Journal of Engineering Research

Volume 2
Issue 2 *(December)*

Article 14

2018

Discrete Wavelet Transform Based Cancelable Biometric System for Speaker Recognition

, Basant Abd El-wahab

Discrete Wavelet Transform Based Cancelable Biometric System for Speaker Recognition

Basant Samir Abd El-wahab[1], Heba A. El-khobby[1], Mustafa M. Abd Elnaby[1], Fathi E.Abd El-Samie[2]
1Department of Electronics and Electrical Communications Engineering, Faculty of Engineering, Tanta University, Egypt
2Department of Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University, Egypt
E-mail: b.s.abdelwahab@gmail.com, h_khobby@yahoo.com ,mnaby45@gmail.com,fathi_sayed@yahoo.com

***Abstract*—The biometric template characteristics and privacy conquest are challenging issues. To resolve such limitations, the cancelable biometric systems have been briefed. In this paper, the efficient cancelable biometric system based on the cryptosystem is introduced. It depends on permutation using a chaotic Baker map and substitution using masks in various transform domains. The proposed cancelable system features extraction phase is based on the Cepstral analysis from the encrypted speech signal in the time domain combined with the encrypted speech signal in the discrete wavelet transform (DWT). Then, the resultant features are applied to the artificial neural network for classification. Furthermore, wavelet denoising is used at the receiver side to enhance the proposed system. The cryptosystem provides a robust protection level of the speech template. This speech template can be replaced and recertified if it is breached. Our proposed system enables the generation of various templates from the same speech signal under the constraint of linkability between them. The simulation results confirmed that the proposed cancelable biometric system achieved higher a level of performance than traditional biometric systems, which achieved 97.5% recognition rate at low signal to noise ratio (SNR) of -25dB and 100% with -15dB and above.**

***Index Terms*— Biometric system, Cancelable biometric system, Mel-frequency Cepstral coefficients, Discrete wavelet transform, wavelet denoising, Speech encryption, Chaotic baker map.**

I. Introduction

BIOMETRIC recognition is one of the robust, convenient way and reliable for personal authentication. Accordingly, biometric systems are developed for numerous applications, such as the credit card industry, physical access control, telephone banking, voice mail, voice dialing and telephone shopping [1-3]. With the increasing use of biometric systems, there is a growing disquiet about the privacy and security of biometric data itself. Every person is alleged to have singular biometric (e.g., voice, face, iris and fingerprint). In the case of this biometric data is breached, it is unthinkable to have an alternative. Consequently, the security of biometric data is the critical issue in upgrading feasible biometric system. Recently, the appearance of cancelable biometric mitigates the biometric data issue. It transforms biometric data into deformed format, which can be replaced and cancelled if it is breached (reusability feature). Cancelable biometric must have ability to generate transformed biometric data for every application (diversity feature). It should be unidirectional transformed, consequently it bans minute to restore of biometric data in case of conciliation (irreversible feature). Finally, the cancelable biometric should not aggravation of the performance of system recognition relative to traditional biometric (performance feature).

There are various research in cancelable biometric system, including the proposed cancelable biometric system depend on the probabilistic random projection in [4]. The features are mined utilizing 2-D principal component analysis. Then, the feature vector is envisaged in the random projection process. The resultant vector is applied to a Gaussian mixture model (GMM). Another cancelable biometric system was introduced in [5], which depend on a fuzzy vault scheme. A biometric protection based on adding chaff points to the Mel frequency Cepstral coefficient (MFCC) vector to align difficult for attackers to separate of actual points from MFCC was suggested in [6]. Although, the authentication at the receiver has complexity with separating the chaff points and has not able to compare between templates in transform domain. So, a prime accumulator is utilized to separate the actual points from the chaff points. In [7], binarization method was presented for template protection based on extracting binary features depend on GMM-UBM recognition system. In [8], a binarization method was applied to extract binary features based on universal background models and Gaussian mixture models. Afterwards, the resultant feature vector is protected utilizing fuzzy commitment scheme.

There are many traditional transformations that can be applied to protect data. Generally, there are two kinds of speech encryption techniques, such as digital and analog. The analog speech encryption can be applied for narrow band radio communication and analog communication. This technique founded on permutation the components of the speech signal in the time domain [9], frequency domain [10], both time and frequency domains [11], Hadamard transform domain [12], wavelet transform domain [13] or circulant transform domain [14]. However, The digital speech encryption including the encryption advanced encryption standard (AES) [15], is considered one of the symmetric key algorithms that utilizes the fixed size of the block 128 bits and sizes of key 128, 192 or 256. The rounds number depends on the size of the key with intricate substitution and permutation to produce the

cipher output. These intricate algorithms pose the delay in the real time applications. Hence, these algorithms are not appropriate to encrypt the big amount of data for real time applications. Although, they are very secure and strong, but are delicate to the noise effect caused from high diffusion capability [16]. Chaotic maps are extensively utilized in image encryption owing to its randomness and its touchy to the initial parameters. Furthermore, it has efficient in non-predictability [17]. In this proposed cryptosystem, we attempt to conduct the perception of permutation and masking utilized in the chaotic Baker map for speech signal with various block sizes. Also, the low-complexity cryptosystem and the achieved high level of security are introduced.

