Speech biohashing authentication based on power spectrum

Yibo Huang1*, Tengfei Chen1, Xiangrong Pu1 and Qiuyu Zhang2
1College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou, Gansu, 730070, China
2School of Computer and Communication, Lanzhou University of Technology, Lanzhou, Gansu, 730070, China
*Corresponding author’s e-mail: huang_yibo@nwnu.edu.cn

Abstract. In order to solve hashing sequences leakage in the existing speech authentication algorithms, this paper proposed a speech biohashing authentication based on power spectrum. Firstly, the algorithm pre-processed the original speech, then used the improved covariance method to extract the power spectrum feature of the speech, and generated a one-dimensional feature vector from the feature matrix. A pseudo-random matrix was generated by using Sinc function transform, the feature vector and pseudo-random matrix are projected and mapped to construct the biometric template. Finally, binarization process was used to generate hash index, and upload it to the cloud. The experimental results show that the algorithm has good discrimination, robustness and authentication performance. At the same time, the biometric template has good security and ensures the irreversibility of the hashing sequence.

1. Introduction

With the rapid development of network technology, multimedia information ushered in explosive growth. As one of the forms of multimedia, speech has an important application in authentication, retrieval and identification, which has attracted widespread attention from scholars.

Hash technology has good discrimination and robustness, and is widely used in speech signal processing. Among them, perceptual hashing has shortcomings in terms of security and privacy, and quantum hashing has high complexity, low efficiency, and strict requirements on equipment. Therefore, speech authentication system based on biohashing came into being.

At present, the methods for extracting speech feature values of the biohashing authentication system have cochleagram [1], linear prediction residual coefficient [2], wavelet transform and so on. Ref. [3] proposed an extraction scheme based on biological hashing and spectral subtraction, which not only improves the performance of the algorithm, but also ensures the security of the system. Ref. [4] proposed an encrypted speech retrieval method based on syllable-level perceptual hash, which embeds perceptual hashing into speech as a digital watermark. In the retrieval scheme, there is no need to search or decrypt, but the robustness and security of the algorithm is weak. Ref. [5] proposed a robust audio hashing elastic mask method, which uses the elastic mask obtained when constructing the hashing database to improve the hash matching performance, but does not consider the security. Ref. [6] proposed an efficient retrieval algorithm under big data encryption. The algorithm uses Feistel encryption (FES) to generate keys and Advanced Encryption (AES) to encrypt plaintext data. This scheme has better security, but poor discrimination and robustness. Ref. [7] proposed a biometric template protection method. The biohashing algorithm reshapes the fingerprint, which can accurately judge the original fingerprint and the forged fingerprint, which proves that the biohashing has better protection and recognition.
capabilities. Ref. [8] proposed a retrieval algorithm based on fast Fourier transform and measurement matrix. The algorithm improves the security of the speech signal, but the robustness needs to be improved.

Aiming at the appeal problem, this paper proposed a speech biohashing authentication based on power spectrum, which can effectively ensure the security of the authentication system, and ensure the irreversibility of the biohashing sequence; at the same time, the algorithm has the advantages of good robustness and discrimination.

2. Related theories

2.1. Improved covariance method

The improved covariance method estimates the autoregressive model (AR) parameters by minimizing the mean of the forward and backward prediction error power estimates, and then estimates the power spectrum.

