Multimedia Tools and Applications (2022) 81:2873–2899
https://doi.org/10.1007/s11042-021-11708-z

Long sequence biometric hashing authentication based on 2D-SIMM and CQCC cosine values

Yi-bo Huang[1] · Hexiang Hou[1] · Tengfei Chen[1] · Hao Li[1] · Qiu-yu Zhang[2]

Received: 18 November 2020 / Revised: 7 October 2021 / Accepted: 22 October 2021 /
Published online: 8 November 2021

Abstract

The existing speech authentication algorithms hash extracted speech features directly and saved them to the cloud, which is easy to cause speech feature leakage. In the process of constructing hashing, the utilization efficiency of speech feature is poor, and the short hashing sequence will lead to the lack of discrimination of hashing sequence and the deviation of authentication. In order to solve the above problems, a long sequence biometric hashing authentication algorithm based on two-dimensional Sine ICMI Cmodulation map (2D-SIMM) and constant Q cepstral coefficients (CQCC) cosine was proposed. First, this algorithm extracts the CQCC of the speech signal, then obtains the eigenvalue of the space cosine distance of the adjacent speech frame CQCC, and finally performs projection mapping between the eigenvalue and the pseudorandom matrix generated by 2D-SIMM to construct a biometric hashing sequence. This paper evaluates the proposed robust feature schemes of MFCC and CQCC space cosine distance through experiments. The experimental results show that CQCC spatial distance combined with 2D-SIMM biometrics characteristics can reach 10^{-21}. when the threshold is 0.35. The BER mean was only 0.0383 for maintaining the robustness of operation for different contents. When the SNR is -5 dB, the matching rate of different noises can reach 45%. At the same time, it also improves the security of the biological template, and the overall performance is greatly improved compared with the existing algorithm.

Keywords Speech content authentication · Biometric security template · 2D-SIMM · CQCC cosine · Discrimination · Unidirectional

✉ Yi-bo Huang

Hexiang Hou
17853487160@163.com

Hao Li
lhnwnu@163.com

Qiu-yu Zhang
zhangqylz@163.com

[1] College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou 730070, China

[2] School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China

1 Introduction

In recent years, the storage of unprotected biometric data poses a serious privacy threat. Due to the scarcity of personal biometrics, once lost, sensitive information about users will be exposed, leading to security risks [3, 6]. At present, Security vulnerability can be identified in the speech biometric authentication algorithms right from data capture up to data storage into the speech biometric database. At the same time, the hashing sequence constructed is relatively short, and the same hashing sequence may come from different user speech features. The low discrimination among users leads to high error rate and poor authentication effect. Therefore, the research on the security and differentiation of speech biometric content authentication becomes an important challenge.

Biometric authentication algorithms widely use biometric features such as human face [2, 14], palm print [9, 20, 31], fingerprint [1, 18], signature [10, 35], iris [11, 24], but rarely involve speech features. In recent years, speech perceptual hashing authentication algorithm can not only achieve good authentication results, but also resist noise interference during channel transmission, but the speech authentication algorithm lacks security. Due to the computational efficiency and security of biometric hashing, it is widely used to protect privacy of biometric features [21, 32]. Therefore, the combination of speech perceptual hashing and biometric hashing can not only improve the authentication effect, but also ensure the security of speech features. The most widely used speech signal features include short-term cross-correlation [26], short-term zero-crossing rate [38], Mel-frequency cepstral coefficient (MFCC) [16], Linear prediction cepstrum coefficient (LPCC) [12], Modified discrete cosine transform (MDCT) [17], discrete wavelet transform (DWT) [27], spectral entropy [25], measurement matrix [28], Modulated complex lapped transform (MCLT) [19], spectrogram [13, 29] and Multiple fusion features. Li et al. [17] used non-negative matrix factorization (NMF) to obtain the local characteristics of MDCT coefficients, and then used the mean to construct a binary hash sequence. This algorithm has good robustness to various content preserving operations, but its efficiency is low and it lacks security. Zhang et al. [27] simply compared the influence of different lengths of hashing sequences on the discrimination of the algorithm, but the algorithm only adopted 250 bits of hashing sequence length. Although the algorithm has a strong summarization, but its discrimination needs to be improved, there is no further study of the features of hashing long sequence. In Ref. [25], a feature fusion method for linear prediction of minimum mean square error and improved spectral entropy was proposed, and the constructed hashing sequence was only 266 bits. The algorithm has poor discrimination and MP3 compression robustness, but it has high efficiency. Zhang et al. [28] used the measurement matrix after chaotic processing to reduce the dimension of the discrete wavelet coefficient matrix, and then constructed a hashing sequence of 360 bits in length. Although the discrimination of the algorithm has been improved, the improvement is small. The algorithm has some security, but it is different from the biometric hashing algorithm, and its comprehensive performance needs to be improved. Therefore, the discrimination of the algorithm can be increased by increasing the length of the hashing sequence.

Sonnleitner et al. [33] used a two-dimensional filter to extract the local peaks in the spectrogram, and constructed the four-axis spatial features into a translational and scale-invariant hashing sequence. The algorithm is robust to noise and severe time-frequency scale distortion. In [30], the spectral matrix is trimmed by using a threshold based on the average value of the spectral values to generate different versions of the audio signal spectral matrix, which can achieve better robustness against noise interference. Jiang et al. [15] proposes an audio fingerprinting algorithm based on the second-generation wavelet packet and improved optimal-basis search algorithm. Although the algorithm not only robust for the audio which is handled by some kinds of method but also has good distinguishability between different audio. But characteristic of the audio data which reflect by the audio with this algorithm is segmentary.The application of it has limitations.

Khurshid et al. [18] adopted a fingerprint feature vector transformation method based on block hashing (BBH), using the average value of the geometric feature vector of the hand to transform each feature vector. This solution has better performance and higher security for feature templates. In [23], the random standard orthogonal projection technology is used to reduce the computational complexity while ensuring accuracy, and the fuzzy commitment protocol is used to ensure the security of the biometric template. It has a higher authentication rate, but the algorithm complexity is higher and the efficiency is slower. Chen et al. [7] proposed a biometric hashing scheme based on deep security quantification (DSQ), which has a good balance between security and practicality. But it requires a relatively long hashing sequence. When a new user joins, the DSQ neural network needs to be retrained, which reduces the flexibility of the framework.

In order to obtain higher template security and matching performance at the same time, this paper proposes a long-sequence biometric hashing authentication algorithm based on 2D-SIMM and CQCC cosine. In this paper, a long hashing sequence is used to improve the collision resistance performance of the algorithm. The extracted frequency domain spatial distance feature has strong robustness. The pseudorandom matrix generated by 2D-SIMM can ensure the security of biometric features. The irreversibility of biometric hashing sequences.

The remaining part of this paper is organized as follows. Section 2 describes related theory introduction. Section 3 introduces the details of the proposed algorithm. Section 4 gives the experimental results and performance analysis as compared with other related methods. Finally, Section 5 summarizes the thesis and prospects for future work.

2 Related theory introduction

2.1 MFCC

Feature extraction technology based on MFCC is used to capture the most important features of speech, which is very close to the response of human auditory system [4, 8]. Figure 1 is a block diagram of MFCC calculation. The MFCC calculation steps are as follows:

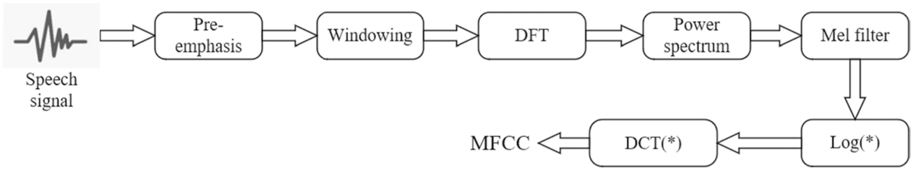


Fig. 1 Block diagram of MFCC feature extraction

2.1.1 Pre-processing

Firstly, pre-emphasize the speech signal $s(n)$ and amplify the high frequency components, which can effectively suppress random noise. Then the signal $x(n)$ is divided into frames using the Hamming window function, and features are extracted in short frames. The Hamming window function can effectively overcome the leakage phenomenon. The speech signal is divided into N frames, each frame has M samples.

$$w(m) = (1 - \alpha) - \alpha cos(2\pi m/(M - 1)) \tag{1}$$

$$x(n, m) = x(n)w(n - m\beta) \tag{2}$$

where, α set parameters for Hamming Window. β represents the time step of the shift in the speech signal. $x(n)$ is the pre-weighted speech signal. n is the index number of the speech signal frame, $n \in [1, 2, \cdots, N]$. m is the index number of each frame, $m \in [1, 2, \cdots, M]$.

2.1.2 Discrete Fourier transform

The speech time domain signal is transformed into the frequency domain signal through the discrete Fourier transform (DFT) to obtain the spectrum $X(m, k)$.

$$X^{DFT}(n, k) = \sum_{m=1}^{M} x(n, m)e^{-j2\pi mk/M} \tag{3}$$

where, $k, m \in [1, 2, \cdots, M]$, $X(n,\ m)$ is the time domain signal after pre-processing, $j = \sqrt{-1}$.

2.1.3 Power spectrum

Take the square of the modulus of the signal spectrum $X^{DFT}(n, k)$ to obtain the power spectrum $P(n, k)$.

$$P(n, k) = \frac{1}{M}|X^{DFT}(n, k)|^2 \tag{4}$$

2.1.4 Mel frequency filter

The Mel scale is linear for frequencies below 1000Hz and logarithm for frequencies above 1000Hz.The Power spectrum $P(n, k)$ is obtained by means of a set of Mel scale triangular filter Banks. At each frequency, the product of $P(n, k)$ and filter $H_l(k)$ is calculated.

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

$$E(n, l) = \sum_{k=0}^{M-1} P(n, k)H_l(k) \tag{6}$$

where, $Mel(f)$ is the Mel frequency scale, $H_l(k)$ is the transfer function $(l = 1, 2, \cdots, L)$ of l filter, and L is the number of filters.

2.1.5 Discrete cosine transform

Logarithmic computation is performed on the Mel spectrum $E(n, l)$ obtained, which is usually used to reflect logarithmic compression of human hearing. The final step is to convert the spectrum value of the L logarithmic filter bank to the I cepstrum coefficient using the discrete cosine transform (DCT).

$$MF_n(i)=\sum_{l=1}^{L}\log(E(n,l))cos\left[\frac{i(l-0.5)\pi}{L}\right] \tag{7}$$

where, i $(i = 1, 2, \cdots, L)$ is the MFCC of each frame of speech signal.

