

Research Article

Open Access

Multimodal Biometrics Based on Identification and Verification System

Osamah Al-Hamdani¹, Ali Chekima¹, Jamal Dargham¹, Sh-Hussain Salleh^{2*}, Fuad Noman², Hadri Hussain², AK Ariff² and Alias Mohd Noor²

¹Computer Engineering Program, School of Engineering and Information Technology, University Malaysia Sabah, Jalan UMS 88400, Kota Kinabalu, Sabah, Malaysia

²Center for Biomedical Engineering, Universiti Teknologi Malaysia, 81300 Skudai, Johor, Malaysia

Abstract

The need for an increase of reliability and security in a biometric system is motivated by the fact that there is no single technology that can realize multi-purpose scenarios. Experimental results showed that the recognition rate of Heart Sound Identification (HSI) model is 81.9%, while the rate for Speaker Identification (SI) model is 99.3% from 20 clients and 70 impostors. Heart Sound-Verification (HSV) provides an average Equal Error Rate (EER) of 13.8%, while the average EER for the Speaker Verification model (SV) is 2.1%. Electrocardiogram Identification (ECGI), on the other hand, provides an accuracy of 98.5% and ECG Verification (ECGV) EER of 4.5%. In order to reach a higher security level, an alternative multimodal and a fusion technique were implemented into the system. Through the performance analysis of the three biometric system and their combination using two multimodal biometric score level fusion, this paper found the optimal combination of those systems. The best performance of the work is based on simple-sum score fusion, with a piecewise-linear normalization technique which provides an EER of 0.7%.

Keywords: Unimodal biometric; Heart signal detection; Health care security; Speaker verification

Introduction

Biometric recognition is an automatic identification and verification of a person based on his or her physiological behavioural characteristics. Biometric include, facial imaging, fingerprints, hand geometry, signature, and voice [1-5]. Alternative to this traditional approach is medical biometric, in which the personal medical features with different formats, such as image signals, Blood Volume Pressure (BVP), pulse oximetry, ECG and Phonocardiogram (PCG), exhibit identity discrimination power [6-8]. The traditional approach of the biometric authentication system is well established, but suffers mainly on defenceless nature against falsification [9]. It is argued here that it is hard to fabricate the ECG, as well as difficult to mimic the signal. The advantage of this bio signal is that it gives some sort of liveliness measurement to avoid spoof attack, which is not inherent in the traditional biometrics. Another signal that is related to the heart is the recording movements of the heart valves (mitral, tricuspid, aortic, and pulmonary), that produce heart sounds. Phua et al. [7] validated exponentially uniqueness of the heart, where two heart sounds form different characteristics. In a control environment, the variability of the heart is constant, and achievable performance can be obtained from this signal. The ECG and PCG signals are not only useful for medical purposes, but can also be applied for biometric identification and verification. These bio signals are useful as it can be easily combined with the traditional biometric, such as speech, to provide a liveliness evaluation, without any additional cost. There is no single technology that can claim to provide best performance in all its applications, even though, there are some technologies which are matured and available for applications, but that does not mean they are flawless. Thus, the work presented here uses a multimodal biometric recognition based on speech, ECG, and PCG. The advantages of using such input sensors are as follows:

- Use of multimodal and fusion techniques enhances performance
- The usage of speech and its capability of leveraging telephony infrastructure, with the ever growing numbers of cellular phone and system complexity; speaker recognition have a promising future
- Speech is more reliable to spoof attack, while the PPG and ECG are not. Furthermore, both of the biosignals are not easily mimic, and

guaranteed liveliness

The results also clarify the effectiveness of the use of a multimodal biometrics approach to the Automatic Client Recognition (ACR) task. However, further investigation using ECG for an individual revealed obvious characteristic that may not be present in recordings from other individuals. Agrafioti [10] proposed a fiducial feature extraction algorithm, with a 12 lead ECG system. The system was evaluated with 20 subjects, varying from age. The fiducial features used was P wave onset duration, QRS and set duration, QRS wave deflection and ST segment. With the 12 lead ECG, there are 360 (12*30) features obtained from every subject, with different combination of features. Leads were also studied using the best speaker identification bearing the accuracy of 100% obtained from 50 subjects. Su et al. [11] developed an algorithm that is capable of detecting atrial fibrillation episode in random ECG automatically. Patient with similar cardiac arrhythmia usually have similarities in their ECG characteristics. Template based algorithm were developed to find a positive match between the template created and the ECG tested. Before the system could detect the peak, a threshold value was applied to determine which part of the ECG can be considered as the peak value. The threshold value was calculated using the max value and the mean value of the ECG. The window length was based on an approximate method used through testing a number of different ECG. The PR interval is set to 0.12 to 0.2 seconds long, with the QT interval of 0.40 second and less. The proposed algorithm is capable of detecting positive fibrillation episodes, with an accuracy of just 78.6%. Shen et al. [12] proposed a speaker identification scheme based on 7 features trained with a neural network classifier. Their assumption is that QRS complex is less affected by varying the heart rates, thus it

***Corresponding author:** Sh-Hussain Salleh, Center for Biomedical Engineering, Universiti Teknologi Malaysia, 81300 Skudai, Johor, Malaysia, Tel: +607-5535208; Fax: +607-5535430; E-mail: hussain@fke.utm.my

Received December 30, 2012; **Accepted** March 08, 2013; **Published** March 14, 2013

Citation: Al-Hamdani O, Chekima A, Dargham J, Salleh SH, Noman F, et al. (2013) Multimodal Biometrics Based on Identification and Verification System. J Biomet Biostat 4: 163. doi:[10.4172/2155-6180.1000163](https://doi.org/10.4172/2155-6180.1000163)