The AR parameter estimation can be written in matrix form

\[
\begin{bmatrix}
    c(1,0) \\
    \vdots \\
    c(p,0)
\end{bmatrix}
+ \begin{bmatrix}
    c(1,0) & \cdots & c(1,p) \\
    \vdots & \ddots & \vdots \\
    c(p,0) & \cdots & c(p,p)
\end{bmatrix}
\begin{bmatrix}
    \hat{a}(1) \\
    \vdots \\
    \hat{a}(p)
\end{bmatrix}
= \begin{bmatrix}
    0 \\
    \vdots \\
    0
\end{bmatrix}
\]

or

\[
c_p + C_p \hat{a} = 0
\] (2)

where

\[
c(j,k) = \frac{1}{2(N-p)} \left( \sum_{n=p}^{N-1} x^*(n-j)x(n-k) + \sum_{n=0}^{N-k-p} x(n+j)x^*(n+k) \right)
\] (3)

From Equation (2), the AR parameter is estimated as

\[
a = -C_p^{-1} c_p
\] (4)

The white noise variance is estimated as

\[
\hat{\sigma}^2 = c(0,0) + \sum_{k=1}^{p} \hat{a}(k)c(0,k)
\] (5)

It can be seen that the improved covariance is the same as the covariance except for the definition of the autocorrelation estimator \(c(j,k)\). From the estimation of AR parameter, the power spectral density (PSD) estimate is formed as

\[
\hat{P}_{MCOV}(f) = \frac{\hat{\sigma}^2}{\left| 1 + \sum_{k=1}^{p} \hat{a}(k)e^{-j2\pi f k} \right|^2}
\] (6)

2.2 Sinc function transform

To improve the security of the biohashing algorithm, the pseudo-random matrix generated by the Sinc function transform and the feature vector formed by the biometric feature are used to iteratively generate the authentication summary. This paper constructs pseudo-random matrix based on Sinc transform.

The one-dimensional Sinc function is the product of the sine function \(\sin(x)\) and the monotonically decreasing function \(1/x\). In digital signal processing, the normalized Sinc function is defined as:

\[
\text{Sin}(c(x)) = \frac{\sin(\pi x)}{\pi x}
\] (7)
Although the Sinc function is a simple one-dimensional model, it can generate complex motion trajectories. Its Sinc function transformation is defined as follows:

$$x_{i+1} = a \cdot \sin c(x_i) \quad (8)$$

Among them, when $a = 1$ and $x_i = 0.1$, a pseudo-random matrix with strong randomness and sensitivity to initial values can be generated.

3. Biohashing authentication algorithm

The biohashing authentication algorithm includes two stages: biohashing construction and biohashing matching. The flow chart of this algorithm is shown in Figure 1.

A Biohashing structure

Step 1. Pre-processing. Firstly, authenticated users provide original speech $s(n)$. After pre-emphasis, framing, and windowing, speech signal $s'(n)$ is obtained, and the number of frames is $N$.

Step 2. Feature extraction. The power spectrum feature of the speech signal $s'(n)$ are extracted, using the improved covariance method to obtain the coefficient matrix $H(u,v)$. Then the column vector of $H(u,v)$ is averaged to obtain the one-dimensional feature vector $W(l,v)$.

Step 3. Biometric template.

a. Set the key $a = 1$, the initial value $x_i = 0.1$, using the Sinc transform to generate a pseudo-random matrix $f(l,v)$, and then reorganize the rows and columns to obtain $f'(v,l)$. Finally, Schmidt orthogonalization is performed to obtain $f''(v,l)$.

b. Carry out the inner product of the matrix $W(l,v)$ and $f''(v,l)$ to obtain the matrix $\xi(v,l)$. Then $\xi(v,l)$ is processed for dimensionality reduction to obtain biological feature matrix $\xi'(l,v)$.

c. Binarizing biometric template $\xi'(l,v)$ to generate hash index $Z_i = \{Z_i(1,n)|v = 1,2,\cdots,N\}$, $N$ is the length of the hashing sequence in this paper.

Binarization processing method: Suppose $Z_i(1,1) = 0$, the $v$-th data of template $\xi'(l,v)$ is greater than the $(v-1)$-th data, then the $v$-th data of the hash sequence is 1, otherwise it is 0.