This paper presents cancelable biometric system, which the transformation implemented in the signal domain. This proposed system applied the Mel frequency Cepstral coefficient (MFCC) for feature extraction from the encrypted signal in time domain combined with the encrypted speech signal in DWT. Afterwards, the feature vector is used in the classification process using the ANN. Since the MFCCs are not robust enough for noise effect, the MFCCs features from DWT is added to the feature of MFCC which extracted from time domain. Since the DWT possess the compaction of decomposing each sub-band for extracting features from various bands to improve the performance of the system. At the receiver side, wavelet denoising is used for enhancing the corrupted encrypted speech signal. In this system, the proposed cryptosystem is based on the permutation utilizing chaotic Baker map and substitution utilizing masks in various transform domains to blow remaining intelligibility resulting from permutation and masking in time domain. The substitution is applied for disposal of the silent periods in the speech signal and devastates pitch and format information. The performance of this proposed is measured based on recognition rate. The simulation results proved that this proposed achieved high performance and security level than traditional biometric system.

This paper is conduced as follow; section 2 introduces the proposed methodology and describes each technique utilized in this proposed system. Section 3 introduces the simulation results from our experimentation. Finally, section 4 introduces the conclusion and future work.

II. Methodology

The proposed cancelable biometric adopted the subsequent procedure: (i) the speech signal is encrypted utilizing the proposed cryptosystem, (ii) feature extraction using MFCC from the encrypted speech signal in time domain combined with the encrypted speech signal in DWT, (iii) ANN is applied for classification, and (iv) wavelet denoising for enhancing the corrupted received signal at recognition phase.

A. Cryptosystem

The proposed cryptosystem is utilized for permuting and masking the segments of the speech signal in various transform domains. The cryptosystem steps can be summarized as follows:

Step 1: Framing the speech signal and convert it from 1-D to 2-D blocks
Step 2: Generate the mask
Step 3: First round
- Permutation using a chaotic map
- Adding the mask

Step 4: Second round
- Convert to transform domains DST, DCT or DWT
- Permutation using a chaotic map
- Adding the mask
- IDST, IDCT or IDWT

Step 5: Third round
- Permutation using a chaotic map

Step 6: Reshaping format from 2-D into 1-D

The elaborated description of each step is as follows.

***Chaotic baker map*:** In the encryption, the Chaotic baker system [17-21] depends on permutation by reordering the elements within the signal block. This system is touchy to the initial parameters. So, if the system used different parameters, it will run in different paths, which are intricate to be calculated and analyzed. The encrypted output from this system has efficient in non-predictability, low correlation and randomness. If the applied secret key is $[n_1, n_2, \ldots, n_L]$, where n_i is sub-key, L is a number of sub-key and $N= n_1+n_2+\ldots+n_L$, the chaotic permutation steps of an N*N square matrix consist of i) dividing the square N*N matrix into N rectangles with N elements and width n_i, and then ii) permuting each rectangle element by rearranging to a row by scanning the rectangle element from the bottom left corner to upper one. Rectangles are taken from left to right beginning with upper rectangles to the lower ones. As reveal in figure 1, an example for the secret key [2,4,2].

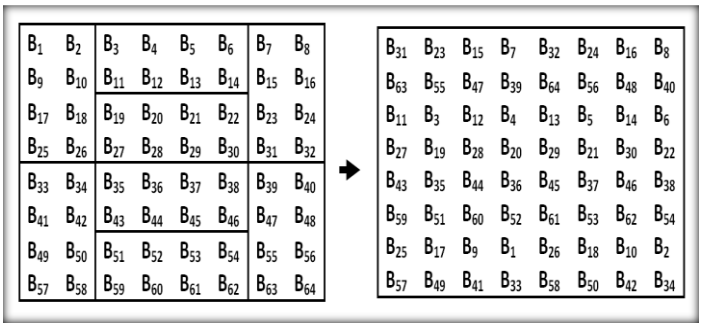


Fig. 1. Chaotic baker map.

Permutation: is used to rearrange and map every element of the block into new position within the same block without varying the values of the elements. In the proposed, the permutation in each round is performed by using the chaotic baker map, which has severe allergic for initial parameters and good randomness.

Masking: is vital to serve known-plaintext attacks by altering the energy of the signal within silent periods. The secret key is applied to generate the mask. The mask is generated by introducing the certain number of ones to blank block and then permuted by chaotic map. The deliverable mask is inserted to

every block after the step of permutation. For example, as shown in figure 2, the mask is generated by applying the secret key [4,2,2,2,2]. Therefore, the total of sub-key results in 12×12 size of the block and the sub-key number is 5, so the mask can be generated by performing the steps as follow:

1. The prim n_1 rows are filled by the number of ones equal to the sub-key number. So, in every of prim 4 rows in the matrix is filled by 5 ones.
2. The resultant matrix is mapped by chaotic Baker map. After that, the resultant mask is inserted to every block of the speech signal and then the resultant signal is circularly shifted between -1 and 1.

Substitution step: permutation of the signal in the time domain will lead to distortion in the time envelope of the speech. Consequently, the intelligibility of the speech will decrease. Nevertheless, parts of the signal keep proper, which may permit cryptoanalysis attacks to construe the encrypted speech. To overcome this issue, the substitution process is utilized to vary the speech power spectrum by varying the elements values in every block. In this proposed, the substitution is conducted by permuting and masking the signal in various transform domains like discrete wavelet transform (DWT), discrete sine transform (DST) and discrete cosine transform (DCT) to estimate the best suits for this mission.