Copyright: © 2013 Al-Hamdani O, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

is appropriate for a design of a robust system to stress the condition framework. The fiducial features consist of RQ amplitude, QS duration, ST amplitude, QRS triangle area, RS amplitude, QT duration and the RS stage. There were 20 subjects taken from the MIT-BIH database. The system performs with an accuracy of 95% for template based, and 80% for decision based neural network (DBNN). The work carried out in this paper is also based on 12 features vector, consisting of magnitude and duration only. Israel et al. [13] considered the slope, as well as the magnitude and duration with 12 fiducial points, while Shen et al. [12] utilized 7 fiducial features, mostly related to the QRS complex. The approached related to fiducial point's detection was not given by these authors. The algorithm proposed here is designed towards detecting the fiducial points of ECG created from ECG wave shape, which is discussed in detail later. The new proposed algorithm is mainly based on the theory and observation mode of the ECG wave-shapes.

The Automatic Client Recognition (ACR) task using large amounts of labelled speech data is not only time consuming, but also requires a huge amount of training data for better generalization of Neural Network (NN) structure and the training methods. On the other hand, hidden Markov model (HMM) [14,15] based system shows better performance than the Vector Quantization (VQ) base, or the (NN) approach for SV tasks. The benefit, however, must be weighed against the high computational requirements using HMM during training, as more features are added into the processing stage. Thus, the design of a uni-multimodal biometric system, as proposed here, used these specific criteria to achieve the results:

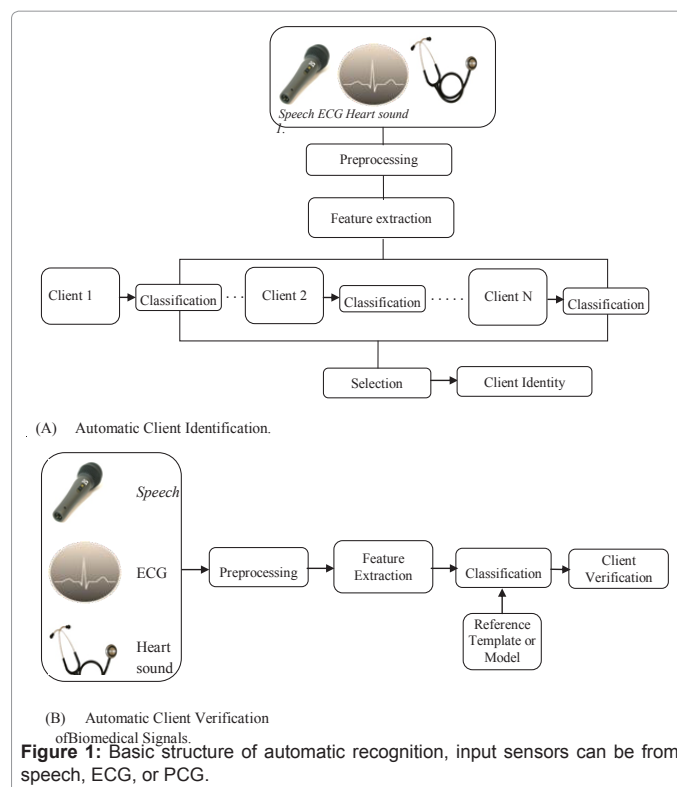
- i. Use of simple algorithm for ease of implementation, and without an increase in computational effort.
- ii. Capable of achieving high performance with limited training data.
- iii. Reasonable handling of the temporal information.

Design of Uni-multimodal Biometric System

Identification/verification system

The basic structure of Automatic Client Recognition (ACR), can be seen in figure 1, where the input sensors can be from speech, ECG or PCG. The elements of the client recognition systems are feature extraction and classification. The feature extraction analysis computes a set of parameters from the speech, ECG or PCG signal. This initial stage is common to the learning, as well as the test operating phase. Sometimes modification is done at this stage which extracts the high level features, aimed at representing a client with limited number of features. These new features are compiled into training and test data structure, which is more economically represented. These data can then be used to build and test the client recognition model for the client. The next stage will perform the classification, which involves the comparison of the feature vector derived from the unknown client, with the reference vector with a threshold set, the resulting distance above or below this threshold will determine the end result.

The fundamental difference between the Automatic Client Identification (ACI) and Automatic Client Verification (ACV) is the decision alternatives. In ACI, the numbers of decision alternatives increases as the population size increases, while there are only two decisions to be made for the ACV task. The ASV system has only to reject or accept the unknown Client, regardless of the population size. Looking at this perspective, this makes ACI a more difficult task as performance decreases, when the population size increases. In



either case, when the unknown speaker does not match the model, an additional threshold can be set to determine whether the decision is close to be accepted or a retrieval is needed [16]. Several reviews of this field have already appeared, such as Rosenberg [15], Rosenberg et al. [16], Bennani et al. [17], Ting et al. [18], Ting et al. [19].

