Evaluation on Score Reliability for
Biometric Speaker Authentication Systems

Muhammad Hazim Hasan, Haryati Jaafar and Dzati Athiar Ramli
Intelligent Biometric Research Group (IBG),
School of Electrical and Electronic Engineering, USM Engineering Campus,
Universiti Sains Malaysia, 14300, Nibong Tebal, Pulau Pinang, Malaysia

---

Abstract: Problem statement: Fusion weight tuning based on score reliability is imperative in order to ensure the performances of multibiometric systems are sustained. **Approach:** In this study, two variant of conditions i.e., different performances of individual subsystems and inconsistent quality of test samples are experimented to multibiometric systems. By applying multialgorithm scheme, two types of features extraction method i.e., Linear Predictive Coding (LPC) and Mel Frequency Cepstrum Coefficient (MFCC) are executed in this study. Support Vector Machine (SVM) is used as a classifier for both subsystems for the pattern matching process. Scores from both LPC and MFCC based sub systems are fused at score level fusion using fixed weighting and adaptive weighting approaches. For fixed weighting, sum-rule method is employed while for the adaptive weighting, sum-rule based on weight adaptation and sum-rule with weight produced from fuzzy logic inference are executed. The performances of single, fixed and adaptive systems are then compared. **Results:** Experimental results show that at 40dB and 20dB SNR signals, EER performances of single systems are 1.1730 and 38.2695% respectively. Consequently, the EER performances are observed as 2.7355 and 1.1359% for the sum-rule based on weight adaptation and sum-rule with weight produced from Fuzzy Logic. **Conclusion:** The results show that fusion system based on fuzzy logic gives advantage due to its capability in adjusting the weight based on the subsystem performance and quality of the current data.

Keywords: Linear Predictive Coding (LPC), Multialgorithm, sum-rule, fuzzy logic, Mel Frequency Cepstrum Coefficient (MFCC)

INTRODUCTION

Speaker recognition is a biometric system that uses individual's voice for recognition purpose and has become one of the premier applications for machine learning and pattern recognition technology. The speaker recognition process relies on feature influenced by both physical structure of an individual's vocal tract and behavioral characteristics of the speech. The biometric speaker recognition system has co-evolved with the technology of speech recognition because of the similar characteristics and challenges associated with each other. Hence, this system uses specific information contained in speech signal for authentication and identification purposes. In authentication system, the systems verify either to accept or reject the claimed identity by approving the genuine otherwise rejecting the imposter while, for identification systems, the task is to determine the unknown user for authorizing intention.

The advantages of using speech signal trait for biometric systems are that the signal is natural and easy to produce, requiring little custom hardware, has low computation requirement and is highly accurate (in clean noise-free conditions) (Ramli *et al*., 2008). However, sometimes a single biometric system fails to authenticate the identity of a person due to insufficient information or by spoofing. For instance, the major setback utilizing speech signals for biometric systems is due to the severe degraded performance as the Signal to Noise Ratio (SNR) of the speech signal drops in noisy conditions. In addition, since voice is categorized as a behavioral signal, the information has a tendency to be different due to the change of speaking rate and environment, for instances, sickness (e.g., head cold can alter the vocal tract), extreme emotional state (e.g., stress or duress), long interval between enrolment and verification process, poor or inconsistent room acoustic and aging (Campbell *et al*., 2003; Samad *et al*., 2007).

One of the solutions to overcome these limitations is by comparing different existing algorithms on the

---

Corresponding Author: Muhammad Hazim Hasan, Intelligent Biometric Research Group (IBG), School of Electrical and Electronic Engineering, USM Engineering Campus, Universiti Sains Malaysia 14300, Nibong Tebal, Pulau Pinang, Malaysia

specific problem and selecting the best of the algorithms that is able to be applied. However, selecting the best algorithm is not an easy task. Hence, combining multiple algorithms that employ multiple feature extraction and/or multiple matching algorithms on the same biometrics is executed as an alternative approach due to the supplementary information from multi algorithm also helps to improve the performance. Moreover, utilization of new sensor is not required thus it is cost effective. Many researchers have proved that the implementation of the fusion approach can help to improve the performance of biometric system (Ramli *et al*., 2009). It is also imperative to assign different weighting in fusion to each biometric trait in order to vary the contribution of matching scores of each biometric trait since the optimum weight can maximize the performance of multibiometric system.

This study evaluates the score reliability of multialgorithm approaches by fusing the data at match score level. The database consists of 2220 audio data which obtained from 37 speakers from three recording sessions. The experiment is conducted based on clean, 40, 35, 30, 25, 20, 15, 10, 5, 0 and -5dB Signal to Noise Ratio (SNR) of audio signal. Two features based on Mel Frequency Cepstrum Coefficient (MFCC) and Linear Predictive Coding (LPC) is executed in this study. The Support Vector Machine (SVM) is used as a classifier for both subsystems in the pattern matching process. The objective of this research is given as follows. First, to develop single biometric systems based on two different feature extraction algorithms which are MFCC and LPC. Subsequently, both of the MFCC and LPC features will be combined at the score level fusion namely as multi algorithm speaker authentication system. The second objective is to evaluate the performances of this fusion system based on fixed and adaptive weighting schemes. For fixed weighting, sum-rule method is employed while for adaptive weighting, sum-rule based on weight adaptation and sum-rule with weight produced by fuzzy logic inference system are executed. The third objective of this study is to compare the performances of the single system, fixed and adaptive weighting fusion systems.