2.2 CQCC

Constant Q Cepstral coefficients (CQCC) were recently introduced into the deception detection of ASV [37]. CQCC was extracted by combining constant Q transform and cepstrum analysis. CQCC is based on constant Q transform and adopts variable time-frequency resolution [34]. Compared with DFT, CQT frequency resolution is higher at lower frequencies and time resolution is higher at higher frequencies. Therefore, CQCC tend to capture more spectral details at lower frequencies and more temporal details at higher frequencies, which are usually lost through more traditional time-frequency analysis methods. Figure 2 is the block diagram of CQCC calculation. The calculation steps of CQCC are as follows:

2.2.1 Constant Q transform

Speech signal $s(n)$ is pre-processed to get $x(n, m)$ according to **Pre-processing**. Let $x(m)$ be the data value of each frame, and the following equation is the constant Q transformation of $x(m)$.

$$X^{CQT}(m,k)=\sum_{j=m-\lfloor\frac{N_k}{2}\rfloor}^{m+\lfloor\frac{N_k}{2}\rfloor} x(j)a_k^*(j-m+\frac{N_k}{2}) \tag{8}$$

where, m is the sample index. $k = 1, 2, \cdots, K$ is the frequency bin index. N_k are the length of the variable window. $\lfloor\bullet\rfloor$ denotes the sign of rounding down. The basis function $a_k(m)$ is the complex time-frequency atom, $*$ is the complex conjugate.

$$a_k(m) = g_k(m)e^{i(2\pi m\frac{f_k}{f_s}+\Phi_k)} \tag{9}$$

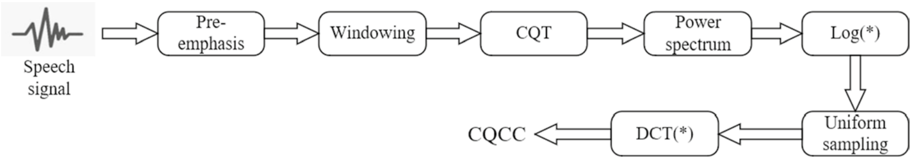


Fig. 2 Block diagram of CQCC feature extraction

where, f_k is the center of k frequency bin. f_s is the sampling rate. $g_k(m)$ is zero-centred window function. Φ_k is a phase offset. Due to the need to adjust the scale between frequency storeys, the central frequency f_k of frequency storeys k is defined as: $f_k = f_1 2^{(k-1)/B}$. f_1 is the center frequency of the lowest-frequency bin. B is the number of bins per octave.

Q is a filter selection metric that reflects the ratio between the center frequency and the bandwidth. For the CQT transform, Q is constant for all frequency bins k; filters are logarithmically spaced.

$$Q = \frac{f_k}{\delta_f} = (2^{1/B} - 1)^{-1} \tag{10}$$

$$N_k = \frac{f_s}{f_k} Q \tag{11}$$

where, δ_f is the bandwidth.

2.2.2 Power spectrum and uniform sampling

The spectrum of each frame is summed to obtain the spectral signal $X^{CQT}(n, k)$ of the speech. The power spectrum $R(n, k)$ is obtained by taking the signal spectrum $X^{CQT}(n, k)$ and squaring its mode. Then take the logarithm of the power spectrum to get $\log(R(n, k))$. The logarithmic power spectrum is uniformly resampled to realize the conversion from non-linear octave to the linear scale.

2.2.3 Discrete cosine transform

$\log(R(n, l))$ performed DCT to obtain CQCC.

$$CQ_n(p) = \sum_{l=1}^{L} \log(R(n, l)) cos\left[\frac{p(l - 0.5)\pi}{L}\right] \tag{12}$$

where, $p(p = 1, 2, \cdots, L)$ is the CQCC of each frame of speech signal.

2.3 Cosine similarity theorem

When cosine similarity measures the similarity of two directions in a space, when the angle between two vectors in the space is smaller, the cosine value will be closer to 1, which proves that the similarity between the two vectors is higher [5, 36].

For example, in three dimensions there are three points, R, Q and P. Where RO, QO and PO are vectors in three different directions in the space, and the included Angle between the three vectors is marked in Fig. 3. The cosine similarity theorem requires you to compute the cosine of the Angle between vectors. The smaller the angle, the bigger the value, the closer it is to 1, which means the vectors are going in the same direction. For two vectors x and y, the cosine similarity between x and y is:

$$\cos(\varphi) = \frac{x \cdot y}{\|x\| \|y\|} = \frac{x_1 y_1 + x_2 y_2 + \cdots + x_n y_n}{\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \sqrt{y_1^2 + y_2^2 + \cdots + y_n^2}} \tag{13}$$

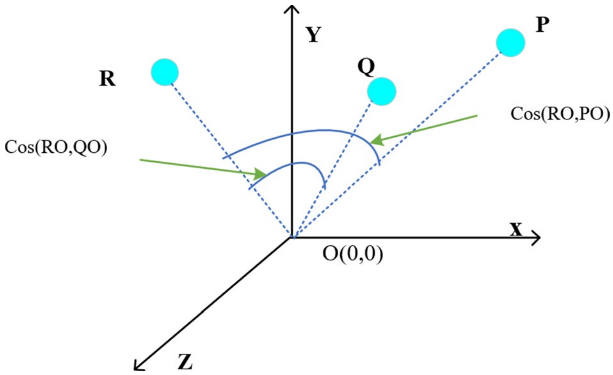


Fig. 3 The cosine of the space between two points

In this paper, each frame of speech signal is regarded as a vector in high-dimensional space, the coefficients of each frame number MFCC and CQCC represent the vector, and the included angle cosine of the adjacent two frames are calculated as biometric.

2.4 2D-SIMM

In order to improve the security of biometrics, a pseudorandom matrix was constructed by using two dimensional sinusoidal ICMIC modulation mapping (2D-SIMM). The mapping has good ergodicity, hyperchaotic behavior, high security and low time complexity. 2D-SIMM is defined as:

$$x_{n+1} = a\sin(\pi y_n)\sin(b/kx_n) \tag{14}$$

$$y_{n+1} = a\sin(\pi x_{n+1})\sin(b/y_n) \tag{15}$$

where, n is the length of the set matrix, corresponding to the number of frames of speech biometric. x_0, y_0 are the initial value set for x_n, y_n. a, b are the parameters set for the system, while $a, b \in (0, +\infty)$. In Ref. [22], the matrix can be completely chaotic. k is the parameter added on the original model to further improve the security of the biometric template.

3 Biometric hashing authentication scheme

The block diagram of the long sequence biometric hashing authentication algorithm based on 2D-SIMM and CQCC cosine values proposed in this paper is shown in Fig. 4.

3.1 Registration phase

Registered users feature the original voice, then construct a biometric security template, and finally store the binary hashing sequence to the cloud.

Step 1: Pre-processing Pre-processing includes pre-emphasis, framing and windowing. The speech signal $x(n)$ is obtained by pre-emphasis the input signal $s(n)$. Then the pre-emphasized speech signal is framed and windowed, in which the Hamming window is used to smooth the frame edges. The speech $x(n)$ is divided into N frame, and signal $x(n, m)$ is obtained, where

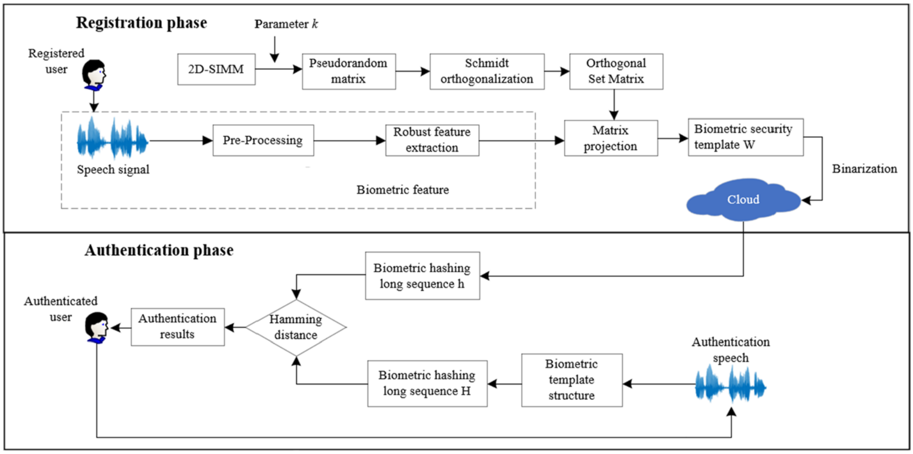


Fig. 4 Block diagram of the proposed long sequence biometric hashing algorithm

$n(n = 1, 2, \cdots, N)$ is the index number of the speech frame, $m(m = 1, 2, \cdots, M)$ is the index number of a frame of signal data.

Step 2: Robust feature extraction

1. MFCC feature extraction

DFT converts frequency domain signals into frequency domain signals to obtain frequency domain signal $X^{DFT}(n, k_1)$ $(n = 1, 2, \cdots, N; k_1 = 1, 2, \cdots, M)$. Then obtain the power spectrum and the Mel filter transformation to obtain the Mel spectrum $P(n, l_1)$ $(n = 1, 2, \cdots, N; l_1 = 1, 2, \cdots, L_1)$. Finally, take the logarithm and DCT transform to get the Mel cepstrum coefficient $MF(n, i)$ $(n = 1, 2, \cdots, N; i = 1, 2, \cdots, L_1)$.

2. CQCC feature extraction

CQT also transforms the time domain signal to get the spectrum signal $X^{CQT}(n, k_2)$ $(n = 1, 2, \cdots, N; k_2 = 1, 2, \cdots, K)$. Then obtain the rate spectrum, take the logarithm and uniform sampling, and get the transformed feature $R(n, l_2)$ $(n = 1, 2, \cdots, N; l_2 = 1, 2, \cdots, L_2)$. Finally, use the same method as MFCC to perform DCT transformation to obtain constant Q cepstrum coefficient $CQ(n, j)$ $(n = 1, 2, \cdots, N; j = 1, 2, \cdots, L_2)$.