The work presented in this paper has three biometric systems, where the input sensors are based on speech, (PCG), and (ECG). Further enhancement of this multimodal biometric system is the use of two level fusion score. As for speech, the word speaker was used to present the identity for identification or verification. Speaker recognition is used to identify a talking person automatically. It can be divided into two main areas, Automatic Speaker Identification (ASI) and Automatic Speaker Verification (ASV). The Speaker Identification process is used to test speaker by comparing the results with the multiple registered speakers, using specific information retained in the speech. The claimed identity is checked by automatic means based on the acoustic of his or her sound, which act as an automatic speaker verification. Both of ASI and ASV systems will process the raw speech data into the input features, and then be compared to the specific classification techniques identified or verified. Thus, identifying an unknown speaker by an automatic means can be termed as Automatic Speaker Recognition (ASR). In order to avoid confusion, the word client is used to represent an individual to be identified or verified with different input sensors. First, the design of uni-modal biometric system was evaluated and compared. Then, the design of the multibiometric system, with one that provide actual biometric data (speech) and the other which give liveliness (ECG, PCG), have two different input sensors to the same biometric system. Finally, the output score of each of the biometric system was fused and evaluated, providing the following benefits:

- Enhanced in system performance behavioural
- Avoid failure to enrol

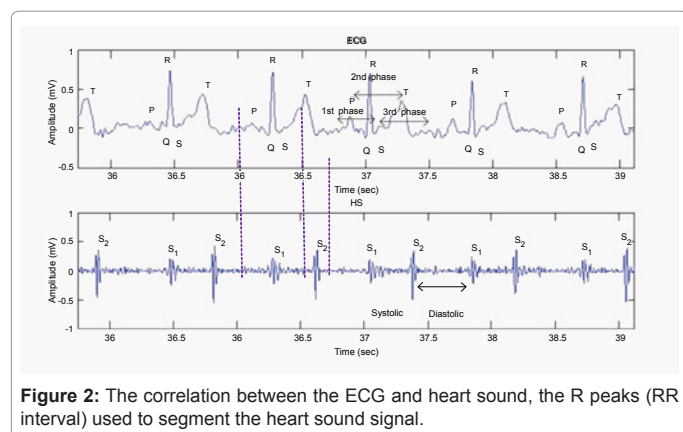


Figure 2: The correlation between the ECG and heart sound, the R peaks (RR interval) used to segment the heart sound signal.

- Overcome the system to spoof attack
- Increase the universality factor

Heart sound segmentation

Two channel hardware electronic designs are used to extract ECG and heart signal, simultaneously. The importance of ECG [20] signals is that it cannot be overemphasized, as local cardiologists rely on these signals to record heart sounds of S_1 and S_2 , in order to determine the systolic and diastolic murmurs. The P-wave is linked with the blood that is being pushed by a trial contraction into the lower ventricle chambers. The QRS complex composes of Q, R and S peaks. It originates from ventricle depolarization that triggers contraction. This process allows the blood to be pushed out of the heart, into the arterial vessels. This information can be extracted during the cardiac cycle of S_1 and S_2 sounds, where S_1 relates to the closing of the mitral M1 and the tricuspid valve T1, while S_2 is related to the plutonic and aortic valve. S_1 corresponds to the QRS complex and the T-wave that is related to S_2 cycle, shown in figure 2. Manual segmentation can be a tedious task, and would not be reliable in real practice. Automatic segmentation of the heart signals into individual cycles provides useful information, such as S_1 , S_2 , systolic and diastolic, that plays a significant role to this work. The ECG signal characteristic of R to R wave is used to determine the one-minute cycle lengths of both the ECG and heart sound data.

Database

Collecting heart sound database is important in the development of an automatic heart diagnosis system. The database collected at the Centre for Biomedical Engineering (CBE) is used in clinical work, research and teaching of cardiac auscultation. The heart sounds are recorded using phonocardiography sensors amplified with a pre-processing circuit. These signals are then digitized and stored. However, care must be taken in collecting data, in order to ensure no data is corrupted with unwanted noise or artefacts, which can affect the performance of the system. The collected sample of heart sounds can also be corrupted with sudden movement of the stethoscope and from background noises.

Heart sound and ECG: The heart sound and ECG database is collected from a large numbers of subjects. There are 20 clients and 70 impostors. The database is divided into groups of training and test data. A group of 20 subjects are modelled by the system, and for heart sound, all end points data is detected with ECG segmentation. Half of the heart sound cycles are used as training data and the remaining half

is for testing. Roughly, 4700 cycles of heart sounds of impostor subjects are used for testing.

Speech: The same set of clients and impostors are used in speaker recognition. The performance of VQ varies strongly with the amount of tokens available. The database used to evaluate the system is trained with 5 training tokens of each digit for the construction of the codebook, consisting of 20 clients. The system is tested with 100 true client tokens and 100 impostor tokens; each speaker has ten digits. The database consists of Arabic isolated digits from a large number of speakers. Ten isolated digits (digits ‘Sefer-0’ to ‘Teseh-9’) are used in the experiment. Then, all endpoint data are detected to remove excess silence, and to minimize storage. The frame sizes are 20 ms with 15 ms overlap.

Feature Extraction and Vector Quantization

By examining the spectral characteristic of the speaker heart sound, additional information, such as amplitude and timing are also important. Timing information is important, as it shows events that correlate to the underlying activity of the heart and speech. An increase or decrease in the amplitude of the signal corresponds to the loudness of the heart sound and speech signals. This loudness is caused by spectral irregularities that show changes of pitch, associated with frequency shifts from average frequencies.

