New Electronic Business Reforming—Study on New Voice Based Biometric Security System

Tianyi Luo

Department of Electrical and Computer Engineer, College of Engineering, The Ohio State University, Columbus, USA
Email: luo.390@osu.edu

Abstract

Along with the rapid development of informatization, people habitually rely on electronic devices. For the convenience and efficiency provided by various devices, people could dispose almost everything by their phones and PC. In this respect, the security of the electronic devices that people use in daily life becomes much more important. In this area, exerts has tried many ways to keep data and uses safe from attacks. This paper has designed a Voice Based Biometric Security System, which is able to reinforce the security level of the target device. With its help, the electronic devices will be more security.

Keywords

Electronic Devices, Biometric Technology, Voice Recognition

1. Different Protections in Different Stages

The method of providing protection for a specific goal changes over time and technological innovation. Through the development of the security, there are three degrees of level of security [1]. The first degree is the physical lock. At the beginning, people used to set the security with what they physically possessed. For example, people use keys to lock doors and use digital cards to enter buildings [2]. It was safe since there is only one matched key to the traditional lock or digital lock. The second degree is the intellectual access control. To improve the security degree and cater the new information era, people started to utilize what they intellectually know to encrypt for instance password [3]. Certainly, the password encryption has so far become a basic guarantee for people's daily life information security. However, the things people owned could be misplaced intentionally or unexpected, and the passwords could be forgotten as well. In the situation where the information security is becoming more and more serious, those two encipherments seem to be far from good enough [4]. People need

another way, a method which is unlikely to be taken or used by other people and does not require extra efforts neither intellectually nor physically, to efficiently strengthen the security.

The third degree of the level of security is the biometric authentication. It uses biometric identifiers which are the distinctive, and measurable characteristics used to label and describe individuals to identify people. It is the most strength and efficient selective restriction of access method so far. It would never be lost or forgotten and nearly impossible to be stolen by anybody [5]. As a result, a Voice Based Biometric Security System becomes necessary. The purpose of the security system is to identify users by their individual behavioral characteristics, or their voice, and permit access for protected files and software, only if they passed the security check. The only key to the biometric locker is based on the pattern of behavior of the user [6].

2. System Functional Design Specifications

The purpose of the voice based biometric security system is to reinforce the security level of the target device. There are three parts of this Voice Based Biometric Security System, which are biometric access control, user database, user interface.

The access control contains a voice recognizer, a parameter analyzer, and a voice comparator software which is based on the MATLAB on Windows 7 operational system. The main analyze method to be used in this design is Mel-frequency Cepstral Coefficients (MFCC), which the latest voice recognition analyzing method. Difference for this method from other similar standard such as LPCS and LPCCs is that MFCC more emphasize on analyzing how the human ear could receive the voice signal information rather than the signal itself [7]. The voice signal from the user will get through MFCC filter which attenuate most signal that cannot be heard by human ear because of too high frequency, and finally display the MFC coefficient value in time domain. This idea makes the analyzing process highly comfortable for comparing MFCC image with voice magnitude signal. Real situation testing will be held to know reliability of the system and give the result.

The identification process would require the users to speak a word or short phrase to the microphone and compare with the record parameter previously stored in the system [8]. The system will identify voice pattern and verdict whether the attempter has the license to the restrict area and the result will be prompted out through GUI. The system would not permit access if users speak different word or phrase from their first time recording stored in the system. Even though it would sacrifice the convenience of the system, the amount of computation is also a main factor which would affect login speed. To have a higher level of security, the security system is designed to have more than one encryptions. The password pin would be deployed inside the system beside the input voice.

Databases are used to store users' profile including username, passwords, and voice comparison key. It is default to allocate space for 500 user profiles but could be modified volume by employer's choice. There is no limitation for anyone to sign up as a user of the system, but the existed user information is accessible for administrator only. Lastly, to simply the instruction, the system has a Graphic User Interface (GUI) built based on the same platform, MATLAB. User could efficiently follow the steps to login and signup to the security system [9].

The complete system runs on Windows 7 platform. It allows up to 500 users to save their preferences in the database, so that these users could login the system anytime. The biological identity will not be determined based on the password, but the users cannot proceed to the voice identification process until correct account and password information are provided. GUI user interface will be included in the system to guide the user. To have high accuracy result from the designed system, the system is recommended to be used under a low noise condition with user speak their own given word.

