ADAPTIVE WATERMARKING TECHNIQUE FOR SPEECH SIGNAL AUTHENTICATION

Dr. Methaq Talib Gaata and Refah Aamer Jaafar

Computer Science Department, University of Mustansiriyah, Baghdad, Iraq

ABSTRACT

Biometrics data recently has become a major role in determining the identity of the person. With such importance for the use of biometrics data, there are many attacks that threaten the security and integrity of biometrics data itself. Therefore, it becomes necessary to protect the originality of biometrics data against manipulation and fraud. This paper presents an authentication technique to achieve the authenticity of speech signals based on adaptive watermarking technique. The basic idea is depends on extracting the speech features from the speech signal initially and then using these features as a watermark. The watermark information embeds into the same speech signal. The short time energy technique is used to identifying the suitable positions for embedding the watermark in order to avoid the regions that used in the speech recognition system. After exclusion the important areas that used in speech recognition the Genetic Algorithm (GA) is used to generate random locations to hide the watermark information in an intelligent manner. The experimental results have achieved high efficiency in establishing the authenticity of speech signal and the process of embedding watermark not effect on features that used in speech recognition.

KEYWORDS

Adaptive Watermarking, Biometric, Speech Processing, Genetic Algorithm

1. INTRODUCTION

Biometric technology has a basic rule for determining the identity of the person, and forms the front of the systems that require precise identity. Also use of biometric systems conveniently significantly (there is no need to carry or remember such passwords and cannot stolen, cannot replace compared with traditional systems). Biometric systems based on physiological and behavioral characteristics of the person such as fingerprint, iris, retina, facial, speech, signature, etc. The biometric systems could work properly only when the biometrics data that entered to the systems are from legitimate source and when sure the biometrics data has not tampered. Therefore, the authentication of biometrics data is of important research issues for protection against tampered attacks [1].

Cryptography and watermarking techniques are among potential techniques to ensure the authentication of biometrics data. The techniques of cryptography introduce a high level of security but need a high time complexity and also do not provide any level of security when the biometrics data is decrypted. On the contrary, the techniques of watermarking based on embedding watermark information into the biometrics data itself without degrading to their features which used through the process of determining the identity of the authorized person. Therefore, it offers authentication of biometrics data even after the decryption process. As a result, a watermarking techniques is usually imperceptible, robust to unauthorized persons and capable of detect any trying to tamper for achieving the authentication of biometrics data [2].

DOI:10.5121/ijcsit.2016.8406

Digital watermarking is one of exciting technique for achieve authentication of digital signal. A digital watermarking techniques divided into two types are spatial domain and frequency domain for embedding watermark data securely into digital signal. The spatial domain techniques are based on direct manipulation of the elements of the digital signal such as the Least Significant Bits (LSB). While in the frequency domain techniques, the digital signal must be converted to frequency domain by several of transformation ways such as the Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT) and etc. Then, the watermark information is embedded into transform coefficients [3, 4].

Recently, there are some of researches that are interested to guaranteeing the authentication of biometrics data based on digital watermarking techniques. *R. Thanki* and *K. Borisagar* [5] proposed a biometric watermarking scheme by using the compressive sensing (CS) theory and the Fast Discrete Curvelet Transform for protection the face and fingerprint images. The sparse measurements of the fingerprint image are embedded as a watermark into the host face image. *M. Ali Nematollahi et al* [6] proposed a semi-fragile and blind digital speech watermarking technique for online speaker recognition systems based on the discrete wavelet packet transform (DWPT) and quantization index modulation (QIM). The watermark bits are inserted within an angle of the wavelet's sub-bands of the speech signal. *Z. Liu et al* [7] proposed a digital speech watermarking authentication and tamper detection and recovery scheme. The watermark represents the number of frames and the compressed speech signal. The number of frames used to locate precisely the attacked frame and the compressed speech signal used for reconstruct the attacked speech signal. The watermark data embedded into segments of speech signal.