Feature extraction

Different techniques, such as wavelet-based approach [21], are used to solve above mentioned issues. The time frequency technique is used to segment the signals, while short windows are used to isolate signal discontinuities, and long windows are used to obtain detailed frequency analysis for feature extraction. Wavelet multi resolution is applied for better time resolution at higher frequencies and better frequency resolution at lower frequencies. An alternative to wavelet-base-approach can be the spectral feature, where the heart sound or speech record can be presented by a set of cestrum coefficients. Mel-Frequency Cestrum Coefficients (MFCC) is used in this paper as the feature representation of the heart and speech signals. Previously, the MFCC was mainly used in the field of speech processing, i.e. for speech or speaker recognition application, and has delivered excellent results due to its robustness under various conditions [22,23]. As heart sound and speech are both acoustic signals, it is possible to use MFCC in heart sound recognition or identification task. The heart signals are pre-emphasized to spectrally flatten the signals. After frame blocking, the signals are hamming windowed to minimize the spectral distortion. MFCC coefficients are then calculated by taking a Discrete Cosine Transform (DCT) of the logarithmic spectrum scale, after it is warped to the Mel scale [24].

$$\text{Mel}(f) = 2595 \log(1 + f / 700)$$

This is similar to perceptual linear predictive analysis of sound signals. In other words, the scaling mimics the human perception of distance in frequency. The number of resulting MFCCs is purposely kept relatively low, normally in the order of 12 to 20 coefficients. In this paper, 12 MFCCs per frame are used for the classification step.

Vector quantization based method

In a vector quantization scheme, the algorithm for the codebook design is executed in two stages: initialization and optimization. Vector Quantization system now has trained sets of codebooks, which serve as the main classification of the Heart Sound Verification (HSV), Speaker Identification (SI), and Heart Sound Identification (HSI). The following steps are required to train VQ codebook using LBG

algorithm. as described by Rosenberg and Soong [25], The steps are as follows (i) design a 1-vector codebook, using centroid as prime for the entire set to train the vectors. As a result, no iteration is required in this step. (ii) The codebook is split into code words. (iii) Perform a nearest-neighbour search for each training vector, find code words in the current codebook, and assign the closest vector (in terms of the minimum distance measurement) to the corresponding cell (associated with the closest centroid). This step is done using the Euclidean distance measurement. (iv) The next step is to update the centroid in each cell, using the centroid of the training vectors assigned. For VQ, the centroid requires updating the codebook by taking the average of the speech or the heart vector in a cell, to find the new value of the code vector. (v) Then, in iteration 1, steps 3 and 4 are repeated until the average distance falls below a present threshold. (vi) Finally in iteration 2, steps 2, 3, and 4 are repeated until a codebook of size M is reached.

During the process of this development, the raw speech and the heart signals are pre-processed to extract the MFCC. Using the above algorithm, a codebook is developed for each client. Detailed results of SI, SV, HIS, and HSV systems will be discussed in the next section.

Evaluation of Client Recognition System Based On Speech and Heart Sounds

Speaker Identification (SI)

As mentioned earlier, the evaluation of SI is based on a set of 20 client speakers enrolled into the system. The 40 impostor speakers are used to evaluate the overall performance of the system. The EER results are based on single digits. The capability of SI to capture the underlying characteristics of the clients' data relies on the number of tokens and size of the codebook. If the codebook is not big enough, then there is limited freedom to form the decision cell. More training data would provide the extra flexibility. Since there is no theoretical guideline for this problem, several experiments are conducted to determine the appropriate codebook size for this task. Identification accuracy is a sensitive indicator of the ability of a parameter to discriminate speakers; for example: results are shown in table 1. The results show speaker

identifications of 12 MFCC features that are calculated for each frame of signal as a feature. A speaker identification experiment is conducted to evaluate the effectiveness of these features. It is found that the SI system provides an average overall performance of 99.3% and Standard Division (SD) of 1.48.

Speaker Verification (SV)

The effect of codebook size begins by assessing size 16, 32, 64, 128 and 256. All experiments show that no improvement is seen as the codebook size increased, and this is probably because of the limited size of training data used to develop the codebook. If more training tokens are used, a better EER performance would be achieved over the entire range of codebook sizes, before settling on a fixed point. Here, 20 client speakers are chosen as a set of client speakers, and are enrolled in the speaker verification (SV) system. The SV system is evaluated with single isolated digits. The performance of each client is listed in figure 3, with EER varying from 0.2% to 10.9%. This EER is for all digits. The overall average EER for the client is 2.1% and Standard Division of 1.49. While each digit displays different performance in isolation, each digit emphasizes a different aspect of time, varying from speech signals and rankings of digits that may vary from clients.

Heart Sound Identification (HSI)

In this experiment, a one minute heart sound is collected at the Center of Biomedical Engineering (CBE), using Welch Allyn equipment, with a sampling frequency 44100 Hz; then, the signal down sample to 16 kHz. Each of the one minutes heart sound recording is been cut into cycles, based on R to R segmentation from ECG for training and testing from each person. The results show that about 13 client subjects have more than 90% accuracy, as shown in table 2. The poor performance by certain clients [15,17-19], may be due to interference of noise and artefacts found in their training and testing samples. The identification system has an average performance of 81.9%, with SD of 21.79, and a performance difference of 17.5%, as compared to the SI model. The approach used for noise elimination was successfully carried out for patients with cardiac murmurs. In future work, we would like to apply

Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	accuracy (%)
Client 1	49	1																			98
Client 2		50																			100
Client 3			50																		100
Client 4				50																	100
Client 5					50																100
Client 6						50															100
Client 7							50														100
Client 8								47	3												94
Client 9									50												100
Client 10										50											100
Client 11											50										100
Client 12												50									100
Client 13													49	1							98
Client 14					1									49							99
Client 15															50						100
Client 16																50					100
Client 17																	50				100
Client 18					1													49			98
Client 19																			50		100
Client 20										1										49	98
AVERAGE																					99.3%, SD (1.48)

Table 1: Speaker identification overall results from the (SI) 20 client speakers.

Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	accuracy (%)
Client 1	30																				100
Client 2		30																			100
Client 3			30																		100
Client 4				30																	100
Client 5					30																100
Client 6						30															100
Client 7							30														100
Client 8								30													100
Client 9									29		1										96.7
Client 10										29	1										96.7
Client 11									1	1	28										93.4
Client 12								2				27		1	1						90
Client 13												1	27					1			90
Client 14													3	21				3	1	3	70
Client 15										4			4		21					1	70
Client 16														4	1	20			3	2	66.7
Client 17												6			9		15				50
Client 18							1							5	1	1		12			40
Client 19										3			2	2	5	7			11		36.7
Client 20	2												2		7	2		2	4	11	36.7
Average																					81.9%, SD (21.79)

Table 2: Heart Sound Speaker Identification (HSI) results.

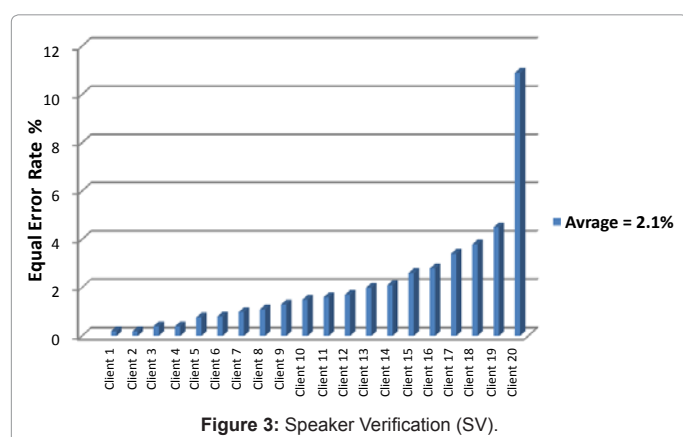


Figure 3: Speaker Verification (SV).

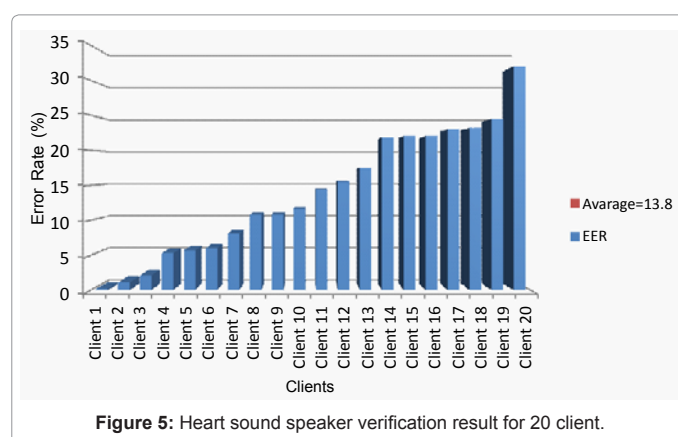


Figure 5: Heart sound speaker verification result for 20 client.

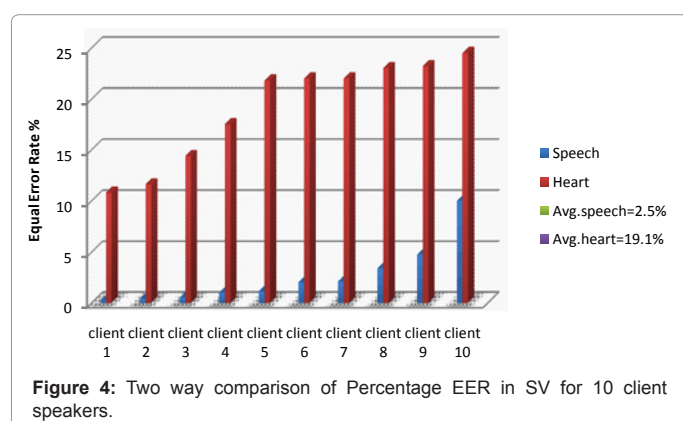


Figure 4: Two way comparison of Percentage EER in SV for 10 client speakers.

such technique for the current biometric system. In order to address the above issue, instead of using measurement samples, the wavelet coefficients are used. As a result, the dimension of the state vector is reduced so as the computational effort, as well. The noisy heart sound measurement Y_t will be transformed into a vector of wavelet coefficients, before estimation of θ_t , which is its de-noised version. The estimation

θ_t will be used to reconstruct the clean heart sound signal. This method has been used by Hadrina et al. [26] and Barschdorff et al. [27]. Discrete wavelet transform of biorthogonal type 5.5 is used with approximation coefficient level of 2.

Heart Sound Verification (HSV)

The performance of HSV model with 10 clients is compared with SV model, as shown in figure 4. The verification scores based on the output of this model are used to apply the EER threshold, to determine whether to accept or reject a client. As with the previous experiment, the threshold is speaker specific. The performance of the 20 clients is shown in figure 5, where there are considerable variations in the performance between the clients. The HSV model has an average EER of 13.8% and SD of 9.09, with a range of 0.2% to 32.2%. The SV model, on the other hand, shows an average EER of 2.1%, with a range of 0.2% to 10.9%. It is expected that the SV model would perform better than the HSV model. In addition, this is mainly due to the noise and the artifacts of the lub and dub sounds, which play a significant role that can degrade the performance of the HSV model. Variations of this lub and dub sound would depend on whether the person is fat or thin, young or old, sick or

healthy. The four auscultation heart points provide different distinctive sounds for each of the cardiac patients [28]. Determining these specific points could enhance the performance in future works.