3. System Implementation

– Graphic User Interface

At the beginning, the author started hand-drawing the initial design of the Graphic User Interface (the GUI) and brainstormed the function the GUI would have to satisficed the whole system's goal. The author came with several different drawings of the initial layout of the GUI windows, and finally decided that the whole system would include five different windows inside the GUI to deliver the best performance to the customer. The interfaces in different situations are set as follows:

In the first window, shown in Figure 1, users could enter the username and password for their account, and able to create a new account if needed. In second interface as shown in Figure 2, the system would ask the user to speak their voice key and match it with the audio reference inside the system's database.

Those two windows, shown in Figure 3 and Figure 4, would appear to the user terminal only if both the password and voice key is inputted to the system. After verification, the system would display result to user through GUI.

This window, shown in Figure 5, would be applied whenever the user would like to sign up a new account to access the system. It would ask user for a unique username for their account and set the validation information, the password and voice key. The system would not create any account if the username is not unique in the system or the password has not been entered differently when it is being confirmed.

– Voice Recognition

In this model, two systems are developed as Speaker Identification and Speaker Verification. The block diagrams of both systems are shown in Figure 6 and Figure 7.
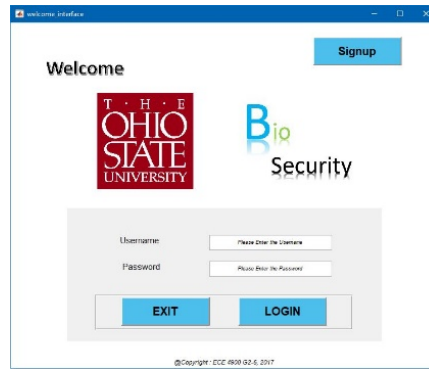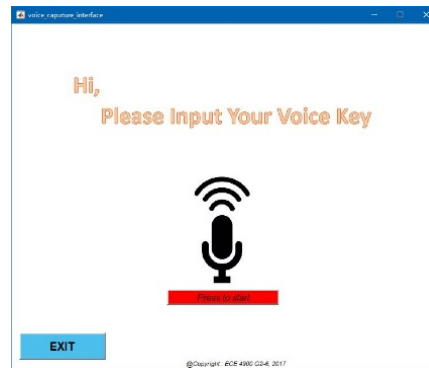
Figure 1. GUI Login & Welcome Interface.

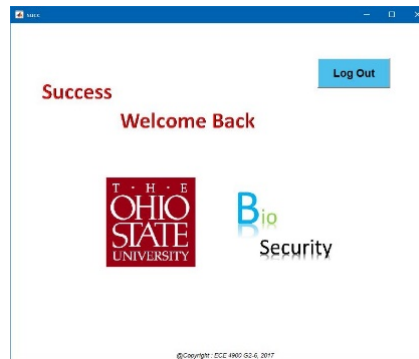**Figure 2.** GUI Voice Capture Interface.

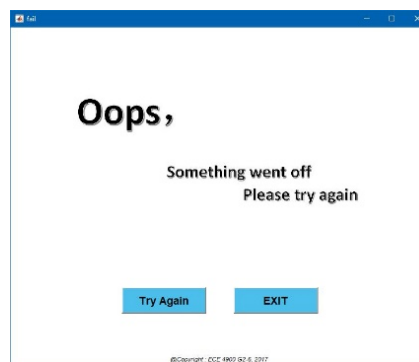**Figure 3.** GUI Login Success Interface.

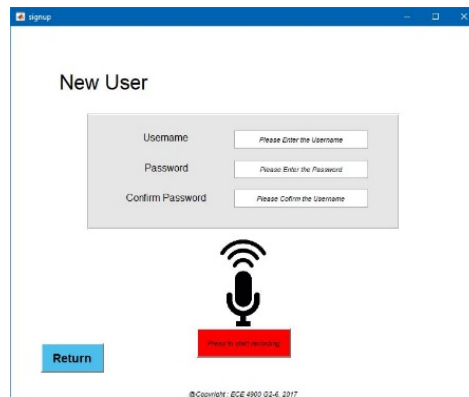**Figure 4.** GUI Login Failure Interface.

Figure 5. GUI New User Signup Interface.

