Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 48 (2006) 1608–1619
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Speaker verification security improvement by means of speech watermarking

Marcos Faundez-Zanuy [a,*], Martin Hagmüller [b], Gernot Kubin [b]

[a] Escola Universitaria Politècnica de Mataró, Telecommunication, Avda. Puig i Cadafalch 101-111, 08303 Mataro, Barcelona, Spain
[b] Graz University of Technology, Signal Processing and Speech Communication Laboratory, Inffeldgasse 12, 8010 Graz, Austria

Received 21 March 2006; received in revised form 8 June 2006; accepted 13 June 2006

Abstract

This paper presents a security enhanced speaker verification system based on speech signal watermarking. Our proposed system can detect several situations where a playback speech, a synthetically generated speech, a manipulated speech signal or a hacker trying to imitate the speech is fooling the biometric system. In addition, we have generated a watermarked speech signals database from which we have obtained relevant conclusions about the influence of this technique on speaker verification rates. Mainly we have checked that biometrics and watermarking can coexist simultaneously minimizing the mutual effects. Experimental results show that the proposed speech watermarking system can suffer A-law coding with a message error rate lower than 2×10^{-4} for SWR higher than 20 dB at a message rate of 48 bits/s.

Keywords: Biometric; Speech watermarking; Speaker verification

1. Introduction

Our previous work (Faundez-Zanuy, 2004) stated the necessity for a constant update in security systems in order to keep on being protected. A suitable system for the present time can become obsolete if it is not periodically improved. Usually, the combination of different systems and/or security mechanisms is the key factor (Faundez-Zanuy, 2005). Thus, in this paper, we propose the combination of a speaker recognition biometric system with a watermarking algorithm that will allow to check the genuine origin of a given speech signal and if the recording has been manipulated (edited), which is useful for forensic applications.

In (Faundez-Zanuy, 2004), we studied the vulnerability points of a biometric system. They are mainly eight, but the most critical ones are the first and second points, which correspond to the sensor level and the transmission of the sensed signal. While the other blocks and transmission lines can be secured in a proper fashion, it will be useless if the first points of the chain are fooled.

Fig. 1 shows the two vulnerability points that we are trying to secure with our new proposed system. It plots the situation where a remote recognition

* Corresponding author. Tel.: +34 937574404; fax: +34 937570524.
E-mail addresses: faundez@eupmt.es (M. Faundez-Zanuy), hagmueller@tugraz.at (M. Hagmüller), g.kubin@ieee.org (G. Kubin).

0167-6393/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2006.06.010

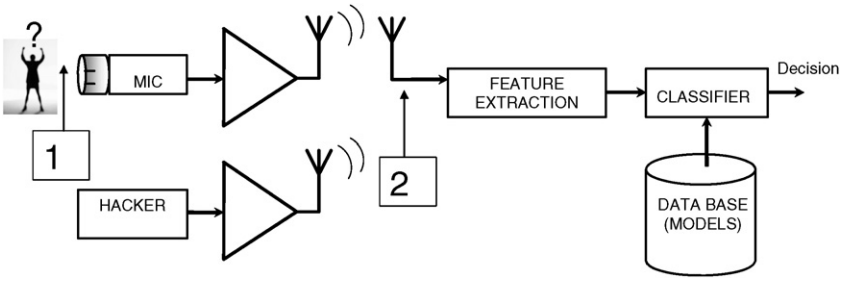

Fig. 1. General biometric system and two possible vulnerability points.

takes place, and the signal is electromagnetically radiated. An example of this situation would be a mobile telephone recognition system or a base station communicating with an airplane. Without loss of generality, this scheme is also valid for an electric wire connecting the sensor with the feature extractor. This last situation can be stated in a remote internet recognition system.

The vulnerability points that we try to secure are

1. *Sensor level*: In this level, a fake biometric characteristic is presented at the sensor. It can be, for instance, a tape recording of an authorized user.
2. *Transmission of the sensed signal*: In this level, the attack consists of digitally stored biometric data belonging to the authorized user, a synthetically generated speech signal, or a hacker trying to imitate a given voice. This situation is similar to the first point, but the biometric data comes from a different origin (marked as "hacker" in Fig. 1), which is more difficult to secure than point 1 (it is difficult to know the origin of the fake speech signal). Obviously this possibility is specially important in remote applications, where there is a client computer that provides the biometric data and a remote host system that performs the biometric authentication. For "on site" recognitions, this point can be almost neglected, because it is weaker to crack the first one.

In this kind of applications the fraudulent acquisition of biometric data by third parties is possible in one of the following ways, which obviously should be avoided with properly security measures:

(a) Acoustic recording of the speech signal. This implies a hidden microphone in the same room than the genuine speaker. This is marked with "1" in Fig. 1. This is the most difficult situation to manage, because the signal acquisition is clean and genuine. However, it is expected that this site will be well-controlled and secured.
(b) For electromagnetic transmission, a radio frequency receiver can listen to the speech signal and store it. For internet applications, the procedure consists of intercepting the electric signal from the electric wire. This is marked with "2" in Fig. 1.

Another new potential use of our proposed system is for forensic applications. Police security forces must prove in front of a court that a telephone recording (which has been previously authorized by the judge) has not been manipulated. Formerly, with analog recordings, they had a technique able to track the recorder's header, but with digital recordings they do not have this possibility.

Using a speech watermarking strategy it is possible to introduce a time stamp code in side the signal. This time-stamp must be encrypted in such a way that this mark can be checked but nobody can replace any speech section with a proper mark. Thus, a watermarked signal can be checked for continuity on this time-stamp, the presence of gaps, some illegal filling speech frames, etc. Thus, our proposed system can solve a drawback produced by the advances in the recording technologies: to demonstrate that a given speech recording has not been manipulated. In addition, this speech watermark must be transparent for a speaker recognizer algorithm, because recognizing the speaker identity is an important issue for forensic experts.

2. Proposed enhanced biometric system

If the communication links are not secured, a replay attack is possible. It consists of resubmission of previously intercepted biometrics or biometric