Multimodal HSV/SV score fusion performance

A VQ output score is important and useful to the biometric task, because of the ability to represent the statistical properties of data. In this work, different biometrics of the same person are acquired and combined to complete and improve the recognition process. There are several levels [29-31], at which fusion can take place, such as Sensor level, Match score level, Rank level, Decision level and Feature level. In this paper, the match score level as shown in table 3 is applied on each individual biometrics process. The table summarizes the average of 10 client of the multimodal system, along with the Standard Division (shown in the parentheses) for different normalization and fusion scheme. The fusion process fuses the speech and heart sound Euclidean distance values into a single score, which is then compared to the system acceptance threshold. The next step is the score normalization stage. The output score from the SV and the HSV are of different numerical ranges. The Simple Sum (ΣSi) method is used to normalize scores of heart and speech biometrics. The normalized scores are obtained by using the following techniques: min-max, Z-score, median-MAD, Double sigmoid, Tanh and pricewise linear. The results from the analysis are shown in table 3, for example: fusion type Simple Sum combine with pricewise-linear, provide the best result with an average equal error rate of less than 0.7% with. The pricewise linear normalization and simple sum fusion gave less than 1% EER for 7 of its client subjects, followed by max-min, z-score and Tanh, which have 4 of the subjects with the same EER. The worst performance is from a combination of double segment and median-MAD in the double sigmoid, where 3 subjects did not show resilience to the errors in estimating the densities. The median-MAD, on the other hand, shows that all of the subjects had an EER of more than 3%.

The result of the speakers' acceptance or rejection stage using

SV system is shown in figure 3. The results are then compared with fusion type base on the Main Rule. Speaker Specific (SS) threshold are used to evaluate the SV system. The threshold is determined by the EER criteria. The use of EER in both forms provides a standard set of measurement that details the performance of the SV system. Figure 6 shows the performance improvement fusion type Simple Sum over the Main Rule, where large differences in the distribution of errors between these two fusions can be seen.

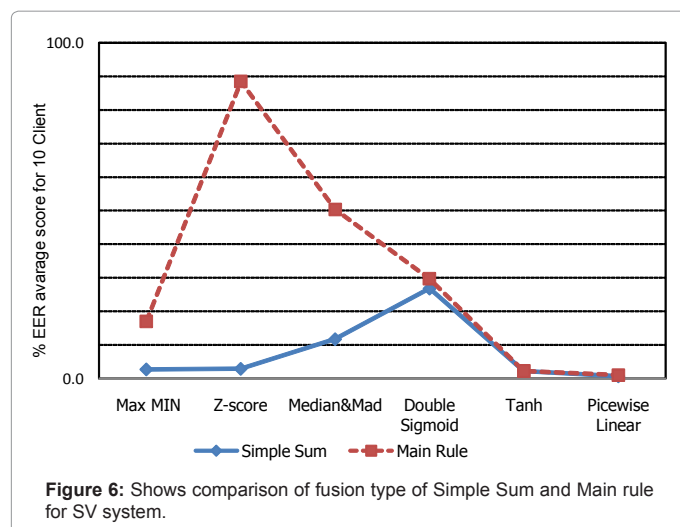
Client recognition with Electrocardiogram (ECG)

Combining the results of different biometric models has shown some significant improvement in the performance of the biometric system. The problems, such as noise and artefacts, as described before for the uni-modal HSV or HIS, significantly reduced the performance of this system. Adding speech as a multimodal system certainly increases the number of clients to be enrolled, but also provides better performance results. The works presented here further investigate the unique and private features of an individual based on ECG. The first part involves the fluttering of the signal and baseline wonder removal. The main reason of the baseline wonder and noise of the ECG is normally caused by changes in the electrode impedance, perspiration or body movement. These unwanted conditions produce artefacts that greatly influence the rich information found in the ECG signals. The baseline wondering can be removed by using high pass digital filter, without changing or disturbing the characteristics of the waveform. The detection of fiducial points increases the complexity of the ECG base identifiers. In fact, there are no definite rules or techniques for localizing wave boundaries, especially when ECG traces can cause anomalies.

In order to analyze the characteristics and wave shape of the ECG, signals are first segmented into windows of cardiac cycle. Each of these window represent cardiac cycle, consisting of P-wave, QRS complex and T-wave. In order to achieve this goal, the system first needs to find the highest peak or the R-peak. The R-peak detection algorithm is based on

Fusion type	% EER before Fusion			Normalization type (% EER after fusion)					
	Name	SV	HSV	Max MIN	Z-score	Median & Mad	Double Sigmoid	Tanh	Picewise Linear
Simple Sum	Client 1	4.77	11.6	5.8	5.2	19.0	1.3	5.2	1.5
	Client 2	3.4	21.8	4.0	3.8	19.3	2.6	3.9	2.9
	Client 3	10	22	8.5	8.0	18.1	1.2	8.0	1.4
	Client 4	0.51	17.5	1.5	1.1	19.5	92.7	1.1	0.6
	Client 5	0.4	23	0.7	0.5	4.3	0.0	0.5	0.0
	Client 6	0.2	22	0.2	0.1	3.3	83.6	0.1	0.0
	Client 7	2.1	14.4	1.5	1.4	4.1	0.3	1.5	0.5
	Client 8	1.1	23.2	1.0	0.8	5.6	0.0	0.8	0.0
	Client 10	0.99	24.5	0.6	0.5	6.3	84.0	0.5	0.0
	Client 9	2	10.8	3.7	7.6	17.8	2.5	1.0	0.0
	Avarage	2.5	19.1	2.7 ± 2.72	2.9 (3.03)	11.7(7.45)	29.8 (41.44)	2.3 (2.59)	0.7 (0.96)
Main Rule	Client 1	4.77	11.6	23.9	89.6	45.1	1.1	5.3	1.0
	Client 2	3.4	21.8	25.1	88.4	45.9	4.7	3.9	4.9
	Client 3	10	22	25.2	83.1	47.0	1.9	8.1	1.8
	Client 4	0.51	17.5	22.7	95.7	53.2	97.1	1.1	0.5
	Client 5	0.4	23	8.7	87.3	45.6	0.0	0.5	0.0
	Client 6	0.2	22	10.5	89.4	54.1	90.9	0.2	0.0
	Client 7	2.1	14.4	9.7	81.0	50.1	0.0	1.5	0.4
	Client 8	1.1	23.2	11.9	92.1	61.2	0.0	0.8	0.0
	Client 10	0.99	24.5	11.9	91.1	52.1	95.9	0.6	0.0
	Client 9	2	10.8	20.7	87.2	49.7	5.7	0.7	1.9
	Avarage	2.5	19.1	17.0 (7.00)	88.5 (4.25)	50.4 (4.99)	29.7 (44.85)	2.3 (2.64)	1.1 (1.53)