MATERIALS AND METHODS

Data acquisition processing: An audio feature is extracted by taking the information of the speech recording based on the speaker's tone and inflection analysis. In this study, the audio is obtained from the Audio-Visual Digit Database (Sanderson and Paliwal, 2003). The digital audio is monophonic, 16 bit 32 kHz and WAV format. The database consists of 2220 audio data which obtained from 37 speakers from three recording sessions has been simulated with Additive

White Gaussian Noise (AGWN). Each of the data undergoes a series of speech processing step that is pre-emphasis, framing and windowing as shown in Fig. 1 (Kisku *et al*., 2010; Daugman, 2000).

The pre-emphasis process is the process to compress the signal dynamic range by passing it through a filter to emphasize the signal to higher frequencies in order to raise the SNR. In this process, the speech signal is filtered with a first order FIR filter whose transfer functions in the z-domain as given in Eq. 1:

$$H(z) = 1 - a.z^{-1} \ 0 \le a \le 1 \tag{1}$$

where, as is the preemphasis parameter (Furui, 2001, Kisku *et al*., 2010). In the time domain, the relationship between output $x'(x)$ and the input $x'(n)$ of the preemphasized signal is given as in Eq. 2:

$$x'(x) = x(n) - ax(n-1) \tag{2}$$

In this study, the value of a is considered as 0.95 where this value can increase the SNR to more than 20dB amplification of the high frequency spectrum (Becchetti and Ricotti, 1999).

The process of digitization is applied to convert the speech samples from Analog to Digital Conversion (ADC). For, speech signal, spectral evaluation can be performed using short time analysis by windowing the preemphasized signal x'(x) into a string of windowed sequence, $x_t(n)$, t = 1,2,...,T, called frames which are processed individually as in Eq. 3 and 4:

$$x'(x) \equiv x'(n - t.M), \quad 0 \le n \le N, \ 1 \le t \le T \tag{3}$$

$$x_t(n) \equiv w(n).x_t'(n) \tag{4}$$

where, w (n) is the impulse response of window.

In this process, the audio signal is divided into frames of N samples where N is the length of each frame. Each frame is shifted by a temporal length M with M<N, makes N-M samples at the end of frame x_t '(n) are duplicated at the beginning of the following frame x_{t+1} '(n). A suitable value for length N is important according to Kondoz (1969). If N is very large, the short time energy will be averaged over a long time hence will not reflect the changing properties of the speech signal. However, if N is small, the short time will change rapidly. 20 ms duration of length N with 50% overlapping is an ideal measurement. Windowing process is then applied to minimize the signal discontinuities at the beginning and end of each frame by zeroing out the signal outside the region of interest.
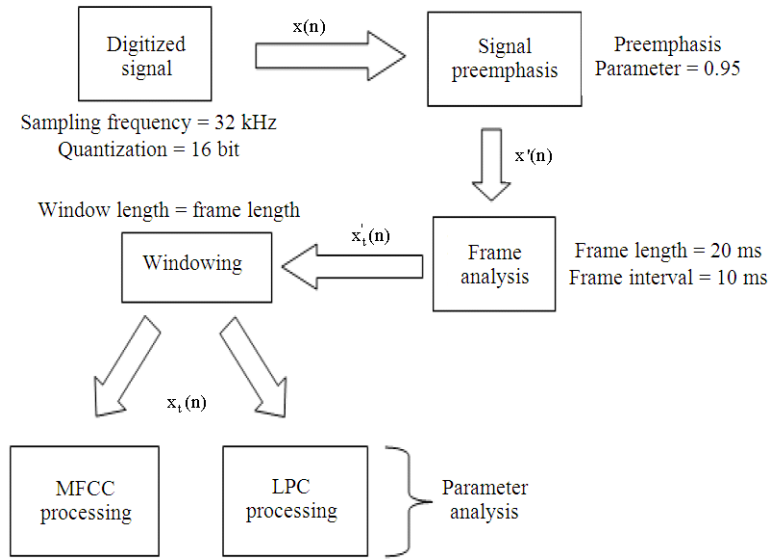
Fig. 1: Speech processing

The Fourier Transform $X_t (e^{j\omega})$ for the discrete time signal $x_t (n)$ can be written as Eq. 5:

$$x_t (e^{j\omega}) = \sum_{n=0}^{N-1} x_t (n) e^{-j\omega n} \tag{5}$$

W (n) and x_t '(n) which are product in the time domain as given in equation 4 becomes a convolution in the frequency domain after introducing the Fourier transform. Considering equation 3-5, the equation of Short Time Fourier Transform (STFT) of x '(n) is initiated as Eq. 6:

$$x_t (e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x'(n - t.M).w(n) e^{-j\omega n} \tag{6}$$

