



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Performances of Qualitative Fusion Scheme for Multi-biometric Speaker Verification Systems in Noisy Condition

Lydia Abdul Hamid and Dzati Athiar Ramli

Intelligent Biometric Research Group (IBG), School of Electrical and Electronic Engineering,
Engineering Campus, Universiti Sains Malaysia, 14300, Nibong Tebal,
Pulau Pinang, Malaysia

Abstract: Fusion of multiple modalities becomes a good strategy to improve the performance in biometric speaker verification as an audio based system accuracy decreases severely under noisy condition. However, for simple sum rule fusion scheme, this approach is only helpful if both systems have equal performances because it considers both models equally regardless of its conditions. As a consequence, a weighted sum rule is then experimented. Instead of varying the weight, this method calculates the sum of Equal Error Rate (EER) percentages produced from each modality for fusion weight computation. In this study, the information in term of Mel Frequency Cepstral Coefficient (MFCC) of speech signal is extracted while Region of Interest (ROI) of lip images has been used as a second modality for the multi-modal systems. The Support Vector Machine (SVM) classifier is executed for the verification system. From the experiment results, EER performances using simple sum rule and weighted sum rule at 35 dB SNR of speech signal and 0.9 noise density of visual signal are observed as 0.1548 and 0.1487%, respectively.

Key words: Multi-modal, simple sum rule, weighted sum rule, mel frequency cepstral coefficient

INTRODUCTION

Biometrics consists of physical or behavioral traits to uniquely recognize a person from others. In computer science, biometric is used as a form of identity access control and access management. Biometrics characteristics can be divided into two classes which are physiological and behavioral. Physiological refers to the characteristic of the body such as fingerprint, face and iris. Behavioral is related to the person's behavior which are voice, gait and rhythm (Reynolds, 2002).

Biometric systems can operate in two types of modes which are verification and identification (Zhang, 2000). For verification, the user's input is compared with the system database. The input is genuine if the degree of similarity is sufficiently high. It can be implemented using two different approaches i.e., text-dependent and text-independent. For text-dependent, the same text is used for both training and testing while for text-independent, different text is used during testing (Khalaf *et al.*, 2011). Identification is different from verification since the system will determine the user's identity from a large set of possible identities. Therefore, the user's biometric input is compared with all the persons enrolled in the

database. As a result, the system will identify which person enrolled in the database has the highest degree of similarity with the user's input else a decision indicating that the user presenting the input is not an enrolled user (Ribaric *et al.*, 2005).

Now-a-days, the recognition system has widely implemented with biometric security system since it is more efficient compared to previous methods which required key, password or smart card. This is because, these methods have several problems such as the cards may be missing and the password may be forgotten (Cambell, 1997; Li *et al.*, 2002).

Besides that, biometric system has improved by applying automatic recognition of a person based on physiological or behavioral characteristics in many areas of applications for examples access control system, internet banking and law enforcement. According to Vijayaprasad *et al.* (2010), one of the common biometric traits that are used for verifying the user identity is the fingerprint images. Many studies have been done to overcome the problems occur due to partial captured fingerprints. However, biometric systems which only consider single modal have several limitations and tend to obtain low performance especially in noisy condition. The

major problem occurs because the physical appearance and behavioral characteristics of a person tends to vary with time (Kittler *et al.*, 1997).

One of the solutions to this problem is to propose multimodal system as reported by Hong *et al.* (1999) and Fox and Reilly (2004). Hong and Jain (1998) have proposed a multi-modal personal identification system which integrates face and fingerprints. Further reports on multi-biometric systems are reviewed by Ross and Jain (2004) and Ross *et al.* (2006). Since multi-modal systems provide significant advantages compared to single-modal systems, a fusion of speech and lip information is experimented in this study.

The objective of this study was to evaluate the performance of the audio and visual fusion systems based on different fusion scheme namely simple sum rule and weighted sum rule. For validation, three types of experiments have been designed in this study i.e., (1) Fusion under noisy audio and clean visual conditions, (2) Fusion under clean audio and noisy visual conditions and (3) Fusion under noisy audio and noisy visual conditions.