Table 3: Fusion verification performance of the HSV/SV for every 10 client.



selective threshold, which says that R peak is significantly larger than the surrounding data. Once the location of R peaks is known in the ECG signal, the proposed algorithm which is used to detect the Q and S points is summarized to a single cycle as follow:

- 1) Set a scanning range on the left and right sides of R peak.
- 2) Split the QRS complex into two different portions (left & Right), and flip the left side to be similar to the right side, in order to make the programming process faster for both sides.
- 3) Use a window of 1.3 ms (50 samples), starting from the scan offset along the curve, to create a polynomial to draw a line from offset point to the left side end point.
- 4) Find the temporary curve cutoff point, using the farthest distance between the line in step 4 and the curve.
- 5) To ensure the ideal cutoff point, repeat steps 4 and 5 replacing the polynomial to draw a line from the resulted temporary cutoff point to the curve end,
- 6) Compare both points; choose the minimum as final curve cutoff point.
- 7) Finally, P and T waves detection is proceed.

It is rather difficult to compare the results from other researchers work because the performance metric are often different from the Equal Error Rate (EER), the size and the type of data. To the best of our knowledge, some papers [7,8,32], reported for medical biometric system that uses biosignals, discussed their performance on identification rather than verification. However, recent review by Akram et al. [6], Agraftoti and Hatzinakos [8], Zayaraz et al. [33], Yang and Li [34], Al-Hamdani et al. [35], in medical biometrics covers work in Deoxyribo-Nucleic Acid (DNA) finger vain, retina vessel, and heart (ECG and PCG), and both of these areas. The uni-multimodal biometric system reported here is based on identification and verification results. For ECGI, the average identification accuracy of the 20 clients is 98.7%. The ECGV has an average EER of 4.2%, with a range of 0.3% to 12%. For identification as well as verification, the uni-biometric system based on speech, provides the overall best performance. In term of multimodal verification, the multimodal HSV has the best accuracy of 0.7%, followed by HECGV with 2.3% and SECGV with 2.7% accuracy. There is a 14.8% different in performance between the HECGV and SECGV multimodal system.

Conclusion

This work presents a reliable system in the multimodal biometrics verification scenario. It is found that fusion strategy is an efficient and reliable way of predicting and correcting erroneous classification decisions in multimodal (speech and heart) systems. It has been observed that the multimodal provides better performance, with the use of simple-sum score fusion and pricewise-linear normalization technique. There is an improvement of 96.3% as compared to the HSV model, and 60% compared to the SV model. Combination of score fusion between the ECG and speech (SECGV) model provides an improvement of 35.7%, when compared with the uni-model of ECGV, only a small improvement of 20.6% over the SV model. Fusion score between HECGV shows significant improvement of 48.1% over the HV model. The EER performance measure supports two basic conclusions:

- Input representation of the multi-modal biometric (HSV) is more appropriate (for the given ASV task) than ECGSV or HECGV based system.
- Experimental study with small scale client and impostors perform better in multimodal than the uni-model biometric system.

Future Work

The results obtained from the analysis of 20 clients and 70 impostors is encouraging, but it is best to acknowledge that these results are still not enough to draw a significant statistical conclusion. Further works will be carried out to maximize data collection with different time variations, to address the noise and artifacts. In addition, segmentation to specific areas like systolic, S_1 and S_2 sounds may contain visible information for an alternative input feature to the current biometric system.

Acknowledgment

We would like to thank the Centre of Biomedical Engineering, University Technology Malaysia (UTM) research grand (GUP-flagship Q.J130000.2436.00G31), and the Ministry of Higher Education (MOHE) research grant that has supported this research work.

References

1. Ayinde O, Yang YH (2002) Face recognition approach based on rank correlation of gabor-filtered images. Pattern Recognit 35: 1275-1289.
2. Han CC, Cheng HL, Lin CL, Fan KC (2003) Personal authentication using palm-print features. Pattern Recognit 36: 371-381.
3. Marupudi N, John E, Hudson F (2006) Fingerprint verification in multimodal biometrics. Region 5 Conference, IEEE 130-136.
4. Yang J (2010) Biometrics verification techniques combing with digital signature for multimodal biometrics payment system. In: Fourth International Conference on Management of e-Commerce and e-Government (ICMeCG), Nanchang, China.
5. Oglesby J, Mason JS (1990) Optimization Of Neural models for speaker identification. IEEE International Conference on Acoustics Speech and Signal Processing 1: 261-264.
6. Akram MU, Tariq A, Khan SA (2011) Retinal recognition: Personal identification using blood vessels. In: International Conference for Internet Technology and Secured Transactions (ICITST), Islamabad, Pakistan.
7. Phua K, Chen J, Dat TH, Shue L (2008) Heart sound as a biometric. Pattern Recognit 41: 906-919.
8. Agraftoti F, Hatzinakos D (2009) ECG biometric analysis in cardiac irregularity conditions. Signal, Image And Video Processing 3: 329-343.
9. Sadaoki F (1997) Recent advances in speaker recognition. Pattern Recognit Lett 18: 859-872.