In order to increases the resolution and no side lobes or frequency leakage; the ideal window function should be a narrow main lobe. In this study, Hamming window w (n) H is used as the window function due to the side lobes of this window are lower compared to other windows. Moreover, a high resolution is not required in speaker recognition since it reduces resolution. Hamming window, $w_H(n)$ is defined as in Eq. 7:

$$w_H (n) = 0.54 - 0.46 \cos\left(\frac{2n\pi}{N-1}\right) n = 0,...,N-1 \tag{7}$$

Feature extraction: In this study, two features which are MFCC and LPC have been used for the development of the multibiometric systems. MFCC feature is based on the known variation of the human ear's critical bandwidths which frequency expressed in the mel-frequency (Chen and Luo, 2009). The mel-frequency is linear spaces below 1000Hz and logarithmic spaces above 1000Hz. The operation of this system is based on two types of filter which are linearly and logarithmically spaced and processes on the Fourier transform of x_t (n): $X_t (e^{j\omega})$. The $X_t (e^{j\omega})$ is evaluated only for discrete number of ω values.

There have several steps in MFCC processing. The first step is computation of the Discrete Fourier Transform (DFT) of all frames of the signal. By considering $\omega = \frac{2\pi k}{N}$, the DFT of all frames of the signal, $x_t(k)$ is obtained as in Eq. 8:

$$x_t (k) : X_t (e^{j2\pi k/N}), k = 0,...,N-1 \tag{8}$$

The computational complexity can also be reduced if the number of samples N is a power of 2. The result obtained after this step is called as signal's spectrum.

A filter bank processing is the second step in MFCC processing. Filter banks properly integrate a spectrum at defined frequency and spectral features are obtained after this process. The outputs of the filter bank are denoted as Y_t (m), $1 \le m \le M$ where M is number of band-pass filters. In general, a set of 24 band-pass filter is used since it simulates human ear processing. Subsequently, computation of the log energy is the third step which computes the logarithm of the square magnitude of the filter banks outputs, y_t (m).
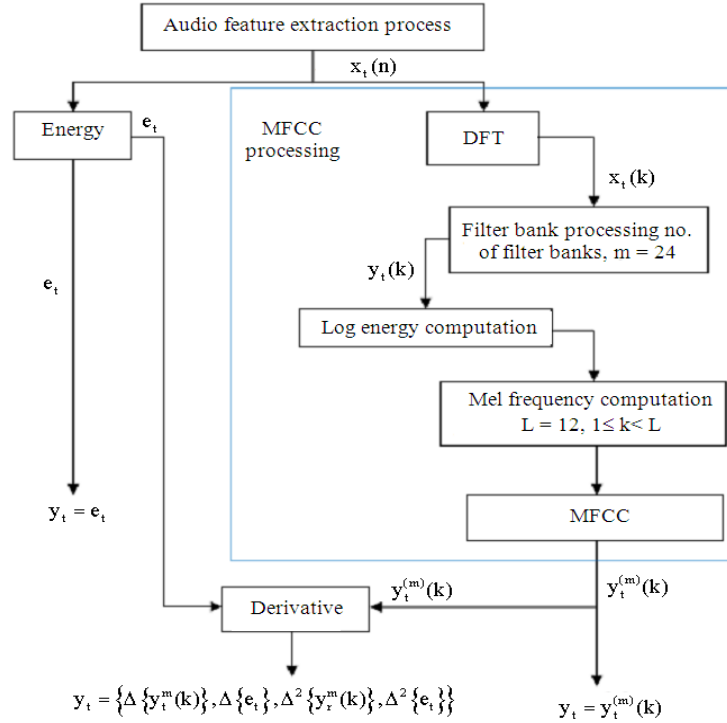
Fig. 2: MFCC block diagram

The final step for MFCC processing is mel frequency cepstrum computation that performs the inverse DFT on the logarithm of the magnitude of the filter bank output given in Eq. 9:

$$y_t^m(k) = \sum_{m=1}^{M} \log\{Y_t(m)\} \cdot \cos\left(k\left(m - \frac{1}{2}\right)\frac{\pi}{M}\right) \quad k = 0,...,L \tag{9}$$

In this study, the database of MFCC features consists of 2220 set of MFCC features from 37 persons with 60 speech signal data per person. There are 12 mel cepstrum coefficients, one log energy coefficient and three delta coefficients per frame. The overall process of the MFCC is shown in Fig. 2.