This paper uses 16 Mel filters to obtain MFCC features, among which $L_1 = 16$. In calculating CQCC, after CQT transformation, the value of K is 8; then, equal interval interpolation sampling is performed, and L_2 is 16.

3. Cosine of adjacent speech frame space is obtained

The extracted MFCC and CQCC eigenvalues are uniformly set as $MQ(n, i)(n = 1, 2, \cdots, N; i = 1, 2, \cdots, L)$, where $L = L_1 = L_2$. The row vectors of the eigenvalues are calculated as $MQ_1(i)$, and then the matrix is spliced to obtain the matrix $\Lambda_1 = [MQ_1, MQ]$, $\Lambda_2 = [MQ, MQ_1]$. Take the cosine of each column of the two matrices and get the final eigenvector $F(n)(n = 1, 2, \cdots, N_p)$.

$$F(n) = \frac{\Lambda_1(n) \bullet \Lambda_2(n)}{\|\Lambda_1(n)\| \|\Lambda_2(n)\|} \tag{16}$$

Step 3: Construct an orthogonal set matrix A pseudorandom matrix is generated by 2D-SIMM. This article sets $a = 1; b = 5$, sets the key k, and the length of the matrix is consistent with the length of the biometric matrix. The initial value is randomly selected and set $x_0 = 0.2; y_0 = 0.3$ to obtain a pseudorandom matrix $v(n, t)(n = 1, 2, \cdots, N_p; t = 1, 2)$. Schmidt orthogonalization is performed on the pseudorandom matrix to obtain an orthogonal matrix $V(n, t)$.

Step 4: Biometric security template construction Extract the row vector of the positive set intersection matrix as $V_1(n)(n = 1, 2, \cdots, N_p)$. The biometric feature $F(n)$ is multiplied by the orthogonal row vector $V_1(n)$ to obtain the square matrix $\Psi(n, n)$.

$$\Psi(n,n) = F(n) \bullet V_1(n) = \begin{bmatrix} \Psi(1,1) & \Psi(1,2) & \cdots & \Psi(1,N_p) \\ \Psi(2,1) & \Psi(2,2) & \cdots & \Psi(2,N_p) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi(N_p,1) & \Psi(N_p,2) & \cdots & \Psi(N_p,N_p) \end{bmatrix} \tag{17}$$

In order to further increase the security of the biometric template, the square matrix $\Psi(n, n)$ is shifted chaotically, and the rows and columns are shifted cyclically in a ring form. In order to reduce the complexity of the algorithm and improve the efficiency, the rows and columns of the square matrix are moved by $0.5N$ positions, and the encrypted square matrix $\Psi^*(n, n)$ is obtained at this time. The row vector $V_2(n)(n = 1, 2, \cdots, N_p)$ in the orthogonal set matrix is projected to reduce the dimensionality of the square matrix $\Psi^*(n, n)$ to obtain the biometric security template W.

$$W(n) = V_2(n) \bullet \Psi^*(n,n) = [W(1), W(2), \cdots, W(N_p)] \tag{18}$$

Step 5: Biometric hashing construction Binarize the biometric security template W to generate a one-dimensional binary hashing length sequence h. Then store the long biometric hashing sequence in the cloud to complete the registration phase.

$$h(n) = \begin{cases} 1, \ if\ W(n) > W(n-1) \\ 0, \qquad Otherwise \end{cases} \tag{19}$$

where, $h(1) = 0$. $h(n)(n = 2, \cdots, N_p)$ is the perceived hashing value of each frame speech signal. Therefore, the length of the hashing sequence in this article is N_p bits.

3.2 Authentication phase

The authenticated user provides speech, constructs a long biometric hashing sequence, performs matching authentication with the biometric hashing sequence in the cloud, and feeds the result back to the authenticated user.

Step 1: The authenticated user provides an authentication speech, and passes the speech through Steps 1-5 in the registration phase to obtain a biometric hashing length sequence $H(n)$.

Step 2: Calculate the bit error rate (BER) of the two sequences through the Hamming distance between the biometric hashing long sequence $H(n)$ obtained by the authentication speech and the biometric hashing long sequence y in the cloud.

$$BER(h,H) = \frac{\sum_{n=1}^{N_p} h(n) \oplus H(n)}{N_p} \tag{20}$$

where, $\oplus$ is the XOR logic operation, and N_p is the length of the biometric hashing sequence. In this paper, BER hypothesis testing is used to describe hashing matching.

T_0: If the h and H of two speech clips have the same content:

$$BER(h, H) \leq \tau \tag{21}$$

T_1: If the h and H of two speech clips have the different content:

$$BER(h, H) > \tau \tag{22}$$

where, τ represents the authentication threshold. By comparing the size between the BER and the set threshold τ, biometric authentication is achieved. If BER is less than the threshold τ, then biometric features are the same and the authentication is passed, otherwise the authentication fails.

Step 3: Feedback the result of the authentication to the authenticated user. This paper uses the False Accept Rate (FAR) and False Reject Rate (FRR) to evaluate the performance of the algorithm. Among them, FAR is used to evaluate the discrimination of the algorithm, and FRR is used to evaluate the robustness of the algorithm.

$$FAR(\tau) = \int_{-\infty}^{\tau} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \tag{23}$$

$$FRR(\tau) = 1 - \int_{-\infty}^{\tau} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \tag{24}$$

where, μ is the expected value, σ is the standard deviation. The smaller the value of FAR and FRR, the better the discrimination and robustness of the algorithm.

4 Experimental result and analysis

The operating experimental hardware platform is Intel(R) Core(TM) i5-7500 CPU, 3.40 GHz, with computer memories of 4G. The operating software environment is MATLAB R2018b of Windows 7 system.

In this study, after lots of experiments, we found that the following parameters are given the best results after applying it to the proposed algorithm: $M = 200; N = 1064; N_p = 1065; K = 8; B = 8; L = L_1 = L_2 = 16; k = 0.5$. Where: M is the length of a frame of speech signal; N is the number of frames after speech framing; N_p is the length of the hashing sequence; K is the number of frequency bands after CQT transformation; B is the number of frequency bins per octave; $L = L_1 = L_2$ is the number of MFCC and CQCC features in each frame; k is the secret key in 2D-SIMM.

4.1 Speech database

The experimental speech datas comes from TIMIT (Texas Instruments and Massachusetts Institute of Technology) speech database and TTS (Text to Speech) speech database. This paper uses 1,200 different speech clips for experiments, each of which has a duration of 4s, a format of WAV, and a sampling frequency of 16kHz.

In order to imitate the interference of the channel transmission environment on the speech signal, the content preserving operation is performed on 1200 speech clips.

The content preserving operation includes volume, resampling, noise, echo and mp3 compression.

In order to detect the effect of the algorithm of this paper under various background noises, the Noise-92 noise database was introduced, and 86,400 noisy speech clips were established. Noise includes 8 different formats such as Pink noise and Gnoisegen noise. The SNR (signal noise ratio) ranges from -10db to 30db, and the interval is 5db.

4.2 Discrimination test and analysis

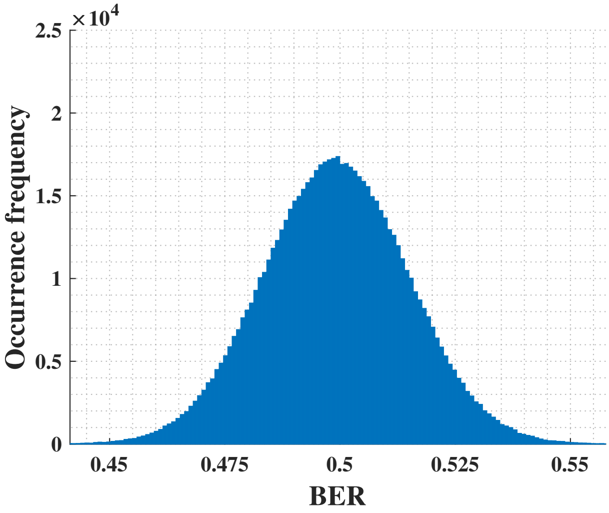
The BER of the biometric hashing value of different speech content basically obeys the normal distribution. There are 1200 different speech clips, using binomial coefficients to calculate the number of all available BER values as $1200 \times 1199/2 = 719400$. Figure 5 shows the BER histogram of a speech clips that matches the other 1199 speech clips. Figure 6 shows the normal distribution of BER of different biometric content hash sequences.

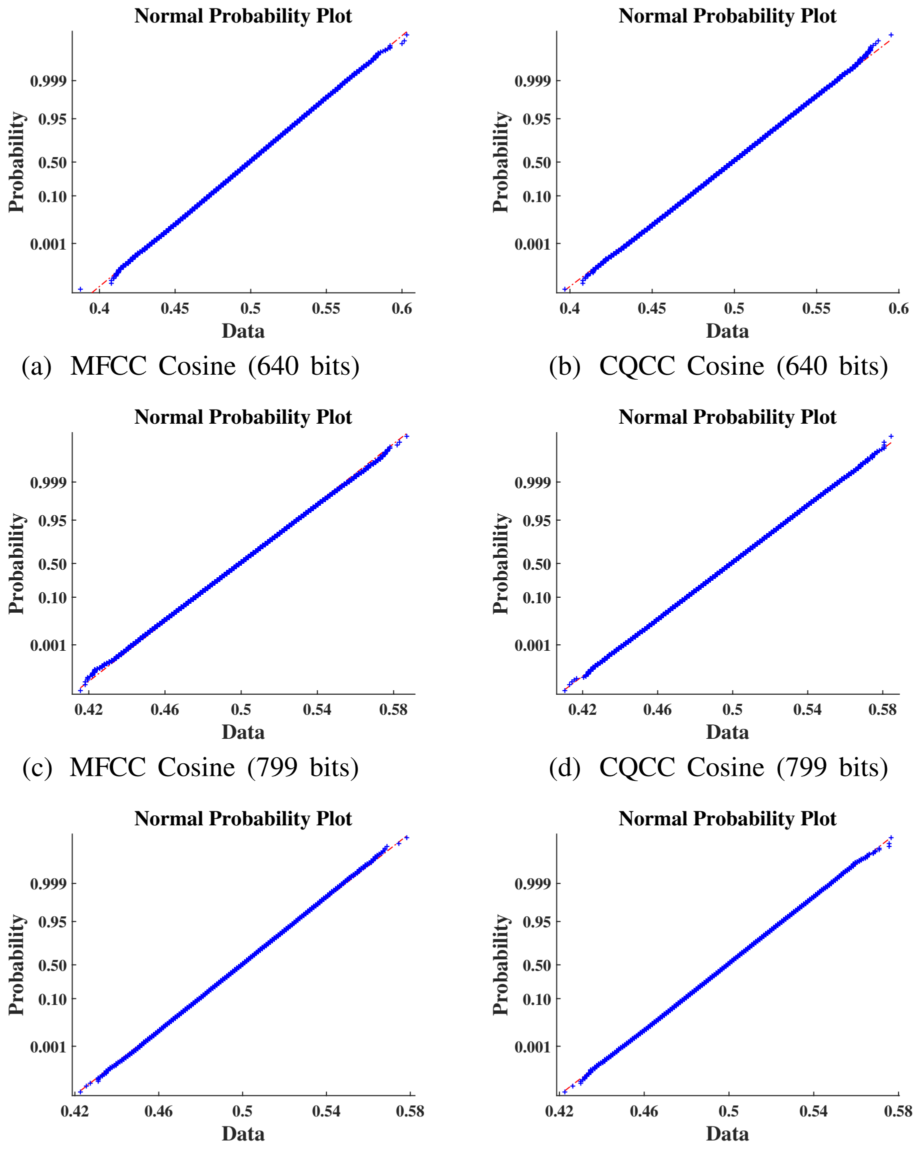
As shown in Fig. 6, the BER value probability of different speeches has a higher degree of coincidence with the probability curve of the standard normal distribution. With the increase of the hashing sequence, the BER range is closer to 0.5, and the value of the distribution is closer to the theoretical value. Compared with 640 bits and 799 bits, the sequence length 1065 bits selected in this article is smaller in BER range and has the best effect. Compared with the MFCC cosine algorithm, the CQCC cosine algorithm has smaller fluctuations in the actual value, and the effect is better.