10. Agrafioti F (2008) Robust subject recognition using the Electrocardiogram. University of Toronto, Toronto, Canada.
11. Su F, Xia L, Cai AN, Wu Y, Ma J (2010) EEG-based Personal Identification: from Proof-of-Concept to A Practical System. 20th International Conference on Pattern Recognition (ICPR) 3728-3731.
12. Shen WT, Tompkins WJ, Hu YH (2002) One Lead ECG for identity verification. Engineering in Medicine and Biology, 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference 1: 62-63.
13. Israel SA, Irvine JM, Cheng A, Wiederhold MD, Wiederhold BK (2005) ECG to identify individuals. *Pattern Recognit* 38: 133-142.
14. Yu K, Mason JSD, Oglesby J (1995) Speaker recognition using hidden Markov models, dynamic time warping and vector quantisation. *IEE Proceedings: Vision, Image and Signal Processing* 142: 313-318.
15. Rosenberg AE (1976) Automatic speaker verification: A review. *Proceedings of the IEEE* 64: 475-487.
16. Rosenberg AE, Lee CH, Gokcen S (1991) Connected Word talker verification using whole word hidden Markov models. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* 381-384.
17. Bennani Y, Soulie FF, Gallinari P (1990) Connectionist approach for automatic speaker recognition. *International conference on Acoustics, Speech and Signal Processing (ICASSP)*, France 1: 265-268.
18. Ting CM, Salleh SH, Tan TS, Ariff AK (2007) Text independent Speaker Identification using Gaussian mixture model. *International Conference on Intelligent and Advanced Systems (ICIAS)*, Malaysia 194-198.
19. Ting CM, Salleh SH, Ariff AK (2009) Malay continuous speech recognition using fast HMM match algorithm. In: 4th IEEE Conference on Industrial Electronics and Applications (ICIEA), Malaysia 2424-2427.
20. Wübbeler G, Stavridis M, Kreiseler D, Bousseljot RD, Elster C (2007) Verification of humans using the electrocardiogram. *Pattern Recognit Lett* 28: 1172-1175.
21. Boles WW, Boashash B (1998) A human identification technique using images of the iris and wavelet transform. *IEEE Trans Signal Process* 46: 1185-1188.
22. Ariff AK, Alwi M, Salleh SH (2005) Malay speaker recognition system based on discrete HMM. 1st International Conference on Computers, Communications, & Signal Processing with Special Track on Biomedical Engineering (CCSP), Malaysia 292-295.
23. Soong FKK, Rosenberg AE, Rabiner LR, Juang B-hH (1985) A vector quantization approach to speaker recognition. *IEEE International Conference on ICASSP* 387-390.
24. Molau S, Pitz M, Schluter R, Ney HJ (2001) Computing mel-frequency cepstral coefficients on the power spectrum. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* 1: 73-76.
25. Rosenberg AE, Soong FK (1987) Evaluation of a vector quantization talker recognition system in text independent and text dependent modes. *Comput Speech Lang* 2: 143-157.
26. Hadrina SH, Salleh SH, Ariff AK, Alhamdani O, Tian-Swee T, et al. (2012) Application of Multipoint Auscultation For Heart Sound Diagnostic System (MAHDS). 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA) 836-841.
27. Barschdorff D, Ester S, Dorsel T, Most E (1990) Neural network based multi sensor heart sound analysis. *Proceedings: Computers in Cardiology* 303-306.
28. Wu H, Kim S, Bae K (2010) Hidden Markov Model with heart sound signals for identification of heart diseases. *Proceedings of 20th International Congress on Acoustics (ICA)*, Sydney, Australia.
29. Jain A, Nandakumar K, Ross A (2005) Score normalization in multimodal biometric systems. *Pattern Recognit* 38: 2270-2285.
30. Jing XY, Yao YF, Zhang D, Yang JY, Li M (2007) Face and palmprint pixel level fusion and Kernel DCV-RBF classifier for small sample biometrics recognition. *Pattern Recognit* 40: 3209-3224.
31. Ross A, Jain AK (2004) Multimodal biometrics: An overview. *Proceedings of the 12th European Signal Processing Conference* 1221-1224.
32. Revett K, Deravi F, Sirlantzis K (2010) Biosignals for user authentication-towards cognitive biometrics. In: *International Conference on Emerging Security Technologies*, UK.
33. Zayaraz G, Vijayalakshmi V, Jagadiswary D (2009) Securing biometric authentication using DNA sequence and Naccache Stern Knapsack cryptosystem. *International Conference on Control, Automation, Communication and Energy Conservation (INCACEC)* 1-4.
34. Yang JF, Li X (2010) Efficient finger vein localization and recognition. 20th International Conference on Pattern Recognition (ICPR) 1148-1151.
35. Al-Hamdani O, Chekima A, Dargham JA, Saleh SH, Noor AM, et al. (2012) Multi-biometrics fusion (heart sound-speech authentication system). In: *International Symposia on Imaging and Signal Processing in Health Care and Technology (ISPHT)*, Baltimore, USA.