LPC feature extraction models the process of speech production and is defined as a digital method for encoding an analogue signal in which a particular value is predicted by a linear function of the past values of the signal (Rabiner and Juang, 1993; Furui, 1981). The most important aspect of LPC is the linear predictive filter which allows the value of the next sample to be determined by a linear combination of previous samples. In other word, linear prediction filters attempt to predict future values of the input signal based on past signals. LPC analysis is based on the assumption that the relation between the current sample x(n) and first-order linear combination of the previous p samples given as in Eq. 10:

$$x(n) \approx a_1 x(n-1) + ... + \alpha_p x(n-p) \tag{10}$$

The linear predicted value $\tilde{x}(n)$ with prediction coefficients, σ_i for x(n) is presented as in Eq. 11:

$$\tilde{x}(n) = \sum_{i=1}^{p} a_i x(n-i) \tag{11}$$

Consequently, LPC cepstrum can be derived through the LPC model. For a time sequence x(n), complex cepstrums $\tilde{c}(n)$ are represented as Eq. 12-14:

$$\tilde{c}(n) = -\alpha_1 \tag{12}$$

$$\tilde{c}(n) = -\alpha_n - \sum_{m=1}^{n-1}\left(1 - \frac{m}{n}\right)\alpha_m \tilde{c}_{n-m} \quad 1 < n \le p \tag{13}$$

$$\tilde{c}(n) = -\alpha_n - \sum_{m=1}^{p}\left(1 - \frac{m}{n}\right)\alpha_m \tilde{c}_{n-m} \quad n < p \tag{14}$$

The database of LPC features in this study consists of 2220 set LPC features from 37 persons with 60 speech signal data per person. 14 cepstrum coefficients per frame are extracted in this method. The overall process of the LPC is shown in Fig. 3.
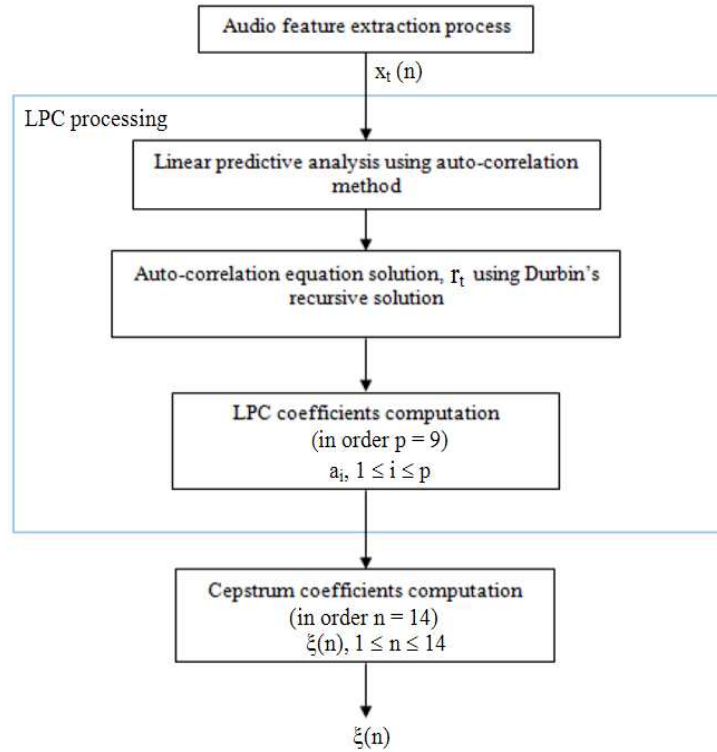
Fig. 3: LPC block diagram

Classification using SVM: SVM is a classifier which can classify sample within two or more classes. In the simplest form, linear and separable case, it is the optimal hyper plane that maximizes the distance of the separating hyper plane from the closest training data point called the support vectors (Gunn, 2005; Wan and Campbell, 2000). The solution of linearly separable case is started by considering a problem of separating the set of training vectors belongs to two separate classes as given in Eq. 15:

$$D = \left\{ \left(x^1, y^1 \right), ..., \left(x^L, y^L \right) \right\} \; x \in \Re^n \; y \in \{-1, 1\} \quad (15)$$

With a hyperplane as in Eq. 16:

$$(w, x) + b = 0 \quad (16)$$

where, w and b are the direction and position in space, respectively and w is normal to the plane. The hyperplane has the same distance from the nearest points from each class and the margin is twice the distance for each direction, w. The support vectors which is a linear combination of a small subset of data, x_s, $s \in \{1, ..., N\}$ is the solution for the optimal hyperplane. Eq. 17 is minimized by the hyperplane that optimally separates the

data which is equivalent to minimizing an upper bound on VC dimension:

$$\Phi(w) = -\frac{1}{2} \|w\|^2 \quad (17)$$

VC dimension is a scalar value that measures the capacity of the learning function. The saddle point of the Lagrange functional (Lagrangian) is used to solve the optimization problem and given as in Eq. 18:

$$\Phi(w, b, a) = -\frac{1}{2} \|w\|^2 - \sum_{i=1}^{L} a_i (y^i \left[\left\langle w, x^i \right\rangle + b \right] - 1) \quad (18)$$

where, a_i is the Lagrange multiplier. The Lagrangian has to be maximized with respect to $a \geq 0$ and minimized with respect to w and b. The solution of the linearly separable case is given by Eq. 19:

$$a^* = \arg \min_{\alpha} \frac{1}{2} \sum_{i=1}^{L} \sum_{j=1}^{L} \alpha_i \alpha_j y_i y_j \left\langle x_i, x_j \right\rangle - \sum_{k=1}^{L} \alpha_k \quad (19)$$