MATERIALS AND METHODS

Both audio and visual data are obtained from audio-visual digit database (Sanderson and Paliwal, 2001). This database contains digitized audio signals which monophonic 16 bit, 32 kHz and in WAV format corresponding to the recording voices of 37 speakers (16 female and 21 male). The recording is done in three different sessions. Each speaker performed 20 repetitions of digit zero for each session. Therefore, 60 audio data for each speaker obtained from each session. It consists of 2220 data for digit zero. The visual data of 37 speakers is stored as a sequence of JPEG images with a resolution of 512×384 pixels. Each speaker consists of 60 sequences of face images (20 sequences from each session) hence in total of 2220 images from entire speakers. Two techniques, face detection and lip localization have been developed in order to identify the lip's location. In face detection process, the colour-based technique is used to differentiate skin region from the non-skin region. Lip region is separated from face skin by using the saturation colour threshold during the lip localization process (Ramli *et al.*, 2008). For the purpose of this study, the first session are used for speaker's model while data from second and third sessions are used for testing data.

In this study, three modules have been developed i.e., features extraction, classification and fusion. For feature extraction, two types of features have been extracted i.e. lip and speech signal information. For lip, the information is extracted in term of Region of Interest

(ROI). According to Samad *et al.* (2007), ROI of a speaker's mouth region view few image pixels of the mouth. ROI for lower face can also provide information which is more computational effective in term of storage compared to the whole face processing. For speech signal, the information based on Mel Frequency Cepstral Coefficient (MFCC). Both techniques can be found by Furui (2001), respectively. MFCC is processed in Fourier Transform (FT). First, all frames of the signal are computed using the Discrete Fourier Transform (DFT). The result obtained is referred as signal's spectrum. The second step is the filter bank processing. The filter bank formed spectral features at defined frequency at its exit. Third, log energy computation which consists of computing the logarithm of the square magnitude of the filter bank outputs is performed. 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.

There are two types of classification procedures which are supervised classification and unsupervised classification. In a supervised classification, the user's claim is compared with the given database which has been learned by the system. These samples are referred to as training areas.

Few classification techniques such as Artificial Neural Network (ANN), Dynamic Time Wrapping (DTW) and Hidden Markov Models (HMM) have been studied by the researchers as described by Ilyas *et al.* (2010) and Al-Hadidi (2012). One of the applications of SVM is in fingerprint recognition where it can learn to recognize handwritten digits by examining a large collection of scanned images of handwritten. SVM has also been successfully applied to an increasingly wide variety of biological applications. One of the application of SVM in biomedical is the automatic classification of microarray gene expression profiles.

Other biological applications of SVMs involve classifying objects as diverse as protein and DNA sequences, microarray expression profiles and mass spectra. Sani *et al.* (2010) reported the use of SVM in face detection application in order to extract the face and non-face regions.

This experiment uses SVM as a classifier. A SVM is a computer algorithm that learns by example to assign labels to objects. The theory of SVM can be found by Wan and Campbell (2000) and Gunn (2005). To summarize, let W be normal to the decision region and the N training examples represented as the pair (W_i, X_i) , $i = 1, 2, \dots, N$ where, $-1 \leq \alpha_i \leq 1$. The points that lie on the hyper plane to separate data satisfy:

$$\langle w, x \rangle + \alpha = 0 \tag{1}$$

where, α is the distance of the hyper plane from the origin. The hyperplane should have the same distance for each class nearest point and the margin distance is twice. The non-linear mapping is imperative in the case of the linear boundary is inappropriate. In practice, the kernel function is introduced to transform the data points in the input space to the feature space. In this study, the polynomial function is used.

The third module involves fusion and two methods of fusion schemes are proposed in these experiments namely fusion simple sum rule and weighted sum rule. Figure 1 summarizes the features extraction, classification and fusion that involves in this experiment. In fusion module, score level fusion scheme is considered for the combination of scores from all subsystems. The scores from each subsystem are then normalized using min-max normalization techniques. Both scores from speech and lips are normalized using Eq. 2.