B. Biohashing matching: The authenticated user provides speech, constructs the hash index $Z_2$ to match with the hash index $Z_1$ in the cloud.
In the authentication process, Hamming distance \( DH(\cdot, \cdot) \) is used to measure whether the matching is successful, the calculation formula is

\[
DH(Z_1, Z_2) = \frac{1}{N} \sum_{i=1}^{N} |Z_1(i) \oplus Z_2(i)|
\]

(9)

where, and \( \oplus \) is the XOR logic operation. When \( DH(Z_1, Z_2) \leq \tau \), represent the match is successful; otherwise, the match is not successful, \( \tau \) is the perception authentication threshold.

In order to evaluate the performance of the algorithm, the False Accept Rate (FAR) is introduced:

\[
FAR(\tau) = \int_{-\infty}^{\tau} f(x | \mu, \sigma) = \int_{-\infty}^{\tau} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx
\]

(10)

where, \( \tau \) is the perception authentication threshold, \( \mu \) is the mean value BER, \( \sigma \) is the standard deviation of BER, and \( x \) is the error reception rate.

4. Experimental results and analysis

The experimental speech data comes from TIMIT (Texas Instruments and Massachusetts Institute of Technology) and TTS (Text to Speech) speech databases. There are 1200 different speech clips in the original speech database. The format of each speech clip is wav, the length is 4s, 16bit PCM, and the sampling frequency is 16khz. The experimental hardware platform is: Intel(R) Core(TM) i5-6500 CPU @ 3.20 GHz, 12GB of RAM. The software environment is Matlab R2018a under Win 10.

4.1 Discrimination analysis

Discrimination is used to distinguish the reliability of different or the same people in reading different speech content. Through the pairwise matching of hashing values of 1200 speech, \( 7.194 \times 10^7 \) data is obtained, as shown in Figure 2. The experimental results show that the BER value obtained by the experiment is roughly distributed normally with \( \mu = 0.5 \), which indicates that the BER of biohashing value of the proposed algorithm basically obeys the normal distribution.

![Figure 2. BER normal distribution](image)

Entropy rate (ER) is used to measure uncertainty of random events. In biohashing authentication experiment, ER is a comprehensive index to evaluate discrimination of biohashing. It overcomes shortcomings of algorithm vulnerable to sequence size. The calculation formula is as follows:

\[
ER = -[q \log_2 q + (1-q) \log_2 (1-q)]
\]

(11)

where, \( q \) is mean value of experiment(\( q = \frac{1}{2} \left( \sqrt{\frac{\sigma^2 - \sigma_1^2}{\sigma^2 + \sigma_1^2}} + 1 \right) \)), \( \sigma \) and \( \sigma_1 \) are standard deviations of BER theory and experiment.

The calculated ER value of this paper is 0.9878, which is very close to 1, which further shows that the proposed algorithm has better discrimination.
In order to further illustrate discrimination of the proposed algorithm, Equation (10) is used to calculate the FAR value. Table 1 shows comparison of FAR value between the proposed algorithm and the Refs. [9,10,11]. It can be seen that compared with other algorithms, the FAR value of the proposed algorithm is smaller, indicating that the proposed algorithm has better anti-collision ability.

Table 1. FAR of different algorithms

| τ (s) | Ref. [9]       | Ref. [10]      | Ref. [11]      | The proposed       |
|------|----------------|----------------|----------------|-------------------|
| 0.10 | 3.6×10⁻⁴²      | 3.1×10⁻³⁵      | 3.2×10⁻³⁰      | 3.2×10⁻¹⁰⁸        |
| 0.20 | 1.4×10⁻²⁴      | 1.6×10⁻²⁰      | 1.3×10⁻¹⁷      | 1.3×10⁻⁶¹         |
| 0.25 | 1.2×10⁻¹⁷      | 9.6×10⁻¹⁵      | 9.6×10⁻¹³      | 2.8×10⁻⁴³         |
| 0.30 | 6.1×10⁻¹²      | 5.3×10⁻¹⁰      | 1.1×10⁻⁸       | 2.6×10⁻⁲⁸         |
| 0.35 | 1.9×10⁻⁷       | 2.8×10⁻⁶       | 5.6×10⁻⁶       | 1.3×10⁻¹⁶         |

4.2 Robustness analysis

In order to evaluate the robustness of biohashing algorithm, 1200 speech in the database are subjected to content retention operations (CPOs), as shown in Table 2, and then the BER mean and max value after CPOs are calculated.