With constraints in Eq. 20:

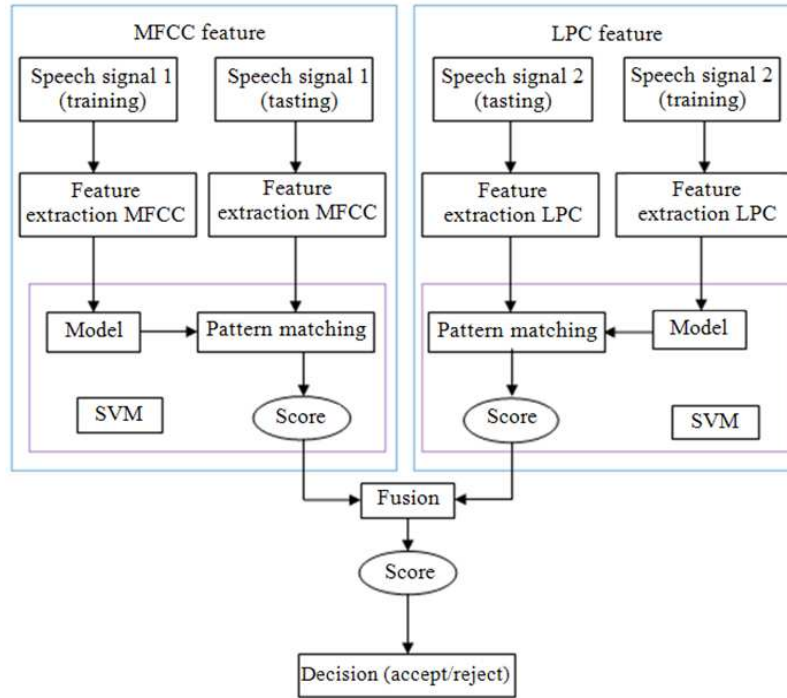$$0 \leq a_i \leq c \; i = 1, ..., L \; \text{and} \sum_{j=1}^{L} a_j y_j = 0 \quad (20)$$

Fig. 4: Fusion system

The nonlinear mapping is used in the case of the linear boundary is inappropriate which the SVM can map the input vector, x into a manifold embedded in a high dimensional feature space z. The SVM construct an optimal separating hyperplane in the higher dimensional space (Chen and Luo, 2009). The non-linear mappings are polynomial functions, radial basis function and certain sigmoid functions. In this study, polynomial kernel is employed. Hence, the optimization problem becomes as in Eq. 21:

$$a^* = \arg \min_\alpha \frac{1}{2} \sum_{i=1}^{L} \sum_{j=1}^{L} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{k=1}^{L} \alpha_k \quad (21)$$

With constraints as in Eq. 22:

$$0 \le a_i \le c \; i = 1,...,L \text{ and} \sum_{j=1}^{L} a_j y_j = 0 \; x(t) = s(t) - 0.95)x(x-1) \quad (22)$$

where, (K x_i, x_i) is the kernel function that performs the nonlinear mapping into feature space. For the polynomial kernel, it is defined as Eq. 23:

$$K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j) = (\gamma x_i^T x_j - s)^a \quad (23)$$

where, $\gamma > 0$ and γ, r and d are kernel parameters.

Fusion System: In this study, both of MFCC and LPC subsystems are combined together as a fusion system as shown in Fig. 4. By taking the benefit of score level fusion as discussed before, the scores from MFCC and LPC subsystem are then fused and the decision is made.

Two types of fusion schemes i.e., fixed weighting and adaptive weighting are implemented and compared at different level of SNR. In fixed weighting approach, the fusion algorithm which is sum-rule fusion scheme is applied while the optimum weight for the weight adaptation fusion system is then computed. The sum-rule fusion method is shown in Eq. 24:

$$\text{fusion} = (w_1 \times \text{LPC}_{\text{score}}) + (w_2 \text{MFCC}_{\text{score}}) \quad (24)$$

where, w is a fusion weight. W is varied from 0 to 1 in steps of 0.1. This study involves the fusion based on clean data. For this purpose, each speaker model is trained using 20 client training data and 720 (20×36) imposter training data. During testing, speaker model from each speaker is tested on 40 client data and 1440 (40×36) imposter data from 36 persons using clean signal. 1480 scores for each type of testing data are obtained.

In the adaptive weighting, the sum-rule with weight adaptation and sum-rule with weight produced from fuzzy logic inference system are applied (Vasuhi

et al., 2010). For sum-rule with weight adaptation, the optimum weight is adapted from the value of optimum weight in fixed weighting system and the audio systems are evaluated based on different SNR levels. Each speaker model is trained using 20 client training data and 720 imposter training data while 40 client data and 1440 imposter data are used as the testing data. The clean data are corrupted into 10 levels of SNR i.e., 40, 35, 30, 25, 20, 15, 10, 5, 0 and -5dB. During testing, speaker model from each speaker is tested on 40 client data and 1440 (40×36) imposter data from the other 36 persons for each level of the corrupted signals. 1480 scores for each type of testing data are obtained.