B. Discrete wavelet transform

Wavelet transformation permits variant resolution in time and frequency. It has time and frequency localization. This functionality permits locality in both time and frequency spaces. The transformation locality of the signal is significant in the recognition process. The noise influences only in few coefficients, where diverse parts of the signal may deliver diverse quantities of information and when the signal ruined by noise in time or frequency. These coefficients constitute local information in time and frequency.

Consequently, the noise effect can reduce the recognition rate based on SNR. So, the wavelet is perfect solution to decrease the unwanted frequencies or noise in speech recognition. The wavelet transform is mathematical procedure, which utilized to split a given speech signal into diverse sub-bands with diverse scales to study separately every scale as observe in figure 2. The main concept of DWT is to decompose a signal into a series of its approximations (lowpass issuance) and its details (high pass issuance) at diverse resolutions. The signal is filtered by utilizing lowpass filter to provide its approximation and filtered by utilizing highpass filter to provide its details. Approximation coefficients carry the model or characteristics of the signal. Detail coefficients include high frequency components that are influenced by noise and include minor information about the speaker identity because they differ significantly with changes in the acquisition or recording conditions and the operative text. Furthermore, indicates that the output from every filter is down-sample of two. According to Nyquist rate, the signal can be perfectly reconstructed using half of its samples, where the greatest number of levels reliant on the signal length.

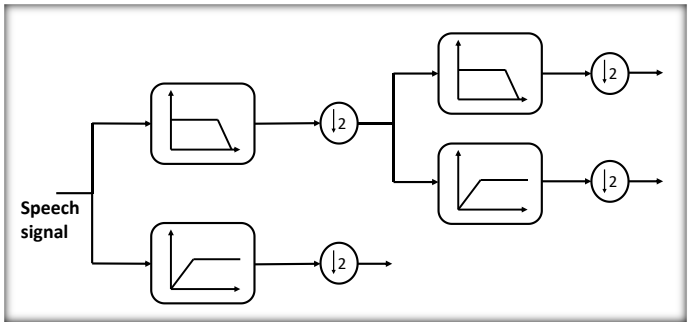


Fig. 2. Two level decomposition for DWT.

C. Feature extraction

The better parametric representation of the speech signal is significant mission to produce best recognition performance. Each speaker has singular characteristic for vocal features. These features are classified into two groups, namely learned and physiological features. Learned features are classified into high level and prosodic features, where the high-level features contents accent, phones, semantics and idiolect. The prosodic features consist of short term features. The disadvantage of learning features is that if the speaker recognition system achieved using it, the system will be very complex. So, the short term applied to implement the speaker recognition system. There are short term features extraction methods [22], such as MFCCs, Linear Predictive Cepstral Coefficients (LPCCs), Linear Prediction Coefficients (LPCs) and Perceptual Linear Predictive (PLP). In this work, the MFCC is used for extracting features, which is short term features and extracted from each frame of speech signal. Figure 3 reveals the MFCC procedure.

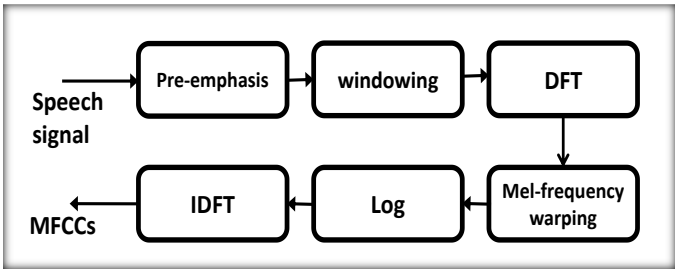


Fig. 3. Extracting MFCCs from the speech signal.

The description of the different steps in Figure 3 can be as follows.

Pre-emphasis: The first-order finite impulse response (FIR) is utilized to pre-emphasize the speech signal. The lower frequencies in speech contain higher energy and thus are preferentially designed with respect to higher frequencies. Consequently, pre-emphasis is applied to reinforce the higher frequencies and to eliminate lip radiation and glottal effects. The transfer function of this filter is as follows [23, 24]:

$$H(Z) = 1 - aZ^{-1} \quad (1)$$

Where $0.9 \leq a \leq 0.99$

Frame blocking and window: The speech signal is quasi-stationary signal. The speech signal in the recognition system is divided into the brief time segments labeled frames. To bring the parameters of the frame differ smoothly, there is usually 50% overlap between neighbor frames. Then, the window is multiplied with each frame like a hamming window to improve the continuity between neighbor frames and make the end points of the framed change smoothly [24], where the hamming windows possess low amplitudes at the edges of the frame in the time domain. Consequently, at higher frequencies, the side lobes become less. The side lobes of this window will destroy the noise in the signal. So, it is desired to design the window with low noise bandwidth. This can be investigated by minimizing the amplitude of side lobe.

Discrete Fourier transform (DFT): is used for analyzing the properties of the frequency spectral of the speech signal. The Fourier transform represent the discrete signal x(n) in the frequency domain as follows [25]:

$$X(K) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nK/N} \qquad 0 \le \mathrm{K} \le \mathrm{N\text{-}1} \tag{2}$$

where n=0, 1,….N-1, and N is the number of the signal samples x(n), and K is the discrete frequency index and $j = \sqrt{-1}$. In addition, the original signal can be reconstructed from its DFT by using IDFT as follows:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(K) e^{j2\pi nK/N} \qquad 0 \le \mathrm{n} \le \mathrm{N\text{-}1} \tag{3}$$

The Mel filter bank: MFCC is predicated on perceptions of human hearing, which cannot envisage frequencies above 1KHZ. In other hand, MFCC is predicated on a known bandwidth variation of the human ear with frequency. Consequently, MFCC utilized linear filter below 1KHZ and logarithmic filter above 1KHZ. Mel frequency scale utilized to represent the speech pitch to extract significant phonetic characteristic in speech as follows [23, 24]:

$$Mel(f) = 2.595 \left(1 + \frac{f}{700}\right) \tag{4}$$

Where f is the frequency in the linear scale and Mel is the Mel scale. MFCCs are extracted by utilizing filter bank. In speech recognition system, the Mel-scale filter bank consist of several triangular bandpass filters deployed within the bandwidth of the signal, where one filter is used for every desired component from Mel frequency.