Table 2. The mean and maximum BER of the proposed algorithm

| CPOs                  | Operation method                                          | Mean   | Max    |
|-----------------------|----------------------------------------------------------|--------|--------|
| Volume Adjustment 1   | Volume up 50%                                             | 0.0061 | 0.0415 |
| Volume Adjustment 2   | Volume down 50%                                           | 0.0152 | 0.0887 |
| Resampling 1          | Sampling frequency decreased to 8 kHz, and then increased to 16 kHz | 0.0536 | 0.1353 |
| Resampling 2          | Sampling frequency decreased to 32 kHz, and then increased to 16 kHz | 0.0102 | 0.0556 |
| Echo Addition 1       | Delay 100 ms, initial strength were 10% of the echo        | 0.0751 | 0.1366 |
| Echo Addition 2       | Delay 300 ms, initial strength were 25% of the echo        | 0.1965 | 0.2695 |
| Narrowband Noise 1    | 30dB narrowband Gaussian noise, center frequency distribution in 0–4 kHz | 0.0993 | 0.2105 |
| Narrowband Noise 2    | 50dB narrowband Gaussian noise, center frequency distribution in 0–4 kHz | 0.0334 | 0.1404 |
| MP3 Compression 1     | Re-encoded as MP3, and then decoding recovery, the rate is 64k | 0.0238 | 0.0777 |
| MP3 Compression 2     | Re-encoded as MP3, and then decoding recovery, the rate is 128k | 0.0176 | 0.0802 |
| Low-pass Filtering 1  | 12 order FIR low-pass filtering, cut off frequency of 3.4 kHz | 0.0916 | 0.1679 |
| Low-pass Filtering 2  | 12 order Butterworth low-pass filtering, cut off frequency of 3.4 kHz | 0.0765 | 0.1278 |

It can be seen from Table 2 that the mean and maximum BER of the proposed algorithm is very small under the 12 kinds of content retention operations, and the maximum does not exceed 0.2695. Therefore, the proposed algorithm has good robustness and can meet the requirements of the biohashing authentication system.

In order to evaluate the overall discrimination and robustness of the proposed algorithm, draw the FAR-FRR curves as shown in Figure 3, and compare the proposed algorithm with other algorithms.

It can be seen from Figure 3 that the FAR-FRR curves of Ref. [10] intersect, which cannot effectively solve the discrimination and robustness. Although Ref. [12] does not cross, when τ = 0.3, the two operations cannot be distinguished. Although there is no intersection in Ref. [9], the judgment interval is smaller than the proposed algorithm. On the whole, the FAR-FRR curves of the proposed algorithm do not intersect, and the judgment interval is large, which can distinguish these two operations well, that is, the proposed algorithm has excellent discrimination and robustness.
4.3 Security Analysis
In order to ensure the security and privacy of the biohashing template, the proposed algorithm uses Sinc function transform to generate the pseudo-random matrix, and the feature matrix and the pseudo-random matrix are projected and mapped to construct the bio-security template.

In order to verify that the biohashing algorithm has the unidirectionality of the trapdoor, this paper randomly extracts 200 speech clips from the database, and calculates the Hamming distance between the features obtained by correct key and wrong key and the original feature. As shown in Figure 4.

It can be seen from Figure 4 that the Hamming distance range between the feature obtained and the original feature by the correct key is \((3 \times 10^{-18}, 7 \times 10^{-18})\), and the Hamming distance range between the feature obtained and the original feature by the wrong key is \((0.10, 0.20)\), indicating that the biohashing algorithm is unidirectionality of the trapdoor, and also proved the security of the biohashing algorithm.