The fusion performance would not be at its best performance due to the matching score for audio and visual are at different numerical range. Therefore, the matching score need to be normalized using the

normalization method (Wang *et al.*, 2007). The minimum and maximum scores are converted to 0 and 1, respectively. Given matching scores $\{X\}$, $k = 1, 2, \dots, n$. The normalized scores are given by (Jain *et al.*, 2005):

$$Y = \frac{X - \min X}{\max X - \min X} \tag{2}$$

Consequently, the descriptions of the fusion schemes are explained as the following.

Simple sum rule: The normalized scores obtained from both subsystems are summed and divided by the total number (N) of subsystems. The simple sum rule is described in Eq. 3:

$$S = \frac{1}{N} \sum_m^n Y_m = Y_1 + Y_2 \tag{3}$$

For the Simple sum rule based, equal weights are used for each subsystem in these experiments. This method need to equally consider both subsystems even though one of the subsystems is in a noisy condition. Therefore, system performances decrease severely when one of the modality is corrupted.

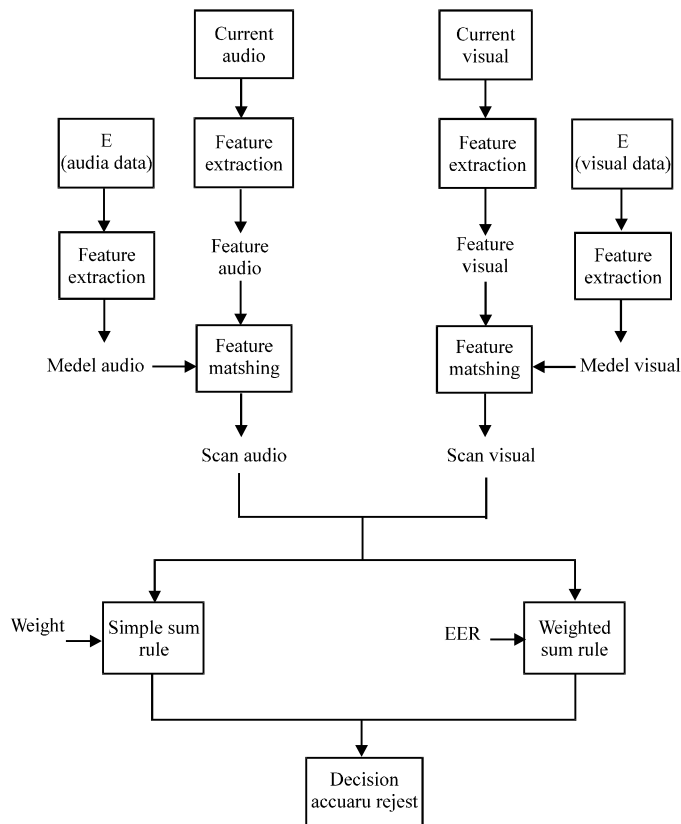


Fig. 1: Architecture of multibiometric system

Weighted sum rule: Preliminary study on weight adaptation can be found by Ramli *et al.* (2009). In this study, the weight adaptation scheme is based on weighted sum rule method is determined based on the performance of each single system by utilizing the Equal Error Rate (EER) value (Scheidat *et al.*, 2007). The weights of each subsystem are calculated by dividing the single EER with sum of all EERs as shown in Eq. 4:

$$W_m = \left[\frac{\frac{1}{EER_m}}{\sum_{m=1}^N \frac{1}{EER_m}} \right] \quad (4)$$

Noted that $0 \leq W_m \leq 1$ and:

$$\sum_{m=1}^N W_m = 1$$

The weight W_m are inversely proportional to the error of the corresponding biometric subsystem.

Therefore, this approach is more accurate for the fusion performances. A property of this weighted sum rule scheme is that the system which is obtained the highest EER is multiplied with the smallest weight and vice versa.