Discrete cosine transform: the final phase includes conduct DCT on the logarithm of Mel spectrum. If the m^{th} Mel filters output are s(m) then the MFCCs are assigned as follows [26]:

$$c_g = \sqrt{\frac{2}{N} \sum_{m=1}^{N_f} \mathrm{Log}\big(\hat{\mathrm{s}}(m)\big) \cos\left(\frac{g\pi}{N_f}(m - 0.5)\right)} \tag{5}$$

Where N_f and C_g are the number of Mel filters and the MFCC respectively, g=0, 1, 2, ….., G-1, G, G refers to the number of the MFCCs. The resulting MFCCs number is selected between 12 and 20. The majority of signal information is represented by the first few coefficients.

Polynomial coefficients: MFCCs are delicate to channel mismatch between training and testing data and they are dependent on the speaker. To overcome this issue, the polynomial coefficients are inserted to the MFCCs to increase the similarity between the train and the testing utterance of the same speaker. The significant of these coefficients stem from that they can maintain worthy information (curvature, mean and slope) about the shape of a time function of Cepstral coefficient in the utterance.

D. Feature matching

Speaker modeling is the next stage in which the models of the speakers are trained to utilize the mined features of the speakers. The competence of the speaker models is reversed in the performance of the system recognition in order to minimize the error rate. There are many techniques for modeling, such as Artificial Neural Networks (ANNs) [27, 28], Vector Quantization (VQ), Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs). The ANNs is good relevant for building an authentication system. The feed forward neural network is used with linear transfer function for input and output layers and sigmoid transfer function for hidden layer. The error back-propagation algorithm is utilized to train this network, where it strengthens training with test and validation vectors. The test vectors are utilized to investigate if the network operates properly. The validation vectors are utilized to stop the network training when it is getting to performance that is minimized to the goal or to the highest epochs. The input layer composed of nodes for every input of the corresponding input speech signals. In the hidden layer, the node number was altered for improvement through trial-and-error and found that used one hidden layer with 125 nodes in the hidden layer obtained minimum square error between the desired output and actual output. The output layer composed of two nodes for taking the decision (accepts or refuses).

E. Wavelet denoising

There are various enhancement methods for getting rid of noise effect from the speech signal, such as spectral subtraction [29, 30], Wiener filter [31] and adaptive Wiener filter [31]. Channel degradation effect can be gotten rid from

the speech signal by using deconvolution methods like regularized deconvolution technique [32] and linear mean square error (LMMSE) technique [33, 34]. This proposed is focused in studying the effect of inserting wavelet denoising for enhancement the corrupted speech signal and make a comparative study between wavelet denoising method and various enhancement methods. The main aim of wavelet denoising is reducing the noise in corrupted speech signal. It is implemented by selecting a thresholding that is adequately a large multiple of the noise standard deviation in the speech signal. The majority of the noise power is eliminated by thresholding the detail coefficients of the wavelet transformed of speech signal. It is two species of thresholding, hard and soft threshold. The hard thresholding equation is given by [29]:

$$f_{hard}(y) = \begin{cases} y & |y| \geq Th \\ 0 & |y| < Th \end{cases} \quad (6)$$

The soft thresholding is given by:

$$f_{soft}(y) = \begin{cases} y & |y| \geq Th \\ 2y - Th & Th\backslash 2 \leq y < Th \\ Th + 2y & -Th < y \leq -Th\backslash 2 \\ 0 & |y| < Th\backslash 2 \end{cases} \quad (7)$$

Where Th is the value of threshold and y constitutes of the high frequency coefficient of DWT. In this experiment, the threshold will be chosen as 80% of the large coefficient value because this is the common threshold used in the literature [29].

F. Proposed cancelable biometric system

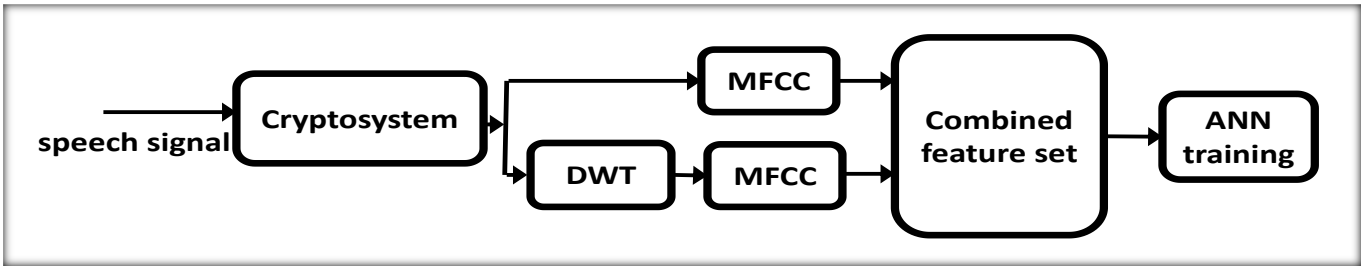
The proposed cancelable biometric system is revealed as in figure 4, which consists of two phases, namely training or enrollment phase and testing or recognition phase. At enrollment phase, the input speech signal is encrypted utilizing the cryptosystem. Then, the features are mined utilizing MFCC from the resultant encrypted signal in time domain combine with the features from the resultant encrypted signal in DWT. Afterwards, the resultant feature vector is used to train ANN. At recognition phase, the corrupted received signal is enhanced by utilizing wavelet denoising. Then the features are mined from the enhanced speech in time domain combined with the DWT. After that the resultant features are applied to ANN in order to recognize the speaker.