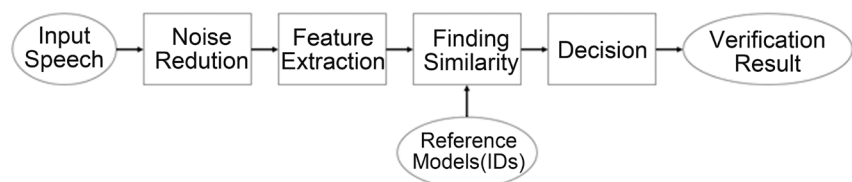**Figure 6.** Speaker enrollment session.

Figure 7. Speaker verification session.

The Speaker Enrollment Session includes Speech Input, Noise Reduction, Feature Extraction, and Speaker Enrollment. In the Speaker Enrollment Session, each user is enrolled with specific parameters which are derived from the Feature Extraction phase. Also, the same specific parameters will be used in the Speaker Verification Session, especially in the Finding Similarity phase. Other specific parameters which are derived from the Speaker Verification Session will be compared with each registered reference model. In addition, in the Speaker Verification Session, the system will decide whether the user gains access to certain profiles or not, which depends on a defined tolerance level for similarity between registered parameters and the newly inputted signal.

In the frequency domain, the proposed design includes a feature extraction algorithm as shown in Figure 8. In this algorithm, the most important part is Mel-Frequency Warping. Mel Frequency Warping will allow us to obtain better representation of sound, especially in audio data compression. It contains several steps shown below:

1) Fourier transform;

2) Mapping the power spectrum using triangular overlapping;

3) Logging the powers at each of the mel frequencies;

4) Discrete cosine transform.

After performing the above four steps, the amplitudes of the resulting Mel

Figure 8. Feature extraction block diagram.

Spectrum shows Mel-Frequency Cepstral Coefficients (MFCCs). Also, these coefficients will be used as characteristic parameters for each user.

3.1. Alternate Designs or Issues and Solutions Considered

The most important phases for the entire system are the noise reduction and the feature extraction phases. There are several concepts and issues to be considered in developing phases for time and frequency domains, such as, noise reduction, different beginning time of voice input, loudness information, and methodological approach in the frequency domain. Also, these concepts and issues should be handled before or during feature extraction for both time domain and frequency domain.

3.2. Time Domain Analysis

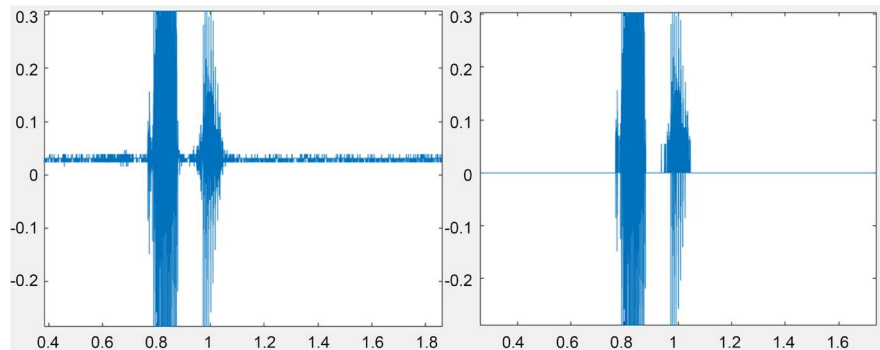### 3.2.1. Background Noise Reduction

Background noise is a major factor contributing to the difficulties in analyzing voice signals. This is attributed to the fact that background noise contributes unintended parameters from the surroundings that becomes unwanted information in the speech input. As shown in Figure 9 the graph (left) includes very low amplitudes of the input signal during the entire recording period. Thus, in the system, very low amplitudes of the signals are considered as background noise, which are removed before performing any other steps as shown in Figure 9 (right). In addition, this would allow the system to have better tolerance of variations like extraneous conditions.

3.2.2. Different Beginning Time of Speech Input

In the system, any user can say the given words at any different time. As shown in Figure 10 (left), one user started saying "apple" at 0.7 second for his reference model. However, as shown in Figure 10 (right), the next time, the user started saying "apple" after 2 seconds. In this case, these two voice signal graphs are not useful for comparison due to the initial time difference. Thus, the author decided to set the region of interest by discarding the unnecessary information and shifting the time domain. Using the MATLAB program, the author obtained time shifted speech signals as shown in Figure 11.