According to the De Moivre-Laplace central limit theorem, the hamming distance is approximate obeying normal distribution ($\mu = 0.5$, $\sigma = \sqrt{0.25/N_p}$). The length N_p of the biometric hashing sequence in this paper is 1065 bits, and the theoretical value of the standard deviation is $\sigma = 0.0153$. Table 1 shows the theoretical and experimental values obtained from BER. Figure 7 shows the FAR curves of robust features with different sequence lengths.

As shown in Table 1 and Fig. 7, as the length of the hashing sequence increases, the actual value of the algorithm is closer to the theoretical value, the actual curve is getting

Fig. 5 BER histogram





(a) MFCC Cosine (640 bits)

(b) CQCC Cosine (640 bits)

(c) MFCC Cosine (799 bits)

(d) CQCC Cosine (799 bits)

(e) MFCC Cosine (1065 bits)

(f) CQCC Cosine (1065 bits)

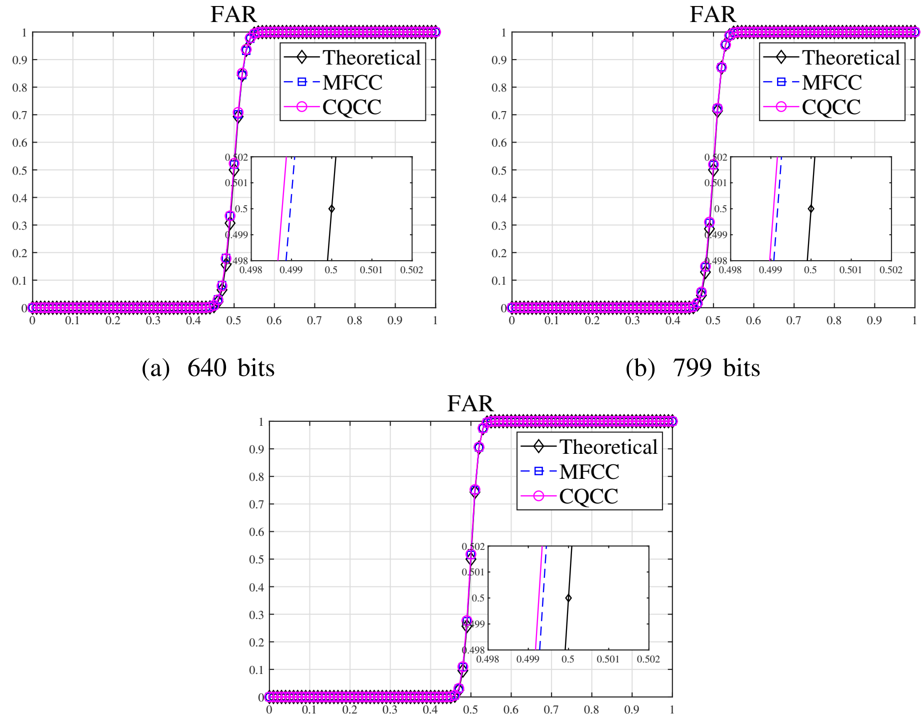
Fig. 6 BER normal distribution with different robust features and different hashing sequence lengths

closer and closer to the theoretical curve. It shows that the algorithm has good randomness and collision resistance. The difference between the CQCC cosine value and the actual value of the MFCC cosine value curve is small, and both are close to the theoretical curve, indicating that the two algorithms have good discrimination.

Tables 2 and 3 compares FAR of different long hashing sequence algorithms and different algorithms. As can be seen in Table 2: when the threshold is the same, as the hash

Table 1 Normal distribution parameters with different robust features and different hashing sequence lengths

Parameter	Hashing sequence length	Theoretical value	Actual value (MFCC)	Actual value (CQCC)
μ	640 bits	0.5000	0.4990	0.4988
μ	799 bits	0.5000	0.4992	0.4991
μ	1065 bits	0.5000	0.4994	0.4993
σ	640 bits	0.0198	0.0208	0.0204
σ	799 bits	0.0177	0.0183	0.0184
σ	1065 bits	0.0153	0.0157	0.0157



(a) 640 bits

(b) 799 bits

(c) 1065 bits

Fig. 7 Robust features FAR Curves with different hashing sequence lengths

sequence increases, the FAR value will be smaller and the algorithm's discrimination will be better. In the case of different thresholds, the difference in FAR values between the two methods is small, which proves that the discrimination effects of the two methods are similar. As can be seen in Table 3: Compared with Refs. [17, 25, 28, 38], the FAR values of the proposed method are all the lowest, and they are all lower than the third power of other algorithms. As for the short hashing sequence used in Refs. [17, 25, 28, 38], the long hashing sequence of the proposed method has a greater advantage in discrimination, and it also proves that the long hashing sequence has strong discrimination.

Table 2 The FAR value of different hashing sequence lengths and different robust features

τ	640 bits	799 bits	1065 bits
Algorithm	MFCC cosine		
0.10	1.8876×10^{-82}	3.5113×10^{-106}	8.1612×10^{-143}
0.20	3.0935×10^{-47}	1.3340×10^{-60}	3.0884×10^{-81}
0.25	2.2082×10^{-33}	1.1216×10^{-42}	5.0071×10^{-57}
0.30	5.0284×10^{-22}	5.4620×10^{-28}	3.3864×10^{-37}
0.35	3.7320×10^{-13}	1.5750×10^{-16}	9.7708×10^{-22}
Algorithm	CQCC cosine		
0.10	9.3958×10^{-86}	7.3441×10^{-104}	1.9846×10^{-142}
0.20	4.4688×10^{-49}	2.7386×10^{-59}	5.5650×10^{-81}
0.25	1.1897×10^{-34}	9.2351×10^{-42}	7.9201×10^{-57}
0.30	7.9063×10^{-23}	2.1269×10^{-27}	4.7640×10^{-37}
0.35	1.3405×10^{-13}	3.4196×10^{-16}	1.2387×10^{-21}

Table 3 The FAR value of different algorithms

τ	Algorithm				
	The Algorithm	Ref. [38]	Ref. [17]	Ref. [25]	Ref. [28]
0.10	1.9846×10^{-142}	7.8834×10^{-50}	1.4855×10^{-43}	1.7668×10^{-28}	3.7200×10^{-43}
0.20	5.5650×10^{-81}	1.2031×10^{-28}	2.5886×10^{-25}	1.9604×10^{-16}	4.3453×10^{-25}
0.25	7.9201×10^{-57}	2.4792×10^{-20}	3.9877×10^{-18}	9.8689×10^{-12}	5.7209×10^{-18}
0.30	4.7640×10^{-37}	1.5705×10^{-13}	3.2038×10^{-12}	6.6409×10^{-08}	4.0419×10^{-12}
0.35	1.2387×10^{-21}	3.1249×10^{-08}	1.3692×10^{-07}	6.1048×10^{-05}	1.5630×10^{-07}

Entropy rate (ER) mainly compares the comprehensive performance of hash algorithms and, unlike other calculation parameters, ER is not affected by the length of the hash sequence. The range of ER value is (0,1). The closer the value is to 1, the better the discrimination.

$$ER = -[q\log_2 q + (1-q)\log_2(1-q)] \tag{25}$$

$$q = \frac{1}{2}\left(\sqrt{\frac{\left|\sigma^2 - {\sigma_1}^2\right|}{\sigma^2 + {\sigma_1}^2}} + 1\right) \tag{26}$$

where σ and σ_1 are theoretical and experimental standard deviation of BERs respectively.

According to the results in Tables 4 and 5, the hash long sequence used in this paper has a high ER value. Comparing different algorithms, the algorithm in this paper also has good results. Experiments have proved that the use of long hash sequences can increase the ER value of the algorithm. Therefore, the proposed algorithm in this paper is highly discrimination.