In the sum-rule with weight produced from fuzzy logic inference system, the range of SNR levels is divided into three levels; high, medium and low level. Hence, the important part of the fuzzy logic is to determine the optimum weight of the fusion systems according to SNR levels and both subsystem performances. For MFCC feature, the range between 25-40 dB is determined as high SNR level, 5-30dB is medium level while the low level is between 5-10dB. For LPC feature, the range of high SNR level is between 34 to 40dB, medium level is between 19-36dB and low level is between-5-21dB.

RESULTS

Performance of single biometric system: System performances based on Equal Error Rate (EER) for MFCC-SVM systems at different levels of SNR are shown in Table 1. A performance based on receiver operation characteristic is presented in Fig. 5.

Table 2 shows the EER performance for LPC-SVM systems based on different levels of SNR. The results based on ROC curve is presented in Fig. 6.
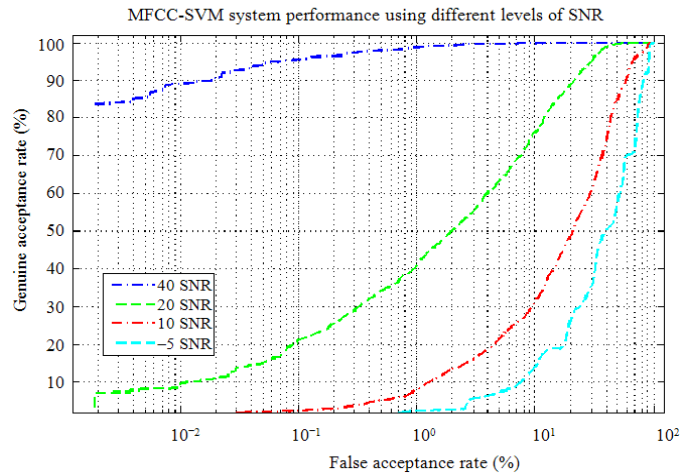

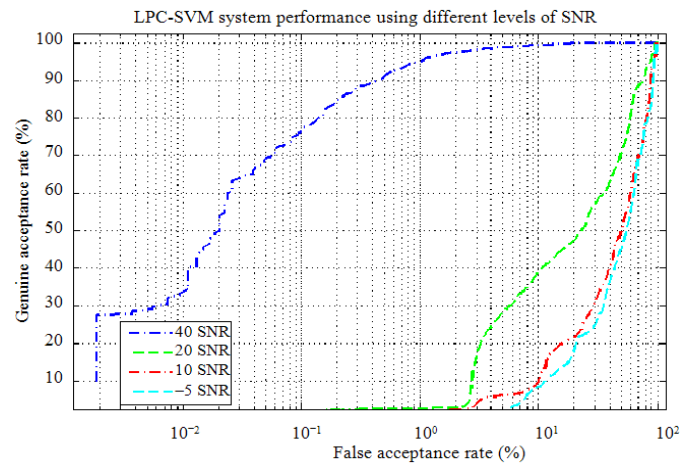Fig. 5: ROC curves for MFCC-SVM for different levels of SNR

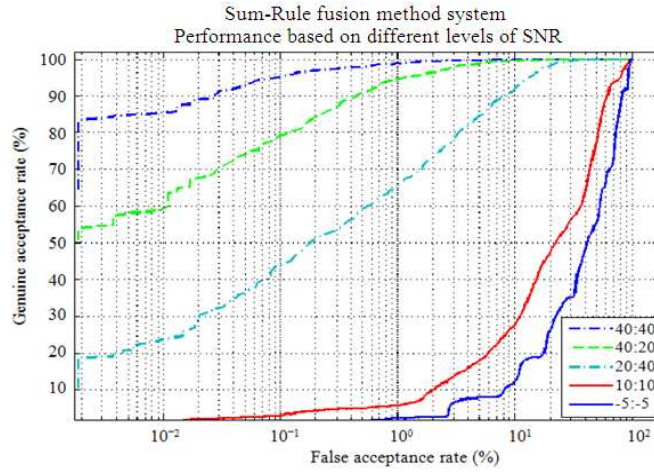Fig. 6: ROC curves for LPC-SVM for different levels of SNR