5. Conclusion
This research proposed a speech biohashing authentication based on power spectrum. The algorithm uses the improved covariance method to extract the power spectrum as biological feature. Analyzing and comparing the performance of the algorithm such as discrimination, robustness and security.
Through experimental results, the following conclusions can be drawn: a. The proposed algorithm has good robustness to regular CPOs, especially in volume adjustment, MP3 compression, etc., which shows that the proposed algorithm can complete speech bio-hashing authentication well in different environments; b. The proposed algorithm has better discrimination and stronger abstract. There is a large judgment interval between the FAR-FRR curves, which shows that the proposed algorithm can effectively balance discrimination and robustness, and can meet the requirements of the biohashing authentication system; c. The proposed algorithm uses Sinc function transformation to construct a pseudo-random matrix, which effectively guarantees the integrity of the biohashing sequence, prevents the leakage of speech data, and improves the security of the biohashing authentication system. In the future work, the algorithm will be optimized, combined with deep hashing to construct an authentication scheme, and a more secure and efficient biohashing authentication system will be studied.

Acknowledgments
This work is supported by the National Natural Science Foundation of China(No.61862041), Youth Science and Technology Fund of Gansu Province of China(No.1606RJYA274).

References
[1] Chen N, Xiao H D, Zhu J. (2013) Robust audio hashing scheme based on cochleagram and cross recurrence analysis. Electronics Letters, 49(1): 7-8.
[2] Prathosh A P, Ananthapadmanabha T V, Ramakrishnan AG. (2013) Epoch Extraction Based on Integrated Linear Prediction Residual Using Plosion Index. IEEE Transactions on Audio Speech and Language Processing, 21(12): 2471-2480.
[3] Zhang Q Y, Li G L, Huang Y B. (2020) An efficient retrieval approach for encrypted speech based on biological hashing and spectral subtraction. Multimedia Tools and Applications, 79(39-40): 29775-29798.
[4] He S, Zhao H. (2017) A retrieval algorithm of encrypted speech based on syllable-level perceptual hashing. Computer Science and Information Systems, 14(3): 703-718.
[5] Jin S S. (2017) A Resilience Mask for Robust Audio Hashing. IEEE Transactions on Information and Systems, E100D(1): 57-60.
[6] Aljawarneh S, Yassein MB, Talafha WA. (2017) A resource-efficient encryption algorithm for multimedia big data. Multimedia Tools and Applications, 76(21): 22703-22724.
[7] Zheng Y, Cao Y, Chang C H. (2018) Facial biohashing based user-device physical unclonable function for bring your own device security. IEEE International Conference on Consumer Electronics (ICCE). Las Vegas. pp. 165-178.
[8] Zhang Q Y, Ge Z X, Zhou L. (2019) An efficient retrieval algorithm of encrypted speech based on inverse fast Fourier transform and measurement matrix. Turkish Journal of Electrical Engineering & Com- puter Sciences, 27(3): 1719-1736.
[9] Zhang, Q Y, Qiao S B, Huang Y B. (2018) A high-performance speech perceptual hashing authentication algorithm based on discrete wavelet transform and measurement matrix. Multimedia Tools and Applications, 77(16): 21653-21669.
[10] Li J F, Wang H X, Jing Y. (2015) Audio Perceptual Hashing Based on NMF and MDCT Coefficients. Chinese Journal of Electronics, 24(3): 579-588.
[11] Zhang Y B, Mi B Q, Zhou L. (2019) Speech Perceptual Hashing Algorithm Based on Short-term Auto-correlation for Speech Authentication. Radio Engineering, 49(10): 899-904.
[12] Zhang Q Y, Hu W J, Huang Y B. (2018) An efficient perceptual hashing based on improved spectral entropy for speech authentication. Multimedia Tools and Applications, 77(2): 1555-1581.