Table 4 ER of the different hashing sequence lengths and different robust features

	ER		
Sequence length	640 bits	799 bits	1065 bits
MFCC Cosine	0.9642	0.9758	0.9813
CQCC Cosine	0.9784	0.9719	0.9813

Table 5 ER of the different algorithms

Algorithm	Ref.[38]	Ref.[17]	Ref.[25]	Ref.[28]
ER	0.9837	0.9112	0.9732	0.9062

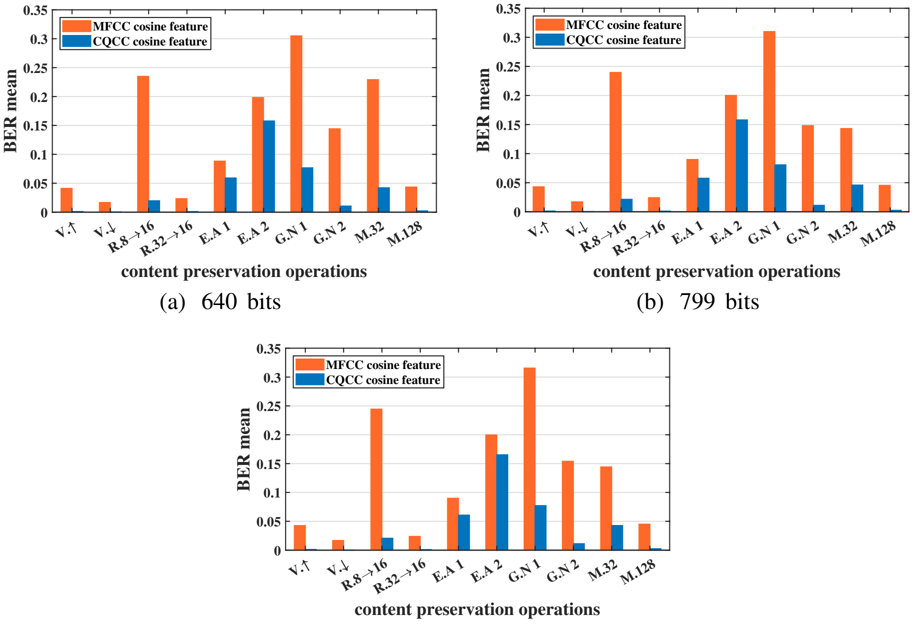
Table 6 Content preserving operations

Operating means	Operation method	Abbreviation
Volume Adjustment 1	Volume down 50%	V.1
Volume Adjustment 2	Volume up 50%	V.2
Resampling 1	Sampling frequency decreased to 8 kHz, and then increased to 16 kHz	R.8→16
Resampling 2	Sampling frequency increased to 32 kHz, and then dropped to 16 kHz	R.32→16
Echo Addition 1	Superimposed attenuation 30%, delay 100 ms,initial strength were 10% of the echo	E.A1
Echo Addition 2	Superimposed attenuation 60%, delay 300 ms, initial strength were 25% of the echo	E.A2
Narrowband Noise 1	SNR=30 dB narrowband Gaussian noise, center frequency distribution in 0 ~ 4 kHz	G.N1
Narrowband Noise 2	SNR=50 dB narrowband Gaussian noise, center frequency distribution in 0 ~ 4 kHz	G.N2
MP3 Compression 1	Re-encoded as MP3, and then decoding recovery, the rate is 32 k	M.32
MP3 Compression 2	Re-encoded as MP3, and then decoding recovery, the rate is 128 k	M.128

4.3 Robustness test and analysis

In order to simulate the interference of speech during channel transmission, Table 6 shows ten different content preserving operations.

Figure 8 shows the BER mean of different robust features and hashing sequence lengths. Compared with the MFCC cosine value feature, the BER average value of the feature value used in this paper does not exceed 0.1653, while the maximum value of the MFCC cosine value reaches more than 0.3, indicating that the algorithm in this paper is very good Robustness for various content preserving operations. As the sequence length increases, the robustness of the content preserving operation decreases, but the decrease is relatively small, which does not affect the overall robustness of the algorithm. In



(a) 640 bits

(b) 799 bits

(c) 1065 bits

Fig. 8 Comparison of BER means of different robust features and different hashing sequence lengths

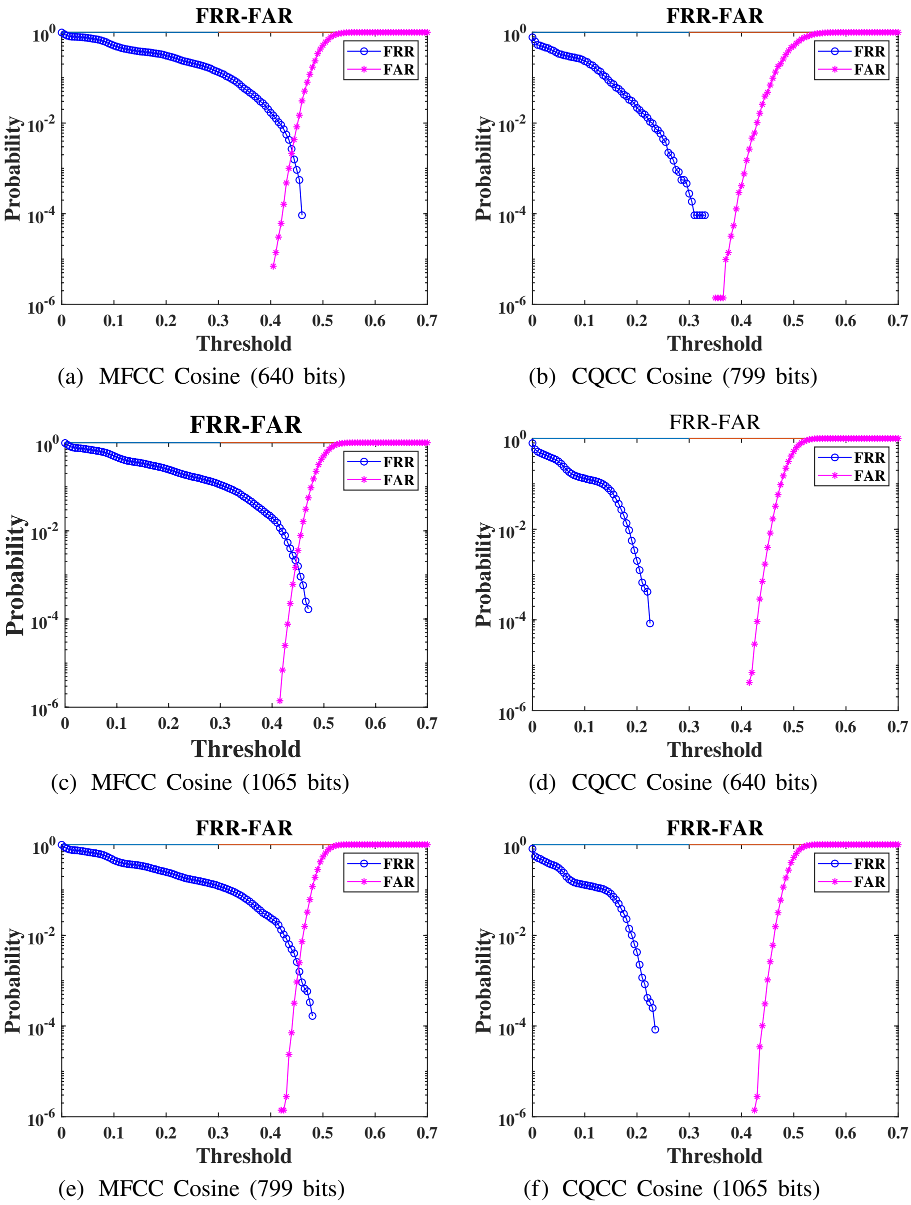
order to balance discrimination and robustness, the hashing sequence uses 1065 bits, which has the best overall effect.

Calculate the FRR values of different features and hashing sequence lengths according to Table 6, and then combine the discrimination to obtain the FAR value. The FAR-FRR curves of different features and hashing sequence lengths are shown in Fig. 9.

As shown in Fig. 9, the FRR and FAR curves of the MFCC cosine values of different hashing sequence lengths intersect, which cannot balance discrimination and robustness. However, the FRR and FAR curves of different hashing sequence lengths in this paper do not cross. The ability to accurately distinguish between content preserving operations and different content speech shows that the Proposed algorithm has good discrimination and robustness. The BER means comparison results of this algorithm and Ref. [38], Ref. [17], Ref. [25] and Ref. [28] are shown in Table 7 and Fig. 10.

It can be obtained from Table 7 and Fig. 10: for different content preserving operations, the algorithm in this paper is superior to other algorithms except the echo operation. Therefore, the proposed algorithm has better robustness. Since echo has a large interference to the original speech waveform, the algorithm is less robust to echo operations. For the interference of MP3 compression, the robustness of the proposed algorithm is better than other algorithms.

Calculate the FRR values of different algorithms according to Table 6, and then combine the discrimination to obtain the FAR value. The FAR-FRR curves of different algorithms are shown in Fig. 11. Figure 11a is the FRR-FAR curve diagram of the proposed algorithm in this paper. The interval between the final drop points of FRR and FAR is [0.235 0.425], indicating that the algorithm in this paper has a good distinction and robustness, capable of accurately identifying content preserving



(a) MFCC Cosine (640 bits)

(b) CQCC Cosine (799 bits)

(c) MFCC Cosine (1065 bits)

(d) CQCC Cosine (640 bits)

(e) MFCC Cosine (799 bits)

(f) CQCC Cosine (1065 bits)

Fig. 9 BER normal distribution with different robust features and different hashing sequence lengths

operations and speech of different content. As shown in Fig. 11b and e, the FRR-FAR curve obtained by Refs. [28, 38] does not cross, and the interval between the final falling point of FRR and FAR are [0.305 0.370] and [0.330 0.350]. Compared with the algorithm in this paper, there is no larger threshold selection space to balance the distinction and robustness. Compared with Fig. 11c and d, Ref. [17] has crossover in the figure, reflecting that the distinction and robustness cannot be solved well. Although

Table 7 BER mean and standard deviation of different algorithms

Algorithm	The algorithm		Ref. [38]		Ref. [17]		Ref. [25]		Ref. [28]	
Abbreviation	Mean	Variance	Mean	Variance	Mean	Variancex	Mean	Variance	Mean	Variance
V.1	0.0012	0.0027	0.0833	0.0512	0.0034	0.0041	0.0090	0.0071	0.0036	0.0039
V.2	0.0004	0.0010	0.0835	0.0391	0.0011	0.0020	0.0070	0.0056	0.0014	0.0023
R.8→16	0.0206	0.0173	0.0854	0.0689	0.0747	0.0583	0.0529	0.0271	0.0670	0.0577
R.32→16	0.0010	0.0023	0.0039	0.0047	0.0043	0.0051	0.0048	0.0061	0.0022	0.0029
E.A1	0.0607	0.0087	0.0771	0.0217	0.0402	0.0113	0.0315	0.0120	0.0414	0.0123
E.A2	0.1653	0.0191	0.1735	0.0268	0.1184	0.0205	0.1171	0.0231	0.1178	0.0207
G.N1	0.0772	0.0388	0.1300	0.0570	0.1312	0.0407	0.1108	0.0357	0.1647	0.0431
G.N2	0.0112	0.0092	0.0223	0.0190	0.0338	0.0218	0.0391	0.0207	0.0383	0.0025
M.32	0.0427	0.0194	0.0734	0.0427	0.1620	0.2017	0.2535	0.0246	0.1834	0.0909
M.128	0.0025	0.0025	0.0183	0.0131	0.0070	0.0067	0.2325	0.0243	0.0406	0.0049