III. SIMULATION AND RESULTS

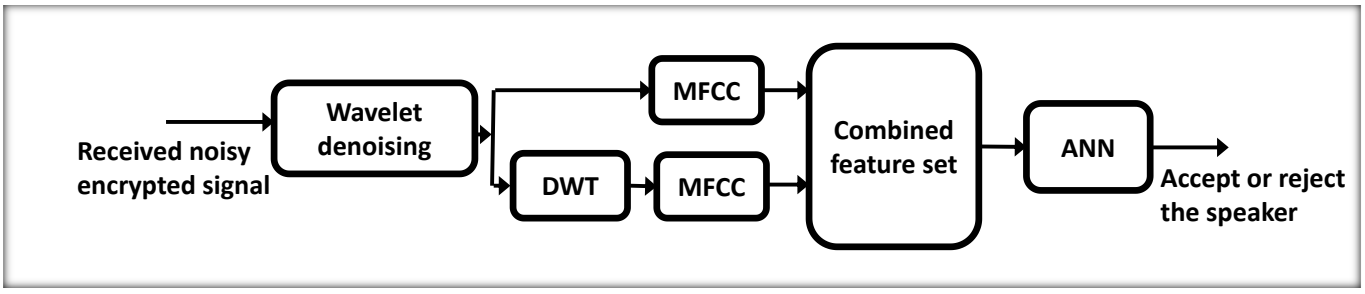
The experiments carried out to test the quality of the cryptosystem in various transform domains to utilize the highest quality in a cancelable biometric system. Subsequently, the experiments carried out to measure the performance of the cancelable system

A. Cryptosystem quality

The quality of cryptosystem in various transform domains is measured based on the correlation coefficient (C_r) and spectral distortion (SD) between the original and encrypted speech signal. The correlation coefficient near to zero and the highest spectral distortion, whenever the highest quality of the encrypted signal. The results for the comparison between the cryptosystem in various transform domains are tabulated in table I.



(a) Training mode



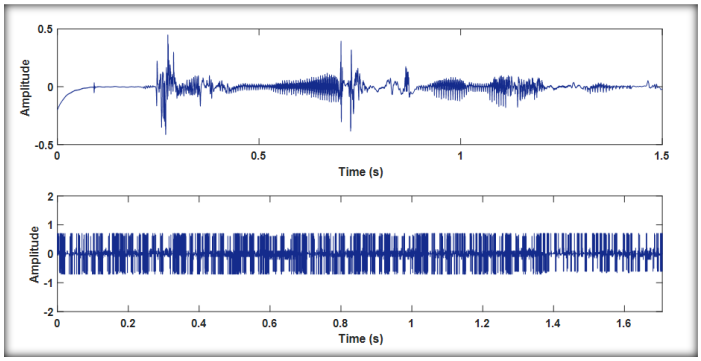
(b) Testing mode

Fig. 4. Block diagram for the cancelable biometric system.

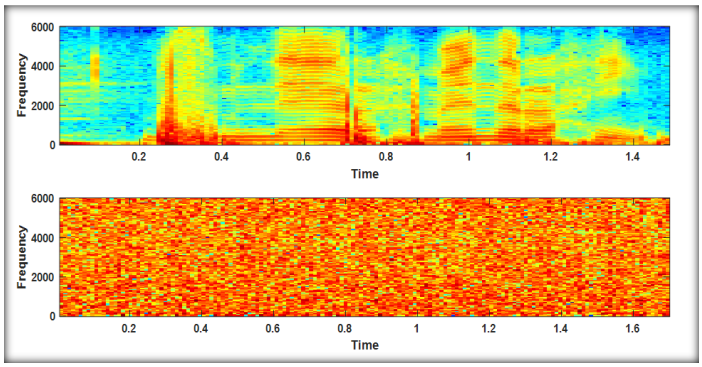
TABLE I
THE QUALITY OF THE PROPOSED OF A CRYPTOSYSTEM.

Symbol	Cr	SD
DCT	0.0033	27.9184
DST	0.0061	13.9352
DWT	-0.0113	36.177

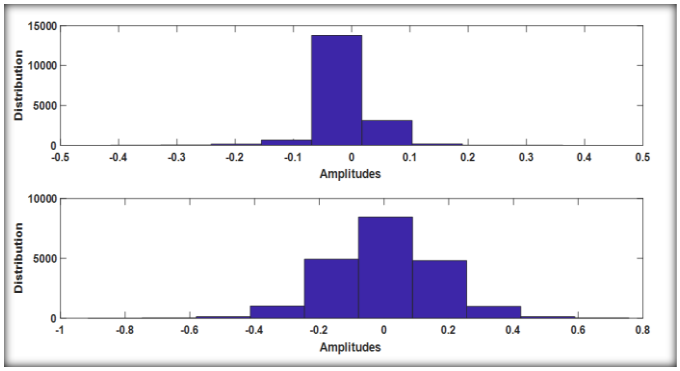
Table I depicts that the DWT is better than DST and DCT, which achieved -0.0113 correlation coefficient and 36.177 dB spectral distortions. Figure 5 (a) through (c) shows the time, the spectrogram and the histogram of the original signal and encrypted signal using cryptosystem in DWT, where the histogram shows that the DWT is uniform.