This paper organized as follows. Section 2 presented the proposed watermarking technique. Section 3 displayed the experimental results. Section 4 give the conclusions of this paper.

2. PROPOSED WATERMARKING TECHNIQUE

The block diagram of the proposed technique is shown in Figure (1).

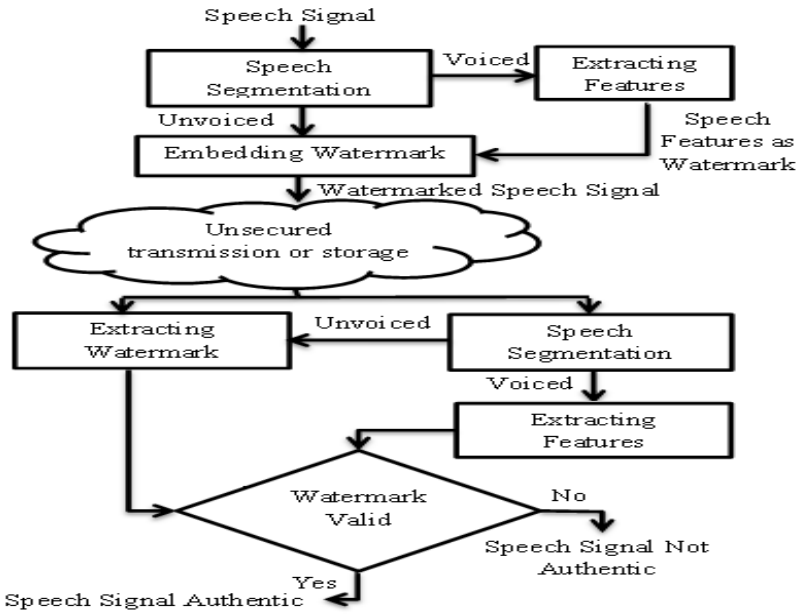

Figure 1. The block diagram of proposed watermarking technique.

As noted in the above figure, the proposed watermarking technique involves embed the speech features as a watermark into the same speech signal after keeping out the important areas that used in speech recognition. In order to prevent the deformation that may be produce during embedding the watermark data into the speech signal that potentially affected on the features which utilized in speech identification system. So, the embedding positions must be selected in an adaptive manner depending on features analysis based on short time energy technique.

2.1. SPEECH SEGMENTATION

The segmentation operation is applied to speech signal in order to segment the speech signal into voiced and unvoiced by calculating the energy of speech signal. The high frequencies of speech signal are suppressed during the human speech production mechanism. Therefore to compensate for these high frequencies the pre-emphasis process is applied to the speech signal by using the digital filter. A digital filter is used as pre-emphsizer to compensate these high frequencies; it's calculated as given in the following equation:

$$y(n) = s(n) - \alpha \times s(n-1), 0 \le \alpha \le 1 \qquad ...(1)$$

where $y(n)$ and $s(n)$ are the output and the input of the filter respectively. In this work the value of a is 0.97.

After pre-emphasis process the speech signal is segmented into group of frames, each frame contains N samples with stationary time duration. The time duration of each frame that is used in this work is equal to 30ms. Following that, the energy of each frame is computed as in following equation:

$$E_n = \sum_{n=1}^{N} [x(n)^2] \qquad ...(2)$$

where E_n is the energy of the frame, $x(n)$ is the frame and N is the number of samples in the frame, if the energy of the frame is larger than the threshold value (T) it is considered as voiced segment otherwise it is considered as unvoiced segment.