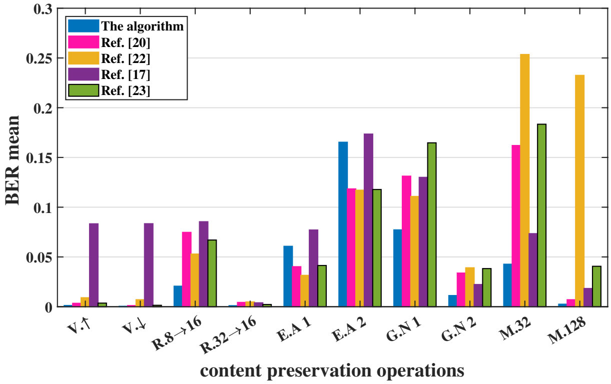


Fig. 10 BER normal distribution with different robust features and different hashing sequence lengths

there is no crossover in Ref. [25], it is also difficult to balance distinction and robustness. The experimental results show that the proposed algorithm has good distinction and robustness, and can accurately recognize content preserving operations and speech clips of different content.

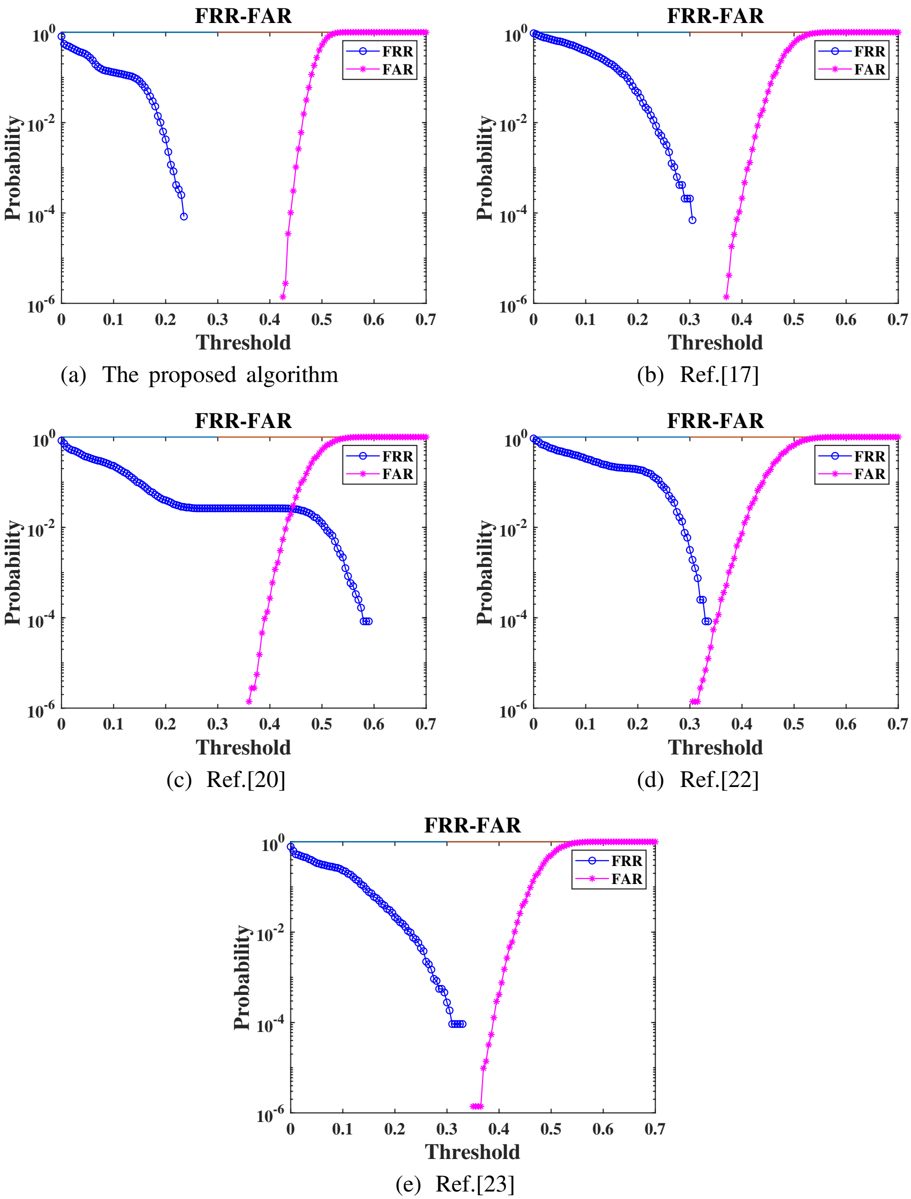
4.4 Verification and analysis of matching rate in complex noise environment

In order to further verify the anti-interference ability of the proposed algorithm against various background noises, the matching rate M_r is introduced.

$$M_r = \frac{T_A}{T_A + T_R + F_A} \tag{27}$$

where, T_A is the algorithm that correctly recognizes the number of speech clips with the same perceptual content, T_R is the algorithm that incorrectly recognizes the number of speech clips with the same perceptual content, F_A is the algorithm that correctly recognizes the number of speech clips with the different perceptual content. The threshold τ is selected as the minimum BER of FAR curve. Different algorithms select different thresholds: the proposed algorithm is 0.4173, that is MFCC cosine, that in Ref. [38] is 0.3677, that in Ref. [17] is 0.3593, that in Ref. [25] is 0.3037, and that in Ref. [28] is 0.3472. Figure 12 shows the comparison of the matching rate between the proposed algorithm, MFCC Cosine and that in Refs. [17,20,22,23] under eight different noise environments.

As shown in Fig. 12, for Factory noise 1, Gaussian white noise, HF channel noise, and Machine gun noise, the matching rate of this algorithm is higher than other algorithms. For all noises, the proposed algorithm has a matching rate of 100% when the SNR is greater than 10db. This is also the MFCC cosine value feature that cannot be compared with Refs. [17, 25, 28, 38]. The proposed algorithm only has a slightly lower matching rate than the MFCC cosine feature under Factory2 noise and Volvo noise. For other noises, the performance of the MFCC cosine feature is poor. When the SNR of the noise is below 0db, the matching rate of Ref. [38] is basically less than 30%, which cannot be well adapted to the speech biometric authentication in complex environments. Ref. [28] only has higher matching rate than the algorithm in this paper in terms of Factory2 noise. Refs. [17, 25] has higher matching rate than the algorithm in this paper in terms of Babble noise, Factory1 noise and Pink noise. On the whole, the algorithm in this paper is more robust than Refs. [17, 25, 28, 38], and can better achieve biometric authentication under extreme noise



(a) The proposed algorithm

(b) Ref.[17]

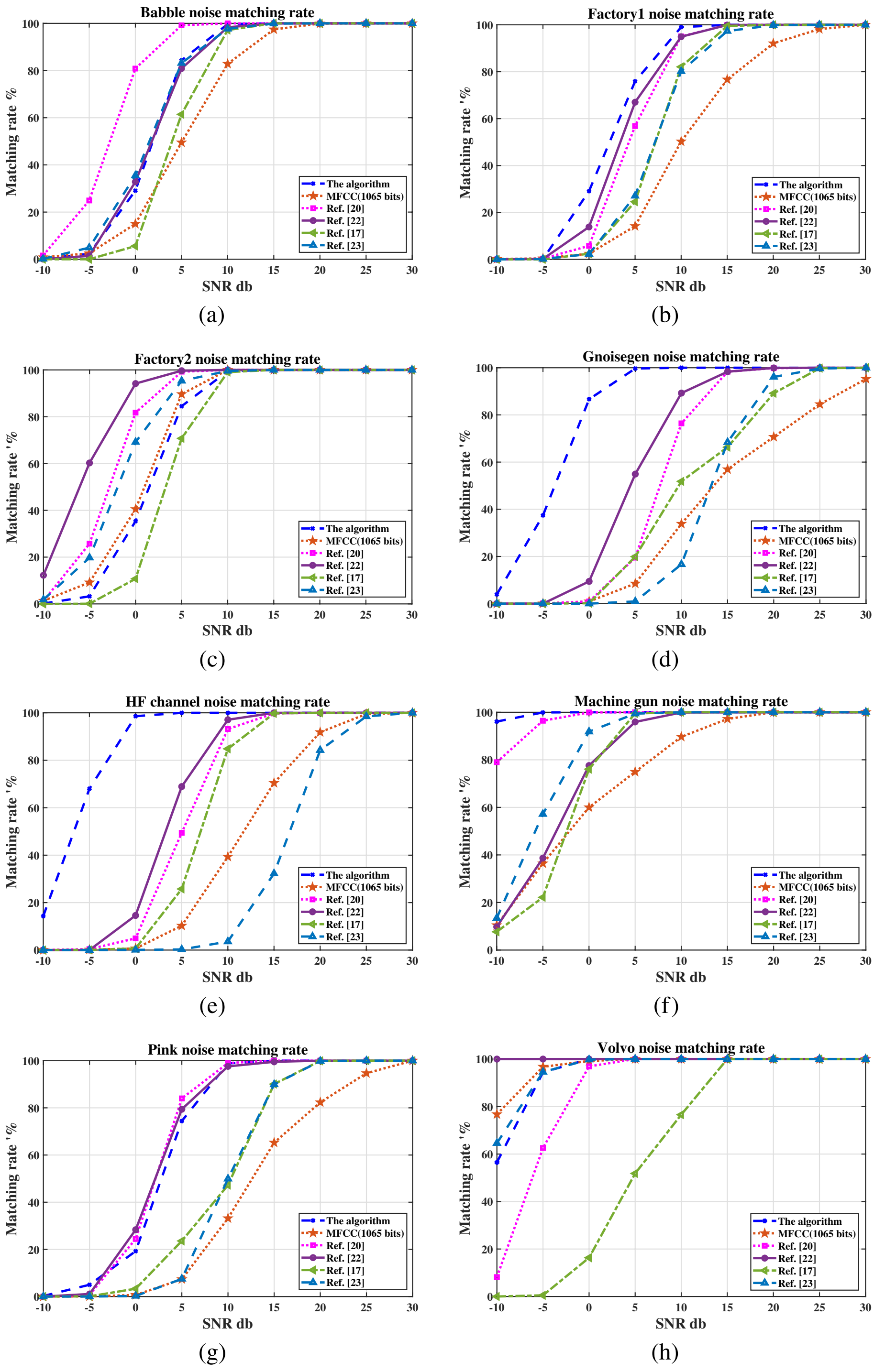
(c) Ref.[20]

(d) Ref.[22]

(e) Ref.[23]

Fig. 11 The FRR-FAR curves of different algorithm

environments. Therefore, the algorithm proposed in this paper has strong robustness to different noises under low SNR, and can meet the needs of speech matching in complex environments.