3.3. Frequency (Sampled) Domain Analysis

In order to extract features from the speech signal, time domain analysis is not sufficient. Thus, it is necessary to consider feature extractions in the frequency domain. In sound processing, Mel-Frequency Cepstral Coefficients (MFCC) is the most popular method. This is due to the fact that "MFCCs are known variation

Figure 9. Test model (left) and background noise reduced model (right) of user A.
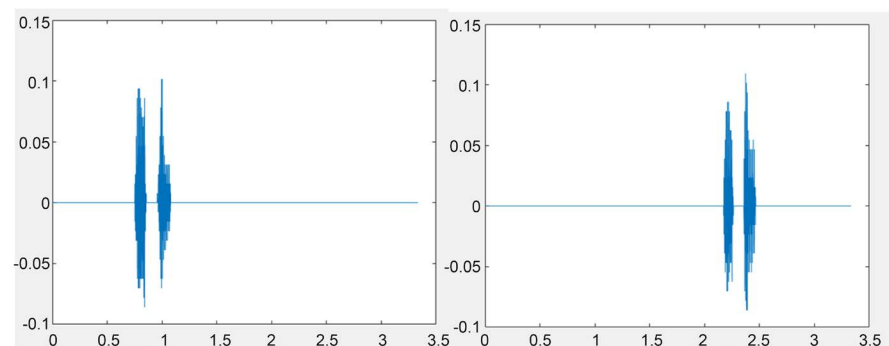
**Figure 10.** Reference model (left) and test model (right) of user B.
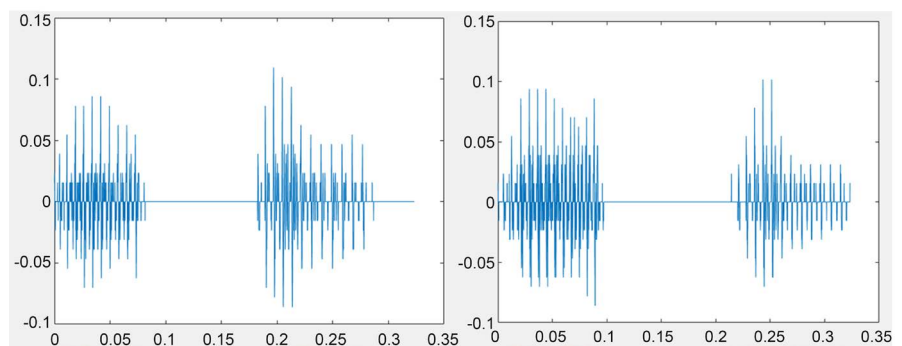
Figure 11. Reference model (left) and test model (right) after time shifting.

of the human ear's critical bandwidths with frequency filters spaced linearly at low frequencies and logarithmically at high frequencies have been used to capture the phonetically important characteristics of speech (Ehkan *et al.*, 2014)". In this respect, a sub-system to obtain Mel-Frequency Spectrum (MFC) and MFC Coefficients to analyze features of voice signals in the frequency domain are built. However, the author faced several issues in the frequency domain.

3.3.1. Another Noise Removal
Entirely scattered background noise has already been reduced in the time domain analysis. However, the microphone which the author bought has several issues. First of all, as shown in **Figure 12**, at times, when the microphone starts recording the voice, it produces unintended noise at the beginning. When the
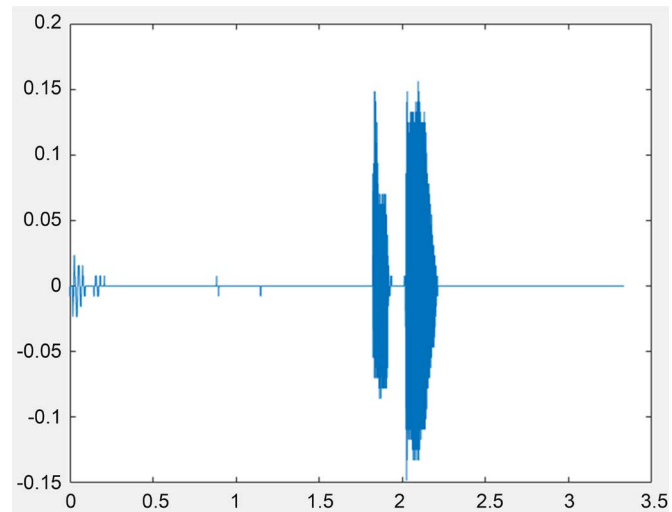
Figure 12. Another noise removal produced by the microphone.

noise is very low, it is removed automatically by the background noise reduction code that the author built. However, when it is higher than the defined value in the code, it still exists. Also, this noise hinders the analysis of the Mel-frequency Cepstrum. In order to solve this problem, the author manually removed all the input values for 0.5 second at the beginning.