2.2. EXTRACTING FEATURES

Speech features have been selected to use as a watermark data because it unique for any individual instead of current digital watermarking techniques which use a digital sample such as a logo image, random numbers, or a digital signature as a watermark. The voiced speech signal that is obtained during speech segmentation process will be used in order to produce speech features. The voiced speech signal is segmented into group of frames; each frame contains N samples with stationary time duration. The time duration of each frame that is used in this work is equal to 30ms. Following that, the frames of speech signal are overlapped to produce new frames; each new frame takes the first segment (15ms) from the prior frame and the second segment (15ms) from the following frame. The process of overlapping is implemented on frames to avoid missing of information that happens during segmentation of speech signal into frames. Then the hamming window is applied to overlapped frames by multiplying the hamming window function with each frame to get best isolation for each frame to make the ability to find the finer features in each frame instead of the whole speech signal. It's calculated as given in the following equation:

$$W(n) = 0.54 - 0.46 \cos\left[\frac{2\pi \times n}{N-1}\right] \quad ...(3)$$

where

$0 \le n \le N-1$

After that, the windowed frames are used in features extraction process to extract a set of features from each frame by using Mel Frequency Cepstral Coefficients (MFCC) method. The MFCC is used as features extraction method because it simulates the human hearing perception. The MFCC features are calculated as in the following steps:

1. The magnitude spectrum vector of each windowed frame is calculated by applying the DFT to each windowed frame by using the following equation:

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{\frac{-j2\pi 2\pi}{N}} \quad ...(4)$$

2. Map the powers of magnitude spectrum that is obtained in the first step onto the Mel scale.

3. The Mel for a certain frequency is computed by applying the following equation:

$$Mel(f) = 2595 \times \log_{10}(1 + \frac{f}{700}) \quad ...(5)$$

4. Compute the Mel spectrum of the magnitude spectrum by multiplying the magnitude spectrum by each of the triangle Mel filter. The number of channels in triangle filter in this work is 20.

5. Take the logarithmic for each one of the Mel spectrum by applying the following equation:

$$E(n) = \log\left[\sum_{k=0}^{N-1} |x(k)|^2 H_n(k)\right], \quad 0 \le n \le N \quad ...(6)$$

6. Transform the output of the previous step into a set of cepestral coefficients by applying the DCT as in following equation:

$$y(k) = \alpha(k) \sum_{n=0}^{N-1} x(n) \cos\left[\frac{\pi(2n+1)k}{2N}\right] \quad ...(7)$$

Then, occupying only the first thirteen of cepestral coefficients which indicate the information of vocal tract and considered as a watermark. The speech features convert into bits stream as follows:

$$\bar{f} = \frac{f - f_{min}}{f_{max} - f_{min}} \quad ...(8)$$

where $\bar{f}$ represented new value of feature, f represented original value of feature, f_{max} is maximum value of features lies in it, and f_{min} is minimum value of features lies in it.

2.3. EMBEDDING WATERMARK

The speech features is embedded in the unvoiced segments of speech signal that were obtained during speech segmentation process. The unvoiced segments are used as embedding locations

because of that in biometric watermarking; the process of embedding watermark should not effect on the features which are utilized in identification systems, thus preventing watermarking of voiced segments which are used for recognition. In more details, the watermark information is embedded in certain locations of unvoiced segments; these locations are selected by using the GA. The GA is used to produce random locations as secret locations for embedding watermark. The steps for applying the GA with speech watermarking can be listed as follows:

1. The unvoiced samples are divided into *M* of blocks; size of each block is the same size of watermark information. Each block represents a solution for embedding the watermark.
2. Select the two blocks randomly.
3. Stop the training process for number of iterations.
4. Apply crossover to the selected blocks and replace with two blocks randomly.
5. Repeat from step 2 to step 5 until termination condition is achieved.

After finished of GA training, one of the blocks in the last iteration is used as secret locations in the embedding and extraction processes. The watermark information embedded in the block resulted from GA training by changing the specific bit of samples. The output of the embedding watermark process is watermarked speech signal.

2.4. EXTRACTING WATERMARK

The steps of extracting watermark process are the same steps of embedding watermark process. The steps of extracting watermark are performed as follows: First, detect the unvoiced segments in watermarked speech signal by using the short time energy technique as in speech segmentation. Second, using unvoiced segments to extract the watermark information from the embedding locations depending on the secret locations that generated by using the GA as in the embedding watermark.

