using digital watermarking and pattern recovery

Chang-Mok Park *, Devinder Thapa, Gi-Nam Wang

Department of Industrial and Information System Engineering, AJOU University, San 5, Woncheon-dong, Yeongtong-gu, Suwon 442-749, South Korea

Received 8 February 2006; received in revised form 25 October 2006
Available online 10 January 2007

Communicated by O. Siohan

Abstract

The objective of this paper is to detect speech forgery using digital audio watermarking and pattern recovery techniques. A digital watermark pattern has been attached with the speech signal to detect three kinds of alterations or forgeries such as substitution, insertion, and removal. The watermark pattern will be modified if some changes have been made to the speech contents. Modification and forgery can be measured and detected by pattern recovery. The proposed method uses the cyclic pattern embedding to overcome synchronizing problems of previous detection techniques. In addition, pattern recovery enhances the robustness to compression. This method has been tested and verified using six recording devices, which was used for collecting verbal data. The speech signals were sampled at the rate of 8 kHz and digitized at 16 bits resolution. Randomly chosen regions were substituted, removed, and compressed in MP3 at the rate of 16 kbps as well as in CELP at the rate of 11.5 kbps. The experiment shows the perfect detection for three kinds of forgeries and it proved the validity of the proposed method.

Keywords: Speech authentication; Digital audio watermarking; Spread spectrum; Speech forgery

1. Introduction

With the rapid advancement of digital storage devices, speech recording has been widely used as evidence in the court of law. This makes the speech authentication task very important as well as vulnerable. The reliability of stored speech required proper security by the inhibition of access to content. Encryption is one way of preserving the digital information; however, digital forgery in signal processing level cannot be protected completely by this method. Digital forgeries like substitution, insertion and removal of the signals are difficult to detect with human ears. It creates the need for some reliable signal analysis method and advance authentication tools. One of the data hiding and extracting techniques is digital watermarking, for digital rights protection (Katzenbeisser and Petitcolas, 2000). Digital watermarking can be used for speech authentication.

Data hiding using audio watermarking is one way of signal modification in the non-critical band of HAS (Human auditory System). In this technique some of the signals which are not audible to the human ears are removed or inserted (Swanson et al., 1998). Spread spectrum was the first technique applied for the digital audio watermarking. This method adds the random signal which is masked by the psycho-acoustic model or the linear prediction filter model, and detection of watermark existence has been accomplished by cross-correlation (Kirovski and Malvar, 2001; Cox et al., 1996; Veen et al., 2001). The limitation of this technique is watermark information can be removed by digital filtering. Echo hiding is another audio watermarking technique based on the concept that humans

* Corresponding author. Tel.: +82 31 219 2429; fax: +82 31 219 1610.
E-mail addresses: cmpark25@hanafos.com (C.-M. Park), debu@ajou.ac.kr (D. Thapa), gnwang@ajou.ac.kr (G.-N. Wang).

0167-8655/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2006.12.010

cannot hear the echo signal when the time period between the original signal and the echo signal is very short (Gruhl et al., 1996; Bender and Gruhl, 1996). This method utilizes the Cepstrum analysis procedure for extraction of the speech signals. It is easy to develop and robust to digital compression; however, the watermark can be easily revealed and removed. Digital filtering methods embed the watermark information using all pass filters (Ciloglu and Karaaslan, 2000). It changes the phase information so that quality of sound is good even though the watermark information can be easily removed by simple signal processing. Some of the existing methods attach duplicate signals as watermark, which can be generated by simple modification of original signals (Petrovic, 2001). This paper has proposed an approach to enhance the robustness and effectiveness of secure speech signal using digital audio watermarking and pattern recovery techniques.

2. Related works

Watermarking is a good technique for speech authentication; however, we can find a few literature studies which deal with this type of problems. Farid (1999) has described a method known as polyspectral analysis. This approach is effective in detecting the presence of digital forgery. In the frequency domain a "natural" signal has weak higher-order statistical correlations, if the signal is passed through digital tampering, "un-natural" higher-order correlations are introduced that can be measured by polyspectral analysis (Farid, 1999). Explicit prior knowledge of the original signal is not needed in this method, but localized and minor tampering are likely unnoticeable.