3.3.2. Different Length of Mel-Frequency Cepstral

After obtaining MFC Coefficients, the author faced a huge problem. In order to compare two different coefficients matrices, both matrices should have same lengths. However, people can say a given word with a different speed at any time.

As shown in Figure 13 and Figure 14, Figure 13 has a length of 18 but Figure x has a length of 24. It means that the user speaks faster in Figure 13 than in Figure 14. In this case, one of them should be stretched out to have the same length.

Figure 15 is the MFC Coefficient matrix stretching code that the author built. First, the lengths of both matrices are obtained in line 23. Through line 24 to 27, the algorithm compares each length of the matrix and then the shorter length of the matrix is stretched out to the same length as the longer matrix. Through line 29 to 31, if both matrices already have the same lengths, they maintain their original lengths. From now on, they have the same lengths meaning that they are ready to be compared (see Figure 16).

3.4. Analysis and Applicability of Constraints

－ Voice Recognition

Both loudness and frequency pattern will be sampled and calculated to have an average similarity with the stored information in the database. When error is higher than the tolerance level, the users will be treated as different people and cannot access to the security data. Password pins are widely used for security systems. However, it does not provide a strong identity check. In contrast, since
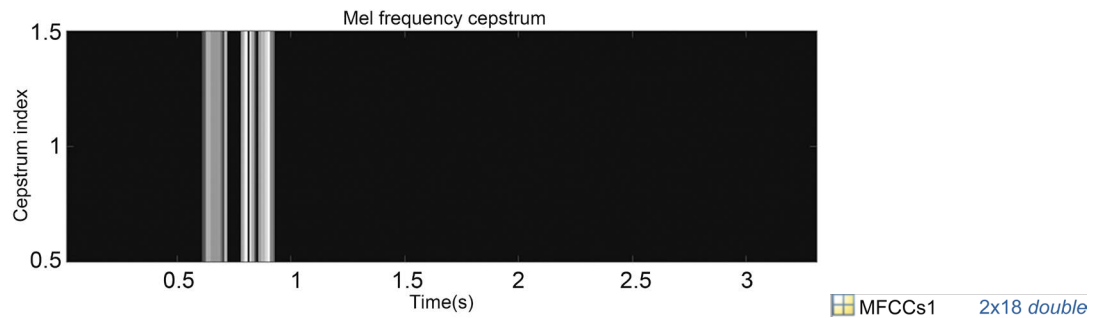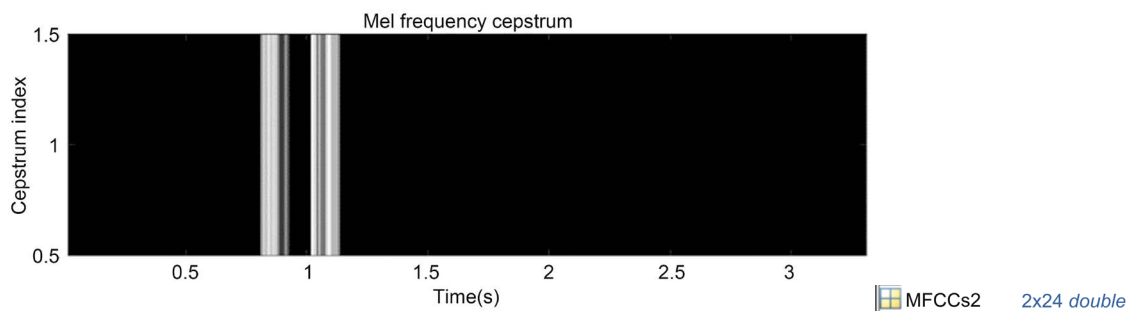
Figure 13. MFC Coefficient example 1.