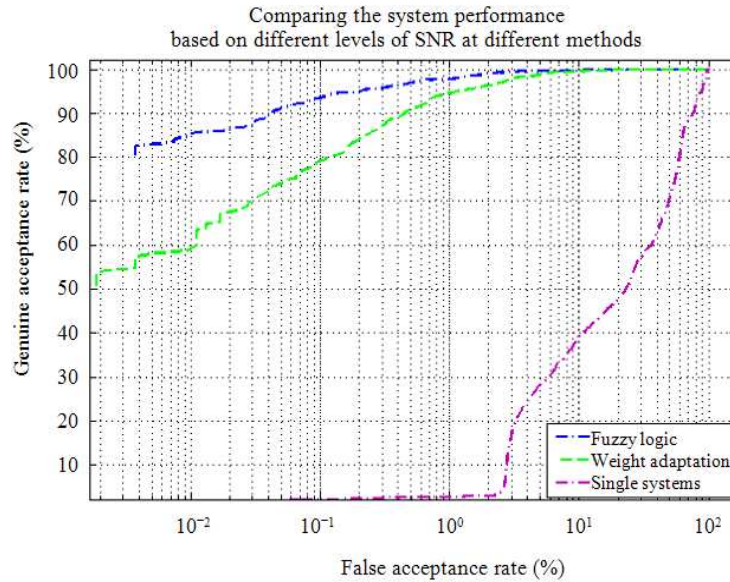Fig. 7: ROC curves for sum-rule fusion method based on different levels of SNR

Fig. 8: ROC curve for speaker authentication systems

Table 1: Performances for MFCC-SVM systems at different level of SNR

SNR levels	40 dB	20 dB	10 dB	-5 dB
EER	1.1730	15.2787	33.9367	46.8553

Table 2: Performances for LPC-SVM systems at different level of SNR

SNR levels	40 dB	20 dB	10 dB	-5 dB
EER	2.3761	38.2695	49.8048	52.4352

Performance of fixed weighting systems: The score ratios between MFCC-SVM subsystem and LPC-SVM subsystems and their corresponding EER performances using sum-rule scheme is shown in Table 3.

Table 3: Score ratio between MFCC-SVM and LPC-SVM features and the corresponding EER performances using SUM-RULE fusion method

Sum rule	w_1	w_2	EER
MFCC only	0.0	1.0	1.1524
MFCC and LPC	0.1	0.9	1.0867
MFCC and LPC	0.2	0.8	1.0313
MFCC and LPC	0.3	0.7	1.0004
MFCC and LPC	0.4	0.6	1.0820
MFCC and LPC	0.5	0.5	1.1909
MFCC and LPC	0.6	0.4	1.2416
MFCC and LPC	0.7	0.3	1.3091
MFCC and LPC	0.8	0.2	1.3476
MFCC and LPC	0.9	0.1	1.5334
LPC only	1.0	0.0	1.6582

Table 4: Combination of different levels of SNR between MFCC and LPC subsystems and the corresponding EER performances using sum-rule fusion method

(MFCC) Level of SNR	Levels of SNR (LPC)			
	40	20	10	-5
40	1.1374	2.7355	3.293	2.6699
20	9.0259	18.6327	21.3082	19.3581
10	24.8658	35.3238	38.247	37.3114
-5	38.4047	46.0548	47.1528	46.5869

Table 5: Comparison performances of the single system and weight adaptation

SNR MFCC (dB)	SNR LPC (dB)	MFCC (EER)	LPC (EER)	Sum-rule with weight adaptation (EER)	Fuzzy logic inference system (EER)
40	40	1.1730	2.3761	1.1374	1.1221
40	20	1.1730	38.2695	2.7355	1.1359
40	10	1.1730	49.8048	3.2930	1.1542
40	-5	1.1730	52.4352	2.6699	1.1633
20	40	15.2787	2.3710	9.0259	2.0940
20	20	15.2787	38.2695	18.6327	10.8112
20	10	15.2787	49.8048	19.3581	12.0199
20	-5	15.2787	52.4352	21.3082	15.7924
10	40	33.9367	2.3710	24.8658	2.2640
10	20	33.9367	38.2695	35.3238	30.5083
10	10	33.9367	49.8048	38.2470	32.8671
10	-5	33.9367	52.4352	37.3114	33.5567
-5	40	46.8533	2.3710	38.4047	2.3579
-5	20	46.8533	38.2695	46.0548	32.0282
-5	10	46.8533	49.8048	47.1528	40.1134
-5	-5	46.8533	52.4352	46.5869	46.5550

Performance of adaptive weighting system compared to other systems: Performance of adaptive weighting system is compared to the fixed weighting system and single system (LPC) at 20dB SNR using the ROC curve as shown in Fig. 8. The overall performances are also illustrated in Table 5.

DISCUSSION

Table 1 and Fig. 5 show the performances of the LPC-SVM systems based on 40dB, 20dB, 10dB and-5dB. At 40dB SNR, the Genuine Acceptance Rate (GAR) is almost 100% at False Acceptance Rate (FAR) of 6%. At the same FAR, the GAR performances for 20, 10 and -5dB are 70, 23 and 7%, respectively. The SNR at 40dB gives the lowest value of EER which signify the highest performance.

Table 2 and Fig. 6 shows the performances of the LPC-SVM systems based on 40dB, 20dB, 10dB and-5dB. For 40dB SNR, the GAR performance is almost 100% at FAR of 12%. At the same FAR, the GAR performances for 20, 10 and -5dB are 59, 30 and 24%, respectively.