Two approaches have been adopted for speech content authentication. The first approach is based on content feature extraction. In this method, the agent passes encrypted features to a receiver and the receiver extracts those content features from the encoded speech. After that, integrity is tested by comparing the decrypted features and the extracted features. It has some limitations such as auxiliary data, which are encrypted features, must be stored and desynchronization between each end must be resolved (Wu and Jay Kuo, 2002a). The second approach embeds a fragile watermark into the speech signal in the frequency domain. The modification uses the scale quantization in that domain. To detect the altering of speech segments, Wu and Jay Kuo (2002a), use a kind of correlation technique, but it cannot detect the intentional cutting of speech segments because there is no temporal information. To detect the cutting of speech segments, Wu and Jay Kuo (2002b) compared the original bit sequences with the extracted bit sequences and determined the authenticity of forgery. This approach has the synchronization problem as two bit sequences must be harmonized.

The proposed method deals with the existing problems of the watermarking method for speech authentication. It embeds the cyclic pattern in the speech signal using watermarking so the extracted pattern does not need to be synchronized as compared to other existing approaches. If the speech content is modified, it consequentially modifies the embedded pattern. The modified region is measured by the distortion of cyclic pattern. The spread-spectrum watermarking method is adopted as an embedding scheme. In the forgery analysis phase, robust extraction and smoothing technique for pattern recovery has been developed. The smoothing technique using curve optimization enhances the robustness of the authentication system.

This paper is organized as follows: the proposed approach has been presented in Section 3, the watermarking method has been explains in Sections 4 and 5 explained about the watermark extraction and recovery pattern, detailed experimental results have been given in Section 6; finally Section 7 concludes with some comparative analysis and references.

3. Proposed approach

The sequential procedure of the proposed approach has been depicted in Fig. 1.

The process starts with the input of the raw speech signal. This can be stored in a PCM format. Pre-detection of the watermark is conducted to avoid redundancy. If the pre-detection does not find any watermark, it embeds the watermark pattern in the speech signal. After watermarking, the watermarked speech signals can be stored in a compressed or a raw format. To detect the forgery or authentication of speech contents, we have to follow a further process. The speech signal should be converted to the PCM signal from stored formats like MP3, WMA, etc. The watermark pattern should be recovered using extraction of bit information and preprocessing of the pattern. Detection forgery of the watermarked speech signal can be checked by the analysis of pattern degradation. A detailed proce-

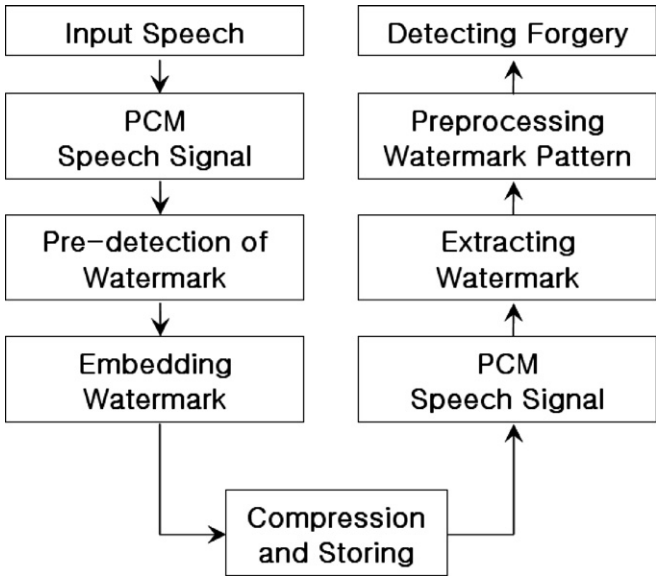

Fig. 1. Overall procedures of the speech authentication system.