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Fig. 12 Comparison of matching rates of different algorithms under different noises

4.5 Unidirectional and security testing and analysis

In order to verify that the biometric hashing has the unidirectionality of the trapdoor, a unidirectivity verification algorithm based on logarithmic ratio method is proposed in this paper.

Randomly extract the speech clips from the speech database, and its original speech feature obtains the biometric security template through the direction of part A in Fig. 13, and then obtains the speech feature through the direction of part B, and finally calculates the difference between the two biometric feature sequences. The logarithmic ratio difference method between two sequences is defined as:

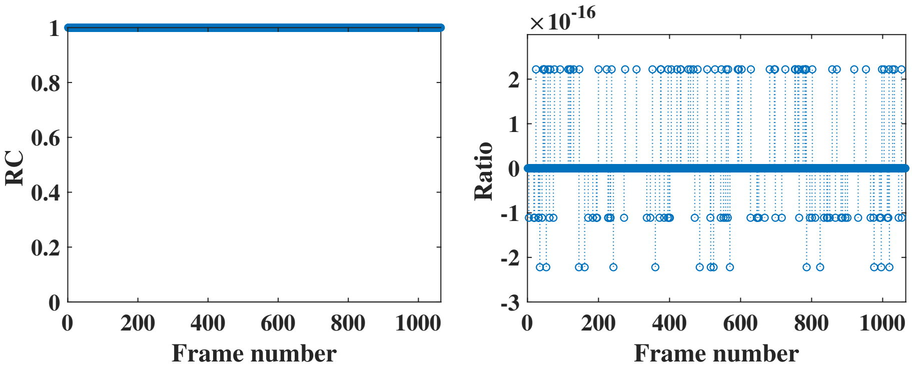
$$RC(n)=\begin{array}{l} 1,-10^{-5}\leq \log\left(\frac{F'(n)}{F(n)}\right)\leq 10^{-5} \\ 0,\quad Otherwise \end{array} \tag{28}$$

where, F' is the feature value obtained from the biometric security template. F is the original feature value. RC is the difference state of the biometric features.

This paper randomly extracts speech clips from the original speech database to verify the unidirectionality of the biometric hashing algorithm with trapdoors. Figures 14 and 15 show the difference between the features F_1' and F_2' obtained by the correct secret key and the wrong secret key and the original feature F respectively.

According to Figs. 14 and 15, there is a slight gap between the feature F_1' obtained by the correct key and the original feature F, and the gap between the two is only distributed in $(-2.2\times10^{-16}, 2.2\times10^{-16})$. The feature F_2' extracted using the wrong key is completely different from the original feature F. The distance between the two is distributed around -4.1. Since the error is only 10^{-8}, the error is too small, so it is shown as a straight line in Fig. 15b. Compared with the correct secret key, the feature sequence generated by the wrong secret key has a larger gap with the original feature sequence, which explains the one-way nature of the biometric hashing with trapdoor.

In order to further verify the one-way nature of the biometric hashing algorithm with trapdoors, this article first randomly extracts 150 speech from the speech database. Calculate the Hamming code distances between F_1', F_2' and F respectively. The Hamming code distance is shown in Fig. 16.



(a) (b)

Fig. 13 Verify the unidirectional block diagram of the biometric security template with trapdoor

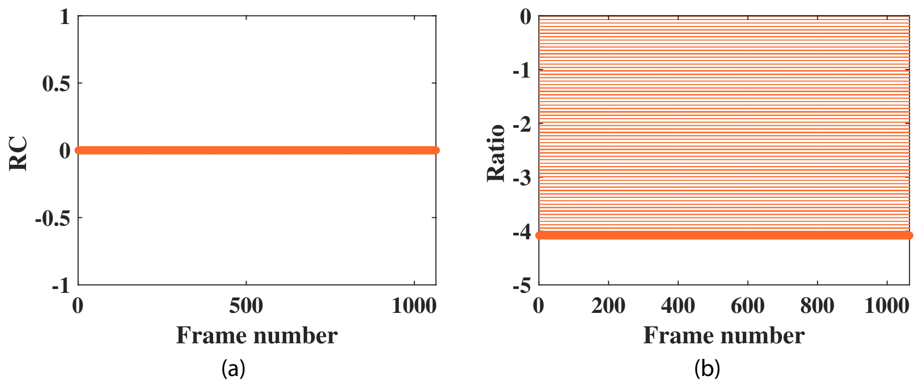


Fig. 14 Features extracted from the correct key

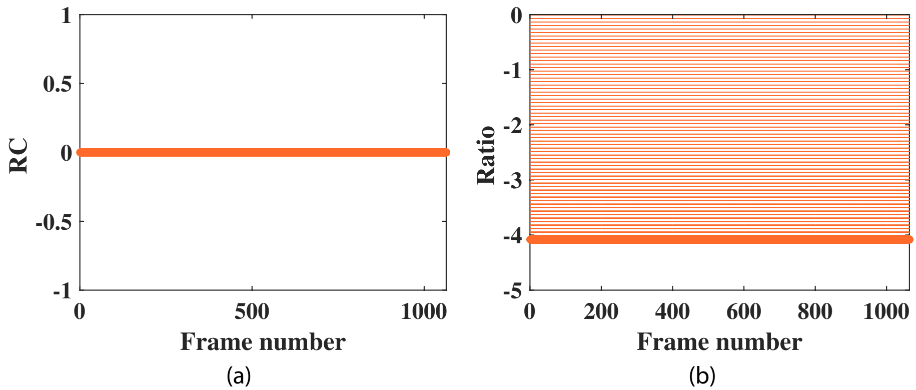


Fig. 15 Features extracted from the wrong key

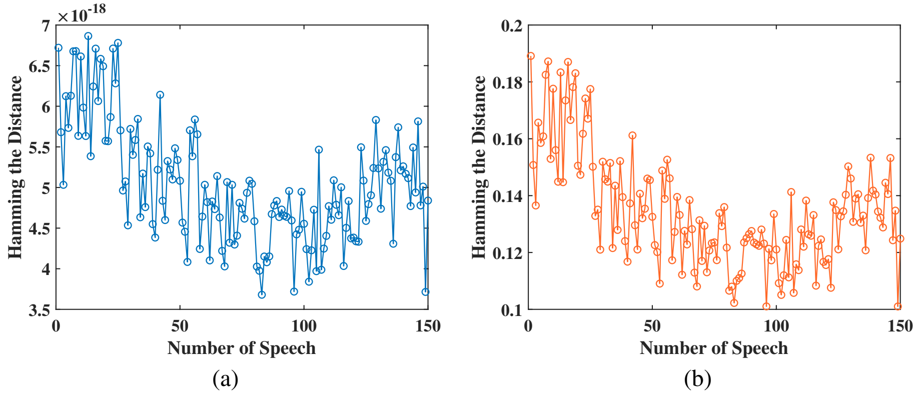


Fig. 16 Hamming code distance between $F_1^{'}$, $F_2^{'}$ and F noises

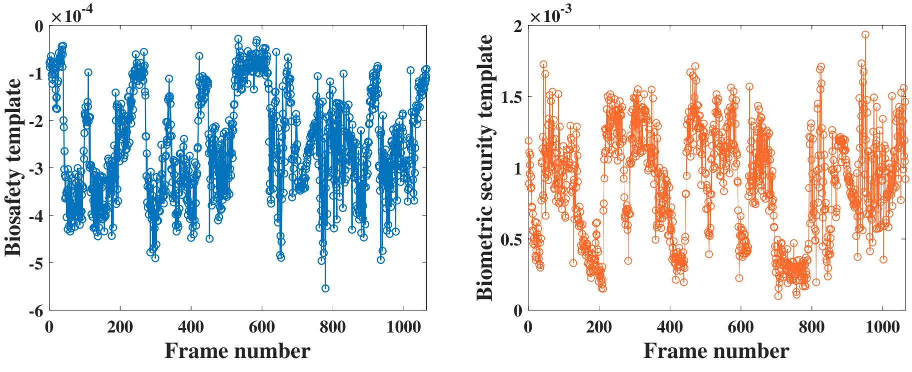
As shown in Fig. 16, the Hamming distance range between the feature F_1' obtained by the correct key and the original feature F is $(-3.7 \times 10^{-18}, 6.9 \times 10^{-18})$, and the Hamming distance range between the feature F_2' obtained by the wrong key and the original feature F is (0.10, 0.19), further verified that the biometric hashing algorithm is one-way with trapdoors, and also proved the security of the biometric hashing algorithm.

In order to enhance the security of the algorithm, this paper uses chaotic shift when constructing the biometric security template. Figure 17 compares the correct chaotic shift and the wrong chaotic shift to obtain the biometric security template. The values of the biometric security template obtained are completely different. The algorithm cannot get the required the biometric security template when the correct chaotic shift is not known. It also proves this article Algorithm security.

4.6 Efficiency testing and analysis

In order to verify the efficiency of the proposed algorithm, 200 speech clips in the speech library were randomly selected and the average running time was calculated. In order to ensure the consistency of the experiment, the same operating environment and speech data clips are used. Table 8 shows the comparison results of the algorithm in this paper, the MFCC cosine value, and the algorithm in Refs. [17, 25, 28, 38].