Based on the fixed weighting system experimental results as given in Table 3, the score ratio between MFCC and LPC subsystems for sum rule fusion

method is fixed to 0.7: 0.3 for this adaptive system. Table 4 and Fig. 7 show the performances of sum-rule fusion method based on different levels of SNR.

For the adaptive weighting system, the fuzzy logic is applied as second approach in order to determine the optimum weight. For comparison, EER performance of the both single systems i.e., MFCC and LPC and fusion systems for adaptive weighting approach i.e., sum-rule with weight adaption and fuzzy logic have been computed.

Table 5 summarizes the performances of single systems (LPC and MFCC subsystems) and fusion systems (sum-rule with weight adaptation and Fuzzy logic inference system). Fusion system using sum-rule with weight adaptation performs well only when both subsystems are in clean conditions or when the LPC subsystem is in high SNR compared to the MFCC subsystems. Otherwise, the performances are worse than the single systems. This trend occurs because of the weight for this fusion scheme is adjusted with ratio MFCC to LPC of 0.7:0.3 which is based on the performance individual subsystems only and not according to the quality of testing data.

Consequently, the advantage of implementing multibiometric systems compared to single systems can be observed through the implementation of Fuzzy based fusion system due to an effective weight tuning, considering the deviation of both subsystems and the data quality.

CONCLUSION

This study reveals that the importance of fusion weight tuning in order to maintain the effectiveness of executing multibiometric systems over single biometric systems. By considering two types of variants i.e., different performances of individual subsystems and inconsistent quality of test samples, the fusion weight tuning is applied so as to ensure the fusion systems is at its best performances. Future research should focus on the other sources that influence the reliability of biometric scores and towards the proper approach in handling weight tuning for multibiometric systems.

REFERENCES

Becchetti, C. and L.R. Ricotti, 1999. Speech Recognition: Theory and C++ Implementation. 1st Edn., Wiley, England, ISBN: 0471977306, pp: 407.

Campbell, J.P., D.A. Reynolds and R.B. Dunn, 2003. Fusing high and low level features for speaker recognition. Proceeding of the Euro speech in Geneva, ISCA, Switzerland, pp: 2665-2668.

Chen, S.H and Y.R. Luo, 2009. Speaker verification using MFCC and support vector machine. Proceedings of the International MultiConference of Engineers and Computer Scientists, Mar. 18-20, Hong Kong, pp: 1-4.

Daugman, J., 2000. Combining multiple biometrics. Cambridge University.

Furui, S., 1981. Cepstral analysis technique for automatic speaker verification. IEEE Trans. Acoustics Speech Sig. Proc., 29: 254-272. DOI: 10.1109/TASSP.1981.1163530

Furui, S., 2001. Digital Speech Processing, Synthesis and Recognition. 2nd Edn., Dekker, New York, ISBN-10: 0824704525, pp: 476.

Gunn, S.R., 2005. Support vector machines for classification and regression. University of Southampton.

Kisku, D.R., P. Gupta and J.K. Sing, 2010. Feature level fusion of biometrics cues: Human identification with Doddington's Caricature. Proceedings of the International Conference of Security Technology, Communications in Computer and Information Sciences, (STCCIS' 10), pp: 157-164.

Kondoz, A., 1969. Digital Speech: Coding for Low Bit Rate Communication Systems. 2nd Edn., Wiley, England, ISBN-10: 0470870087, pp: 458.

Rabiner, L.R. and B.H. Juang, 1993. Fundamental of speech recognition liveness verification in audio-video speaker authentication. University of Canberra.

Ramli, D.A., S.A. Samad and A. Hussain, 2008. Score information decision fusion using support vector machine for a correlation filter based speaker authentication system. Adv. Soft Comput., 53: 235-242.

Ramli, D.A., S.A. Samad and A. Hussain, 2009. A multibiometric speaker authentication system with SVM audio reliability indicator. IAENG Int. J. Comput. Sci.

Samad, S.A., D.A. Ramli and A. Hussain, 2007. A multi-sample single-source model using spectrographic features for biometric authentication. Proceedings of the 6th International Conference on Information, Communications and Signal Processing, Dec. 10-13, IEEE Xplore Press, Bangi, pp: 1-5. DOI: 10.1109/ICICS.2007.4449710

Sanderson, C. and K.K. Paliwal, 2003. Noise compensation in a person verification system using face and multiple speech features. Pattern Reco., 2: 293-302. DOI: 10.1016/S0031-3203(02)00031-6

Vasuhi, S., V. Vaidehi, N.T.N. Babu and T.M. Treesa, 2010. An Efficient Multi-modal Biometric Person Authentication System using Fuzzy Logic. Proceedings of the 2nd International Conference on Advanced Computing, Dec. 14-16, IEEE Xplore Press, India, pp: 74-81. DOI: 10.1109/ICOAC.2010.5725365

Wan, V. and W.M. Campbell, 2000. Support vector machines for speaker verification and identification. Proceedings of the IEEE Signal Processing Society Workshop, (ISPSW' 00), IEEE Xplore Press, pp: 775-784. DOI: 10.1109/NNSP.2000.890157