(a) Original signal versus encrypted signal.



(b) Original versus encrypted spectrogram.



(c) Original versus encrypted histogram.

Fig. 5. Time, spectrogram and histogram of the original signal and encrypted signal utilizing cryptosystem in DWT.

B. Recognition rate study

The comparison study in this section is implemented to compare between a traditional biometric system and the proposed cancelable system. This comparison implemented under the lowpass channel affect with AWGN corruption and SNR range from -25dB to 25dB. In the training phase, a database consisted of 15 speakers. To build this database, each of these speakers repeat specific sentence 10 times. Hence, MFCCs and polynomial coefficients are generated from 150 samples of speech in time domain combined the DWT of these samples to build the feature vectors of the data base. Afterwards, these features are used for training the ANN. In the testing phase, each of the speakers has asked to say the sentence once again and the corresponding speech signal is then corrupted. Subsequently, the features are extracted from these corrupted signals with the same method in training phase to apply them in the matching process. In the cancelable system, each speech signal from the speaker is encrypted as the first step in the training phase and test phase. The performance of the system is measured based on the recognition rate as shown in figure 6.

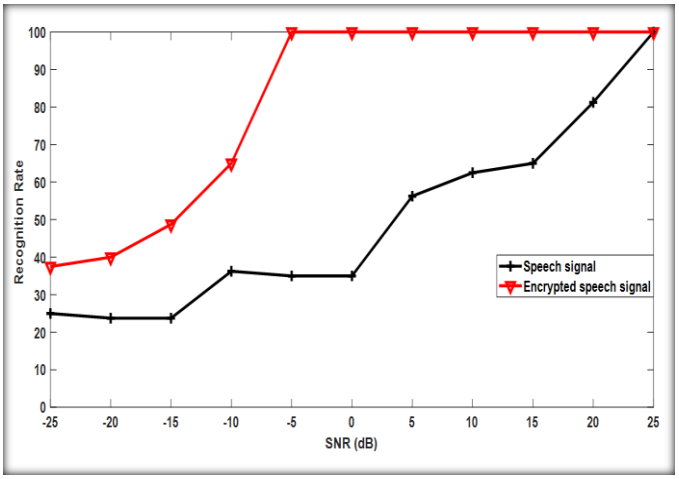


Fig.6 The recognition rate vs. SNR for traditional biometric system and proposed cancelable biometric system.

Fig. 6 illustrates that the cancelable biometric system achieved the best recognition rate than the traditional biometric system. It achieved 38% recognition rate at -25dB and 100% at -5dB and above, while the traditional biometric system achieved 25% at -25dB and 100% at 25dB. These results approve that the cryptosystem is impeccable for noise immunity and furthermore accomplishing a high level of security.

In this section, the performance of the cancelable system is studied with various types of wavelet families for denoising. The wavelets used in the simulation are Haar, Daubechies (db1, db4, db8, db20), Discrete Meyer (demy), Coiflets (coif1, coif3, coif5), Symlets (sym2, sym5). Various trials of experiments are carried out to test the system performance

when the hard or the soft threshold is used. The soft threshold obtained better performance than hard threshold. Consequently, the remaining experiments are implemented based on wavelet denoising with soft threshold. The performance comparison between the different wavelet filters is tabulated in the table II, which indicates that db1 wavelet family achieved the best performance of the system, which achieved the highest recognition rate closed to 87.5% at low SNR -25dB and achieved 100% at -15% and above. Additionally, Table III focused on the performance of the cancelable system when used various levels from db1 wavelet denoising.

TABLE II
THE PERFORMANCE OF VARIOUS WAVELET DENOISING FAMILIES.

SNR (DB)	**RECOGNITION RATE (%)**										
	Haar	db1	db4	db8	db20	dmey	coif1	coif3	coif5	sym2	sym5
-25	87.5	87.5	55	67.5	73.75	75	42.5	60	70	45	57.5
-20	97.5	97.5	68.75	81.25	80	80	56.25	67.5	75	57.5	83.75
-15	98.75	100	90	98.75	90	93.75	87.5	91.25	90	82.5	95
-10	100	100	100	100	100	100	100	100	100	100	100
-5	100	100	100	100	100	100	100	100	100	100	100
0	100	100	100	100	100	100	100	100	100	100	100
5	100	100	100	100	100	98.75	100	100	100	100	100
10	100	100	100	100	100	100	100	100	100	100	100
15	100	100	100	100	100	100	100	100	100	100	100
20	100	100	100	100	100	100	100	100	100	100	100
25	100	100	100	100	100	100	100	100	100	100	100

TABLE III
THE PERFORMANCE OF DB1 WAVELET DENOISING WITH DIFFERENT LEVELS.