Figure 14. MFC Coefficient example 2.

```
23 —      l1 = length(MFCCs1);   l2 = length(MFCCs2);
24 —      if   l1>l2
25 —          MFCCs2=imresize(MFCCs2,[2,l1],'nearest');
26 —      else if   l1>l2
27 —          MFCCs1=imresize(MFCCs1,[2,l2],'nearest');
28
29 —      else
30 —          MFCCs1=MFCCs1;
31 —          MFCCs2=MFCCs2;
32 —      end
33
34 —      comp = MFCCs1-MFCCs2;
```

Figure 15. MFC Coefficient matrix stretching code.

```
MFCCs1    2x24 double
MFCCs2    2x24 double
```

Figure 16. Stretching result example.

every person has a specific voice, voice recognition security systems can be one of the higher level security systems. Voice recognition systems only require an operating system and a microphone. Thus, this system is cost effective. However, the voice-based biometric security system is really sensitive to background noise. In Figure 17, the user said a specific word between 0.7 to 1.3 second. However, as shown in the graph, unnecessary background noise is scattered throughout
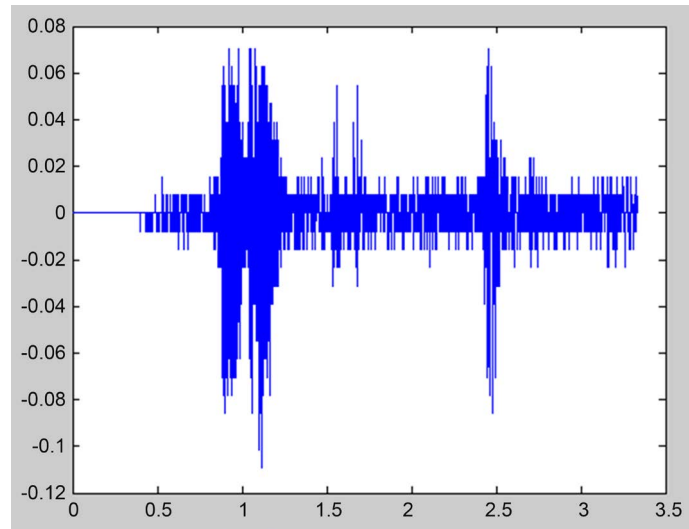
Figure 17. Background noise.

the entire recording. Also, in this case, although the author has built two noise reduction algorithms as mentioned before, there are uncontrollable noise present coming from the recording environment. In other words, the proposed system is really sensitive to noise. Thus, if the voice recording is performed in a noisy place, we cannot extract any important feature information. Therefore, for the entire project and further developments, this constraint should be considered.

4. Validation and Future Improvements

As result, the author successfully build all the windows as designed based on the MATLAB, but the system is not completed without compiling all parts together including the GUI, the Voice Algorithm and the database, which could contain up to 500 users' profile (Figure 18).

To test all program, the author has to first connect the Voice Algorithm with the GUI. The author uses the Callback function again to link those two individual parts together as a system. With several tests run, the author manages to confirm the GUI functional ability to perform the work. Also, the instruction given in the GUI is clear for any users including those with less experience dealing the high tech to follow the procedure.

To obtain validation results, the author performed two validation experiments. One is the comparison results in which each user's reference voice is compared to their own voice input (see Table 1). The other is the comparison results in which each user's reference voice is compared to a different user's voice input (see Table 2). All comparison result values are the summation of all errors from each element of two MFC Coefficients matrices.

As a result, when each user's reference voice is compared to different user's voice inputs, the average errors are slightly higher than when it is compared to their own voice input. These small differences are much lower than the author
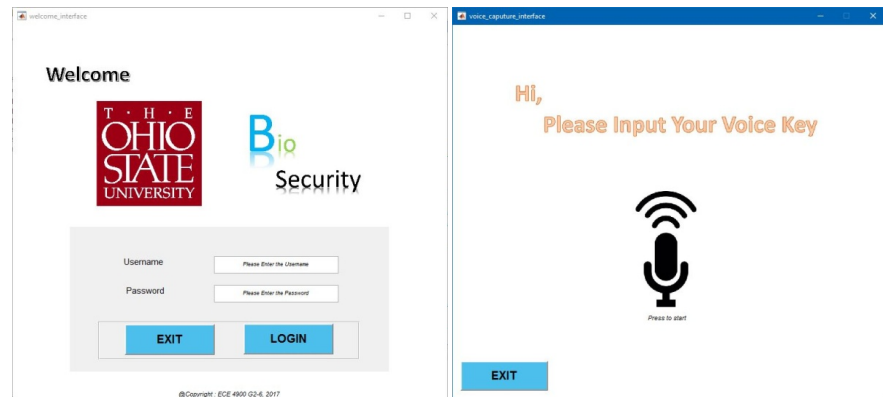
Figure 18. Examples of the GUI windows.