As shown in Table 8, as far as the algorithm in this paper is concerned, as the length of the hash sequence increases, although the efficiency performance of the algorithm is decreasing, the difference is small, which meets the requirements of real-time authentication. The hashing long sequence used by the proposed algorithm is compared with other hashing short sequences of the proposed algorithm, although the efficiency performance is low, the discrimination is greatly improved. Compared with the MFCC cosine value, when the hashing sequence is 1065 bits, the MFCC cosine value is 1.08 times of the proposed algorithm. Compared with other literatures, the efficiency of this algorithm is 2.21 times that of Ref. [17] and 1.07 times that of Ref. [28]. But compared to Refs. [25, 38], Ref. [38] is 5.44 times of the algorithm in this paper, and Ref. [25] is 3.31 times of the algorithm in this paper. Due to the use of long hashing sequences and chaotic shifts of biometric features, this paper has a long running time, so the efficiency performance of this algorithm is low. Although the length of the hashing sequence in this paper is 4 times that of Ref. [25] and 3 times that of Refs. [17, 28,



(a) (b)

Fig. 17 Biometric security templates for correct chaotic shift and false chaotic shift

Table 8 Efficiency of the different algorithms

Algorithms	Hashing sequence length	Working frequency	Average time
The algorithm	640 bits	3.4 GHz	0.0508s
	799 bits	3.4 GHz	0.0693s
	1065 bits	3.4 GHz	0.0827s
MFCC cosine	640 bits	3.4 GHz	0.0469s
	799 bits	3.4 GHz	0.0592s
	1065 bits	3.4 GHz	0.0767s
Ref. [38]	360 bits	3.4 GHz	0.0152s
Ref. [17]	360 bits	3.4 GHz	0.1825s
Ref. [25]	266 bits	3.4 GHz	0.0250s
Ref. [28]	360 bits	3.4 GHz	0.0883s

38], the algorithm in this paper performs very well in efficiency performance and can meet the requirements of efficiency authentication.

5 Conclusions and future work

This paper proposes a long sequence biometric hashing authentication algorithm based on 2D-SIMM and CQCC cosine. The algorithm has good comprehensive performance and solves the problems of poor discrimination, low robustness and security in the existing biometric authentication algorithm. Through experimental analysis, the following conclusions can be drawn: the use of long hashing sequences can effectively reduce the probability that different speech segments are recognized as the same segment, and improve the authentication rate of the algorithm. The extracted biometric features can well deal with the interference of the volume, resampling, MP3 compression and other content preserving operations. For low SNR, Babble and other complex noise environments have better matching rate. This paper uses the ratio method to prove the one-way nature of the biometric hashing algorithm with trapdoor. The biometric security template produced by 2D-SIMM has high security and reduces the risk of biometric features leakage.

Because the hashing sequence is too long, it will occupy a larger storage space, resulting in a decrease in efficiency performance. The speech biometric content is subject to tampering attacks, resulting in missing and forged information, and it is impossible to determine the accuracy of authentication. Therefore, the next work needs to further optimize the length of the hashing sequence to realize the tampering detection and positioning of the speech biometric content.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grant 61862041 and Youth Science and Technology Fund of Gansu Province of China under Grant 1606RJYA274.

References

1. Abdullahi SM, Wang H, Li T (2020) Fractal coding-based robust and alignment-free fingerprint image hashing. IEEE Trans Inf Forensics Secur 15:2587–2601
2. Anjith G, Zohreh M, David G, Olegs N, André A, Sébastien M (2019) Biometric face presentation attack detection with multi-channel convolutional neural network. IEEE Trans Inf Forensics Secur 15:42–55
3. Atighehchi K, Ghammam L, Barbier M, Rosenberger C (2019) Greychashing: combining biometrics and secret for enhancing the security of protected templates. Futur Gener Comput Syst
4. Boujelben Ons, Bahoura Mohammed (2018) Efficient fpga-based architecture of an automatic wheeze detector using a combination of mfcc and svm algorithms. J Syst Archit 88:54–64
5. Cao D, Gao X, Gao L (2017) An improved endpoint detection algorithm based on mfcc cosine value. Wirel Pers Commun 95(3):2073–2090
6. Chang D, Garg S, Hasan M, Mishra S (2020) Cancelable multi-biometric approach using fuzzy extractor and novel bit-wise encryption. IEEE Trans Inform Forensics Secur PP(99):1–1
7. Chen Y, Wo Y, Xie R, Chudan W, Han G (2019) Deep secure quantization: On secure biometric hashing against similarity-based attacks. Signal Process 154:314–323
8. Deng M, Meng T, Cao J, Wang S, Zhang J, Fan H (2020) Heart sound classification based on improved mfcc features and convolutional recurrent neural networks. Neural Netw 130:22–32
9. Dexing Z, Huikai S, Du X (2019) A hand-based multi-biometrics via deep hashing network and biometric graph matching. IEEE Trans Inform Forensics Secur 14(12):3140–3150
10. Gomez-Barrero M, Fierrez J, Galbally J, Maiorana E, Campisi P (2016) Implementation of fixed-length template protection based on homomorphic encryption with application to signature biometrics. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 191–198
11. Hamd MH, Mohammed MY (2019) Multimodal biometric system based face-iris feature level fusion. Int J Mod Educ Comput Sci 11(5):1–9
12. Huang Y-B, Zhang Q-Y, Yuan Z-T, Yang Z-P (2015) The hash algorithm of speech perception based on the integration of adaptive mfcc and lpcc. J Huazhong Univ Sci Technol (Natural Science Edition) (in Chinese) (02):124–128
13. Huang Yi-bo, Wang Yong, Zhang Qiu-yu, Zhang Wei-zhao, Fan Man-hong (2020) Multi-format speech biohashing based on spectrogram. Multimed Tools Appl 79(33):24889–24909
14. Jami SK, Chalamala SR, Jindal AK (2019) Biometric template protection through adversarial learning. In: 2019 IEEE international conference on consumer electronics (ICCE). IEEE, pp 1–6
15. Jiang Y, Chunxue W, Deng K, Yan W (2019) An audio fingerprinting extraction algorithm based on lifting wavelet packet and improved optimal-basis selection. Multimed Tools Appl 78(21):30011–30025
16. Jin-feng L, Wu T, Hong-xia W (2015) Perceptual hashing based on correlation coefficient of mfcc for speech authentication. J Beijing Univ Posts Telecommun 02:89–93
17. Jin-feng L, Wu T, Hong-xia W (2015) Perceptual hashing based on nmf and mdct coefficients. Chin J Electron (in Chinese) 03:579–583
18. Khurshid M, Selwal A (2020) A novel block hashing-based template security scheme for multimodal biometric system. In: Decision analytics applications in industry. Springer, pp 173–183
19. Kim H-G, Cho H-S, Kim JY (2016) Robust audio fingerprinting using peak-pair-based hash of non-repeating foreground audio in a real environment. Clust Comput 19(1):315–323
20. Li H, Qiu J, Teoh ABJ (2020) Palmprint template protection scheme based on randomized cuckoo hashing and minhash. Multimed Tools Appl: 1–25
21. Lifang W, Yukun M, Peng Z, Weishi Z (2016) Review of biometric template protection. Chin J Entific Instrum
22. Liu W, Sun K, Zhu C (2016) A fast image encryption algorithm based on chaotic map. Opt Lasers Eng 84:26–36
23. Nguyen TAT, Dang TK, Nguyen DT (2019) A new biometric template protection using random orthonormal projection and fuzzy commitment. In International Conference on Ubiquitous Information Management and Communication, pages 723–733. Springer
24. Nitin K, Manisha R (2020) Rp-lpp: a random permutation based locality preserving projection for cancelable biometric recognition. Multimed Tools Appl 79(3):2363–2381
25. Qiu-Yu Z, Hu W-J, Yi-bo H, Si-bin Q (2018) An efficient perceptual hashing based on improved spectral entropy for speech authentication. Multimed Tools Appl 77(2):1555–1581
26. Qiu-yu Z, Liang Z, Tao Z, Deng-hai Z (2019) A retrieval algorithm of encrypted speech based on short-term cross-correlation and perceptual hashing. Multimed Tools Appl 78(13):17825–17846

27. Qiu-Yu Z, Peng-Fei X, Yi-Bo H, Rui-Hong D, Zhong-Ping Y (2015) An efficient speech perceptual hashing authentication algorithm based on wavelet packet decomposition. J Inf Hiding Multimed Signal Process 6(2):311–322
28. Qiu-yu Z, Si-bin Q, Yi-bo H, Tao Z (2018) A high-performance speech perceptual hashing authentication algorithm based on discrete wavelet transform and measurement matrix. Multimed Tools Appl 77(16):21653–21669
29. Qiu-Yu Z, Tao Z, Si-Bin Q, Wu D-F (2019) Spectrogram-based efficient perceptual hashing scheme for speech identification. IJ Netw Secur 21(2):259–268
30. Ouali C, Dumouchel P, Gupta V (2016) A spectrogram-based audio fingerprinting system for content-based copy detection. Multimed Tools Appl 75(15):9145–9165
31. Shao H, Zhong D, Du X (2020) Towards efficient unconstrained palmprint recognition via deep distillation hashing. arXiv preprint arXiv:2004.03303
32. Sandhya M, Prasad MVNK (2017) Biometric template protection: a systematic literature review of approaches and modalities. In: Biometric security and privacy. Springer, pp 323–370
33. Sonnleitner R, Widmer G (2015) Robust quad-based audio fingerprinting. IEEE/ACM Trans Audio Speech Language Process 24(3):409–421
34. Tak H, Patino J, Nautsch A, Evans N, Todisco M (2020) An explainability study of the constant q cepstral coefficient spoofing countermeasure for automatic speaker verification. arXiv preprint arXiv:2004.06422
35. Vanita J, Gopal C, Nalin L, Akshit R, Shlok W (2019) Dynamic handwritten signature and machine learning based identity verification for keyless cryptocurrency transactions. J Discrete Math Sci Cryptogr 22(2):191–202
36. Wen-Sheng C, Haitao C, Binbin P, Bo C (2019) Robust nonnegative matrix factorization based on cosine similarity induced metric. In International Conference on Intelligent Science and Big Data Engineering, pages 278–288. Springer, 2019
37. Jichen Y, Rohan KD (2019) Low frequency frame-wise normalization over constant-q transform for playback speech detection. Digit Signal Process 89:30–39
38. Zhang Q-Y, Qiao S-B, Zhang T, Huang Y-B (2017) Multi-format audio perception hashing algorithms based on zero ratio. J Huazhong Univ Sci Technol (Natural Science Edition) (in Chinese) 45(6):33–38