3. EXPERIMENTS AND RESULTS

For evaluate of the results obtained by the proposed technique. We are use the speech signal files are wave format with 16 bit sample resolution, 8000 Hz frequency and mono channel [8]. The Figure 2 shows the results of speech segmentation into voiced/unvoiced signal.

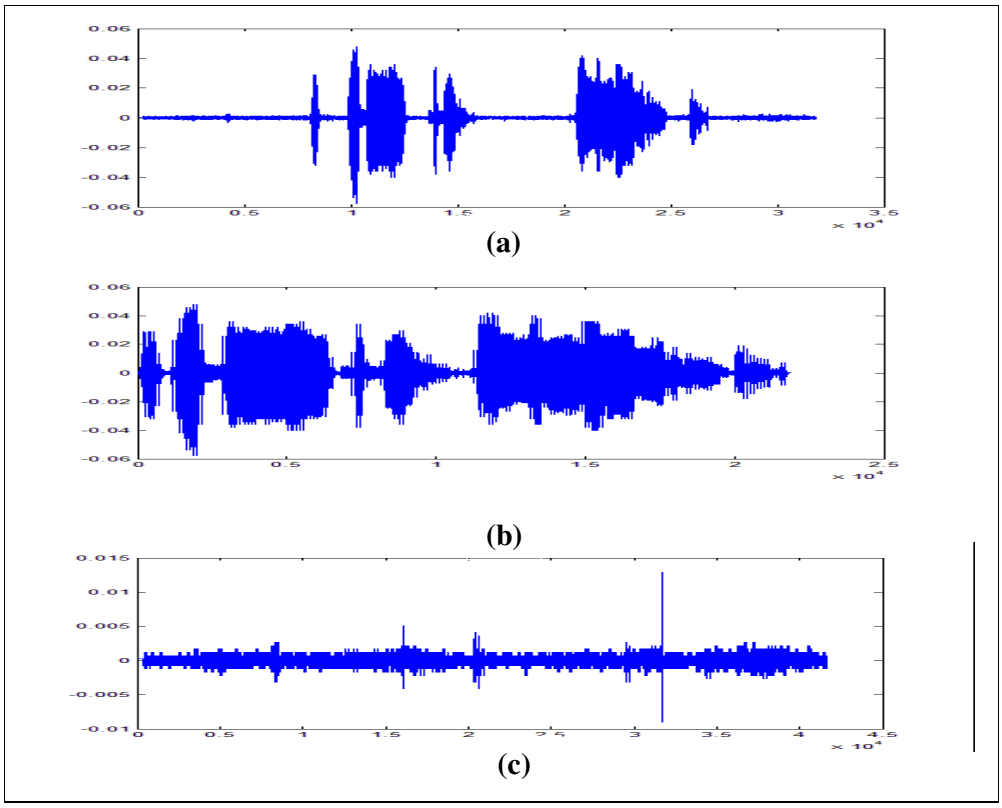

Figure 2. Segmentation of speech signal into voiced/unvoiced signal

As shown in Figure (2), (a) represents the original speech signal while (b) represents the voiced signal and (c) represents the unvoiced signal. To determine the amount of distortion on the host speech signal due to the proposed watermarking technique, the Signal-to-Noise Ratio (SNR) [9] of watermarked speech signals is calculated. The results of SNR are shown in Table 1.

Table 1. SNR values of watermarked speech signals.

Watermarked Speech Signal	Duration Time of Original Speech Signal	Size of Watermark	SNR
Sample 1	6 second	1000 byte	51.9936
Sample 2	12 second	2000 byte	54.0841
Sample 3	25 second	4000 byte	52.2715

For checking the watermarked speech signal is authentic or not authentic, this is done through determining the validity of extracted watermark information from watermarked speech signal. The extracted watermark information from watermarked speech signal compared with extracted features from watermarked speech signal by computes the Tamper Assessment Function (TAF) [10] between extracted watermark information and extracted features by apply the equation as follows:

$$TAF\,(EW\,,EF\,) = \frac{\sum_{i=1}^{NW} (EW\,)_i \oplus (EF\,)_i}{NW} \qquad ...(9)$$

Where

TAF: The floats ratio between 0 and 1.