Table 1. Comparison result 1, compared to their own voice input.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Avg Error
User 1	1.4	1.24	0.56	1.92	1.10	1.57	1.54	1.33
User 2	1.73	1.76	1.52	2.19	2.10	2.30	1.64	1.89
User 3	0.45	0.09	0.75	1.50	0.80	0.82	0.98	0.77
User 4	0.46	2.32	0.31	0.58	1.63	2.65	1.39	1.33
User 5	0.58	0.33	0.38	0.50	0.48	0.38	0.39	0.43

Table 2. Comparison result 2, compared to different user's voice input.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Avg Error
User 1	3.96	3.96	3.25	4.95	3.34	4.28	4.09	3.97
User 2	2.61	2.59	2.15	2.11	0.16	2.39	2.27	2.04
User 3	2.76	2.52	2.25	1.15	1.68	1.78	2.27	2.06
User 4	2.93	2.59	2.03	2.88	1.48	2.16	0.49	2.08
User 5	1.97	1.67	2.06	1.98	1.25	1.57	0.85	1.62

initially expected. Also, these results hinder the author from defining a tolerance level for what is a match or not. Therefore, this constraint is considered as the risk of our proposed system.

There are a few possible reasons why comparison results do not have remarkable differences and why our proposed system has a lower security level.

4.1. Matrix Stretching Error and Its Considered Solution

The first possible reason is the matrix stretching error. As shown in **Figure 19**, when we stretched one of the MFC Coefficient matrices, MATLAB just copies some columns right after their original position to match its length. This characteristic has huge impact when we compare two matrices. This is attributed to the fact that each MFC Coefficient element has its meaning, but the copied columns allow the coefficients to have erroneous information. Although the author has tried the same approach by squeezing the length of the longer matrix to that
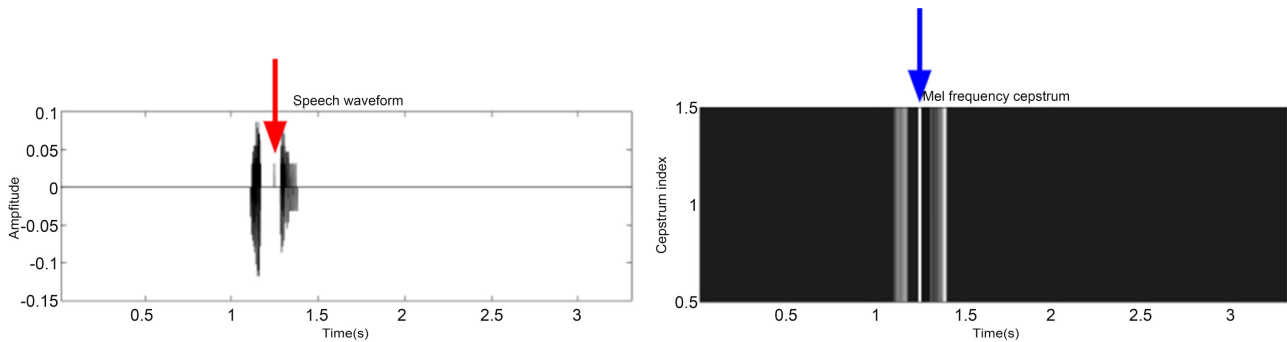
■ Before Stretching

1	2	3	4	5	6	7	8	9	10
0.9992	0.9986	0.9956	0.9958	0.9966	0.9970	0.9924	1.0000	0.9997	0.9993
-0.0397	-0.0537	-0.0936	-0.0913	-0.0823	-0.0780	-0.1234	0.0013	0.0259	0.0377

■ After Stretching

1	2	3	4	5	6	7	8	9	10
0.9992	0.9992	0.9986	0.9956	0.9956	0.9958	0.9966	0.9966	0.9970	0.9970
-0.0397	-0.0397	-0.0537	-0.0936	-0.0936	-0.0913	-0.0823	-0.0823	-0.0780	-0.0780

Figure 19. Matrix stretching error example: (1) Before stretching (upper), (2) After stretching (lower).