SNR (dB)	Recognition rate (%)				
	db1 level 1	db1 level 2	db1 level 3	db1 level 4	db1 level 5
-25	87.5	90	97.5	93.75	97.5
-20	97.5	97.5	98.75	96.25	98.75
-15	100	100	100	100	100
-10	100	100	100	100	100
-5	100	100	100	100	100
0	100	100	100	98.75	100
5	100	100	100	100	98.75
10	100	100	100	100	98.75
15	100	100	100	100	100
20	100	100	100	100	100
25	100	100	100	100	100

Table III depicted that db1 at level 3 achieved the best performance, which achieved 97.5% at -25dB and 100% at -15dB. Furthermore, when the number of levels increased above level 3, the performance of the system corrupted, where the number of levels appropriates to the speech signal length. In this section, the comparison studies are implemented between various enhancement algorithms and db1 wavelet denoising with level 3, which obtained the best performance as illustrated in the above section. The simulation results are given in figure 7 these results reveal that the db1 wavelet denoising with level 3 is best suits for identification score.

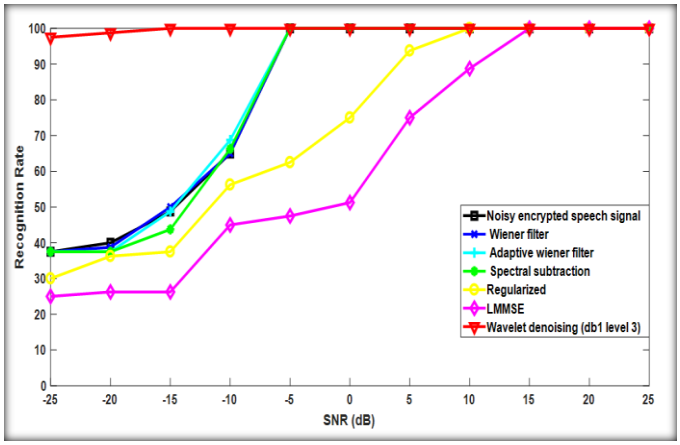


Fig.7 The recognition rate vs. SNR for a cancelable biometric system when used different enhancement methods.

These studies reveal supremacy of the wavelet denoising as the technique of noise mitigation to other techniques for utilizing with the proposed cancelable system. Db1 soft thresholding with 3 levels wavelet denoising achieve the best recognition rate. Consequently, the wavelet denoising can be used with the cancelable system during the feature extraction from the speech signal combined with DWT of the signal to obtain the best performance of the cancelable system in noisy environment. The wavelet denoising and DWT possess the compaction of decomposing each sub-band, which allow for extracting the features from various bands to improve the performance of the system.

The results proved that the proposed system added more security to the traditional biometric systems more and can prevent attackers to spy on the system. In addition, it does not require decryption process at the receiver for time saving. This purposed depends on extracting features from encrypted speech signal; hence the extracted features will be adaptable features and saved from attackers. Because it is phantom features which relies on the initial parameters of the cryptosystem like block size and the secret key (Reusability feature). If the cryptanalysis has samples from the original signal and its encrypted edition and possesses liberty to use them for discovering the key, he must have knowledge about the block size to construct alleged substitution and permutation processes. If he attempts with different block size, this would lend differ results (unlinkability feature). So, the knowledge about plaintext without knowing the block size pointless and in the other hand, it is very hard to guess the key. If the database of the system is stolen, the system can protect itself by re-selecting the initial parameters of the cryptosystem to update the system database. The simulation results confirmed that the proposed cancelable biometric system achieved higher a level of performance than traditional biometric systems, which achieved 97.5% recognition rate at low signal to noise ratio (SNR) of -25dB and 100% with -15dB and above.

IV. CONCLUSION

For speaker recognition system, a proposed cancelable biometric system proposed cryptosystem, which based on permutation and masking utilizing chaotic Baker map. This proposed based on extracting features using MFCC from the encrypted speech signal in time domain combined with the features from the encrypted signal in DWT. After that applied feature vector in ANN for classification. At recognition phase, wavelet denoising is used for enhancing the corrupted received signal with a comparative study between the various wavelet families denoising and various enhancement methods. The simulation result reveals that this proposed improves the performance of a traditional biometric system which achieved high recognition rate at low SNR up to 97.5% at -25dB and 100% recognition rate at -15dB when enhancement this proposed utilizing db1 wavelet denoising. Furthermore, this proposed meets the assessment criteria of the biometric protection, such as diversity and reusability.

V. REFERENCE

[1] D. A. Reynolds, "An overview of automatic speaker recognition technology," in *Acoustics, speech, and signal processing (ICASSP), 2002 IEEE international conference on*, 2002, pp. IV-4072-IV-4075.

[2] J. R. C. de Lara, "A method of automatic speaker recognition using cepstral features and vectorial quantization," in *Iberoamerican Congress on Pattern Recognition*, 2005, pp. 146-153.

[3] B. Choudhury, P. Then, B. Issac, V. Raman, and M. K. Haldar, "A survey on biometrics and cancelable biometrics systems," *International Journal of Image and Graphics*, vol. 18, p. 1850006, 2018.

[4] A. B. J. Teoh and L.-Y. Chong, "Secure speech template protection in speaker verification system," *Speech communication*, vol. 52, pp. 150-163, 2010.

[5] A. Juels and M. Sudan, "A fuzzy vault scheme," *Designs, Codes and Cryptography*, vol. 38, pp. 237-257, 2006.

[6] W. Xu and M. Cheng, "Cancelable voiceprint template based on chaff-points-mixture method," in *Computational Intelligence and Security, 2008. CIS'08. International Conference On*, 2008, pp. 263-266.