EW: Extracted Watermark.
EF: Extracted Features.
NW: Number of bits.

If the TAF value is less from the threshold value (T=0.01), the speech signal considered as authentic otherwise the speech signal considered as not authentic. Table 2 show the TAF values for watermarked speech signal without attack and with attacked.

Table 2. Values of TAF.

Watermarked Speech Signal	TAF	Decision
No attack	0.0005<T	Authentic
Add noise	0.4884>T	Not-authentic
Re-sampling	0.5358>T	Not-authentic
Re-quantization	0.3562>T	Not-authentic
Filtering	0.4830>T	Not-authentic

4. CONCLUSIONS

This paper presented adaptive digital watermarking technique for ensuring the authentication of the speech signals. The benefits of the proposed technique are the watermark information is inserted in a way that the speech features that are used for speech recognition are not changed through the process of embedding watermark. Therefore, the speech features of watermarked speech signal are very similar to that with original speech signal. The watermark information can be extracted from the watermarked speech signal without needing the information of the original speech signal. This technique has a good quality of the watermarked speech signal although the quantity of watermark information is high.

REFERENCES

[1] Maryam Lafkih, Patrick Lacharmey, Cristophe Rosenbergery, Mounia Mikramz, Sanaa Ghouzalix, Mohammed El Haziti, Wadood Abdulk, Driss Aboutajdine, (2015) "Application of New Alteration Attack on Biometric Authentication Systems", First IEEE International Conference on Anti-Cybercrime (ICACC), Riyadh- KSA.

[2] Cameron Whitelam, Nnamdi Osia and Thirimachos Bourlai, (2013) "Securing Multimodal Biometric Data through Watermarking and Steganography", IEEE International Conference on Technologies for Homeland Security (HST), Waltham-MA.

[3] Nadhir Ben Halima, Mohammad Ayoub Khan, Rajiv Kumar, (2015) "A Novel Approach of Digital Image Watermarking using HDWT-DCT", IEEE Global Summit on Computer & Information Technology (GSCIT), Sousse- Tunisia.

[4] Lei Chen and Jiying Zhao, (2015) "Adaptive Digital Watermarking Using RDWT and SVD", 978-1-4673-9175-7/15/$31.00, IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE), Ottawa-ON.

[5] Rohit Thanki and Komal Borisagar, (2016) "Biometric Watermarking Technique Based on CS Theory and Fast Discrete Curvelet Transform for Face and Fingerprint Protection", Springer International Publishing Switzerland.

[6] Mohammad Ali Nematollahi, Hamurabi Gamboa-Rosales, Francisco J. Martinez-Ruiz, Jose I. De la Rosa-Vargas, S. A. R. Al-Haddad, Mansour Esmaeilpour, (2016) "Multi-factor authentication model based on multipurpose speech watermarking and online speaker recognition", Springer Science Business Media New York.

[7] Zhenghui Liu, FanZhang, JingWang, HongxiaWang, JiwuHuang, (2015) "Authentication and recovery algorithm for speech signal based on digital watermarking", Elsevier.

[8] Azarias Reda, Saurabh Panjwani and Edward Cutrell, (2011) "Hyke: A Low-cost Remote Attendance Tracking System for Developing Regions", The 5th ACM Workshop on Networked Systems for Developing Regions (NSDR), New York-USA.

[9] James M. Kates and Kathryn H. Arehart, (2013) "SNR is not enough: Noise modulation and speech quality", 978-1-4799-0356-6/13/$31.00, IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver-BC.

[10] Gaurav Gupta, Amit Mahesh Joshi and Kanika Sharma, (2015) "An Efficient Robust Image Watermarking based on AC Prediction Technique Using DCT Technique" ICTACT Journal on Image and Video Processing, Vol: 06, Issue: 01.