Figure 20. Uncontrollable noise error example: (1) Voice signal (left), (2) Mel-Frequency Cepstrum (right).

of the shorter one in order to solve this problem, it loses more information than the stretching method. Thus, the author has decided to keep using the stretching method. However, this problem should be resolved to have a better security level.

To overcome this challenge, the author decided not to give the same word for every user. Each user can say any one or two words in their tastes. This option will increase the security level of our proposed system.

4.2. Uncontrollable Noise Error

The second reason why our proposed system has lower security level: when we record the voice input in silent environments, occasionally, there is an unintended noise between the two syllables. In **Figure 20**, the noise is indicated by the red arrow (see **Figure 20** (1)). In this case, although we can handle this noise manually, we do not know when and where this noise occurs. Thus, we cannot remove this noise automatically. This problem negatively affects the security level. This is due to the fact that as shown in **Figure 20** (2), it produces an unnecessary spectrum, indicated by a blue arrow. Therefore, this problem should be considered as a risk of our proposed system.

5. Conclusion

The design for biometric security system is successful built by the parts mentioned above. According to the result from final demonstration, the system successfully identified one of our testers in a quiet room. However, when environ-

mental noise becomes relatively higher, the system failed. From the success rate during the overall testing, it showed that the system still needs further development on reliability such as more feature extraction process from the voice signal.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Fong, S. (2012) Using Hierarchical Time Series Clustering Algorithm and Wavelet Classifier for Biometric Voice Classification. *Journal of Biomedicine & Biotechnology*, **2012**, Article ID: 215019. https://doi.org/10.1155/2012/215019

[2] Dustor, A. and Kłosowski, P. (2013) Biometric Voice Identification Based on Fuzzy Kernel Classifier. In: Kwiecień, A., Gaj, P. and Stera, P., Eds., *Computer Networks. CN* 2013. *Communications in Computer and Information Science*, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38865-1_46

[3] Mazairafernandez, L.M., ÁlvarezMarquina, A. and Gómez-Vilda, P. (2015) Improving Speaker Recognition by Biometric Voice Deconstruction. *Frontiers in Bioengineering & Biotechnology*, **3**, 126. https://doi.org/10.3389/fbioe.2015.00126

[4] Motwani, R.C. (2010) A Voice-Based Biometric Watermarking Scheme for Digital Rights Management of Three-Dimensional Mesh Models. University of Nevada, Reno.

[5] Al-Sarayreh, K.T., Al-Qutaish, R.E. and Al-Majali, M.D. (2008) Incorporating the Biometric Voice Technology into the E-Government Systems to Enhance the User Verification. *WSEAS Transactions on Computers*, **7**, 435-452.

[6] Motwani, R.C., Dascalu, S.M. and Harris, F.C. (2010) Voice Biometric Watermarking of 3D Models. 2010 2*nd International Conference on Computer Engineering and Technology* (*ICCET*), Chengdu, 16-19 April 2010, 17.

[7] Kounoudes, A., Kekatos, V. and Mavromoustakos, S. (2006) Voice Biometric Authentication for Enhancing Internet Service Security. 2006 2*nd International Conference on Information & Communication Technologies*, Damascus, 24-28 April 2006, 45.

[8] Chenafa, M., Istrate, D., Vrabie, V. and Herbin, M. (2008) Biometric System Based on Voice Recognition Using Multiclassifiers. In: Schouten, B., Juul, N.C., Drygajlo, A. and Tistarelli, M., Eds., *Biometrics and Identity Management*, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-89991-4_22

[9] Rudrapal, D., Das, S., Debbarma, S., Kar, N., Debbarma, N. and Rudrapal, D., *et al.* (2012) Voice Recognition and Authentication as a Proficient Biometric Tool and Its Application in Online Exam for P.H People. *International Journal of Computer Applications*, **39**, 6-12. https://doi.org/10.5120/4870-7297