[7] M. Paulini, C. Rathgeb, A. Nautsch, H. Reichau, H. Reininger, and C. Busch, "Multi-bit allocation: Preparing

voice biometrics for template protection," in *Odyssey 2016*, 2016, pp. 291-296.

[8] S. Billeb, C. Rathgeb, H. Reininger, K. Kasper, and C. Busch, "Biometric template protection for speaker recognition based on universal background models," *IET Biometrics*, vol. 4, pp. 116-126, 2015.

[9] H. Beker and F. Piper, "Secure speech communications", Academic Press, London," 1985.

[10] L.-S. Lee, G.-C. Chou, and C.-S. Chang, "A new frequency domain speech scrambling system which does not require frame synchronization," *IEEE transactions on communications*, vol. 32, pp. 444-456, 1984.

[11] R. Milton, "A time and frequency-domain speech scrambler," in *Communications and Signal Processing, 1989. COMSIG 1989. Proceedings., Southern African Conference on*, 1989, pp. 125-130.

[12] Y. Wu and B. P. Ng, "Speech scrambling with Hadamard transform in frequency domain," in *Signal Processing, 2002 6th International Conference on*, 2002, pp. 1560-1563.

[13] F. Ma, J. Cheng, and Y. Wang, "Wavelet transform-based analogue speech scrambling scheme," *Electronics Letters*, vol. 32, pp. 719-721, 1996.

[14] G. Manjunath and G. Anand, "Speech encryption using circulant transformations," in *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*, 2002, pp. 553-556.

[15] J. Daemen and V. Rijmen, "Rijndael, the advanced encryption standard," *Dr. Dobb's Journal*, vol. 26, pp. 137-139, 2001.

[16] D. Tseng and J. Chiu, "An OFDM speech scrambler without residual intelligibility," in *TENCON 2007-2007 IEEE Region 10 Conference*, 2007, pp. 1-4.

[17] J. Fridrich, "Symmetric ciphers based on two-dimensional chaotic maps," *International Journal of Bifurcation and chaos*, vol. 8, pp. 1259-1284, 1998.

[18] S. N. Al Saad and E. Hato, "A speech encryption based on chaotic maps," *International Journal of Computer Applications*, vol. 93, 2014.

[19] Y. Zhai, S. Lin, and Q. Zhang, "Improving image encryption using multi-chaotic map," in *Power Electronics and Intelligent Transportation System, 2008. PEITS'08. Workshop on*, 2008, pp. 143-148.

[20] Y. Wang, G. Ren, J. Jiang, J. Zhang, and L. Sun, "Image encryption method based on chaotic map," in *Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference on*, 2007, pp. 2558-2560.

[21] X. Tong and M. Cui, "A novel image encryption scheme based on feedback and 3D Baker," in *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM'08. 4th International Conference on*, 2008, pp. 1-4.

[22] S. S. Tirumala, S. R. Shahamiri, A. S. Garhwal, and R. Wang, "Speaker identification features extraction methods: A systematic review," *Expert Systems With Applications*, vol. 90, pp. 250-271, 2017.

[23] K. L. Neville and Z. M. Hussain, "Effects of wavelet compression of speech on its Mel-Cepstral coefficients," in *international Conference on Communication, Computer and Power (ICCCP'09), Muscat*, 2009, pp. 387-390.

[24] S. Gupta, J. Jaafar, W. F. W. Ahmad, and A. Bansal, "Feature extraction using MFCC," *Signal & Image Processing*, vol. 4, p. 101, 2013.

[25] D. G. Childers, D. P. Skinner, and R. C. Kemerait, "The cepstrum: A guide to processing," *Proceedings of the IEEE*, vol. 65, pp. 1428-1443, 1977.

[26] L. Muda, M. Begam, and I. Elamvazuthi, "Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques," *arXiv preprint arXiv:1003.4083*, 2010.

[27] A. I. Galushkin, *Neural networks theory*: Springer Science & Business Media, 2007.

[28] G. Dreyfus, *Neural networks: methodology and applications*: Springer Science & Business Media, 2005.

[29] N. W. Evans, J. S. Mason, W. M. Liu, and B. Fauve, "An assessment on the fundamental limitations of spectral subtraction," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, 2006, pp. I-I.

[30] P. Krishnamoorthy and S. M. Prasanna, "Enhancement of noisy speech by spectral subtraction and residual modification," in *India Conference, 2006 Annual IEEE*, 2006, pp. 1-5.

[31] M. Abd El-Fattah, M. I. Dessouky, S. M. Diab, and F. E.-S. Abd El-Samie, "Speech enhancement using an adaptive wiener filtering approach," *Progress in Electromagnetics Research*, vol. 4, pp. 167-184, 2008.

[32] B. Macq, J. Dittmann, and E. J. Delp, "Benchmarking of image watermarking algorithms for digital rights management," *Proceedings of the IEEE*, vol. 92, pp. 971-984, 2004.

[33] S. Et-Khamy, M. Hadhoud, M. Dessouky, B. Salam, and F. A. El-Sarnie, "Sectioned implementation of regularized image interpolation," in *Circuits and Systems, 2003 IEEE 46th Midwest Symposium on*, 2003, pp. 656-659.

[34] S. El-Khamy, M. Hadhoud, M. Dessouky, B. Salam, and F. A. El-Samie, "Optimization of image interpolation as an inverse problem using the LMMSE algorithm," in *Electrotechnical Conference, 2004. MELECON 2004. Proceedings of the 12th IEEE Mediterranean*, 2004, pp. 247-250.