RESULTS AND DISCUSSION

In this study, the system performances are evaluated using Genuine Acceptance Rate (GAR), False Acceptance Rate (FAR) and Equal Error Rate (EER). In term of sensitivity or, GAR can be explained as the percentage of authorized individuals is admitted by the system:

$$GAR = \left[\frac{\text{Rejected genuine}}{\Sigma_{\text{genuine}}} \right] 100\% \quad (5)$$

and:

$$GAR = 1 - FRR \quad (6)$$

On the other hand, FAR or imposter pass rate is the percentage in which unauthorized individuals are accepted by the system or in term in specificity, it is the percentage of unauthorized individuals is correctly rejected. FAR is calculated as:

$$FRR = \left[\frac{\text{Accepted imposter}}{\Sigma_{\text{imposter}}} \right] 100\% \quad (7)$$

Studies by Ping *et al.* (2012) discussed that the system is at its best performance when

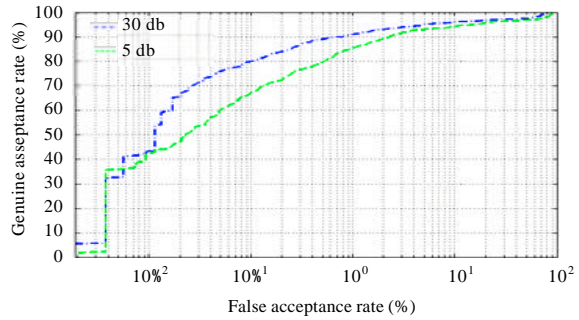


Fig. 2: Performances of MFCC-ROI-SVM system for noisy audio and clean visual for simple sum rule fusion scheme

Table 1: EER performances of MFCC-ROI-SVM for noisy audio and clean visual for simple sum rule fusion scheme

Noisy audio	-5	5	10	15	20	25	30	35	40
EER	24.4942	15.4608	10.2862	4.5195	1.4665	0.1286	0.0056	0.0038	0.0028

Table 2: EER performances of MFCC-ROI-SVM for noisy audio and clean visual for weighted sum rule fusion scheme

Noisy audio	-5	5	10	15	20	25	30	35	40
EER	0.2787	0.2745	0.2721	0.2703	0.2659	0.1401	0.0048	0.0031	0.0019

both FRR and FAR are at the lowest values. Therefore, higher recognition system will be achieved.

Fusion of noisy audio and clean visual (simple sum rule):

Table 1 shows the performances of the MFCC-ROI-SVM systems for noisy audio and clean visual using simple sum rule fusion scheme. For this scheme, the weight is fixed to 0.5.

The system performances based on GAR and FAR for 5 and 30 dB are compared in Fig. 2. The GAR performances for 5 and 30 dB at 1% FAR are observed as 85 and 92%, respectively.

Fusion of noisy audio and clean visual (weighted sum rule):

Table 2 shows the performances of the MFCC-ROI-SVM systems for noisy audio and clean visual using weighted sum rule fusion scheme. For this method, the weight is calculated using Eq. 4.

The system performances based on GAR and FAR for 5 and 30 dB are compared in Fig. 3. The GAR performances for 5 and 30 dB at 1% FAR are observed as 96 and 99%, respectively.

For comparison, Fig. 4 shows the EER performances of the multi-modal system at -5 dB for simple sum rule and weighted sum rule which are 0.028 and 0.0019%, respectively. Study shows that the GAR performances of

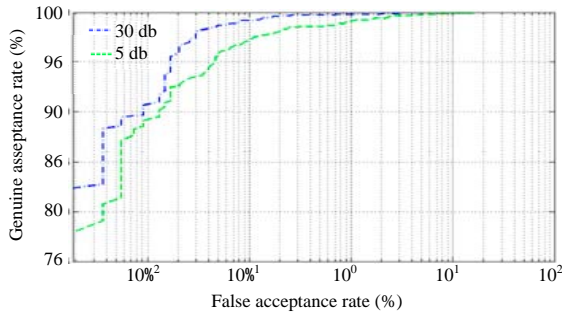


Fig. 3: Performances of MFCC-ROI-SVM system for noisy audio and clean visual for weighted sum rule fusion scheme

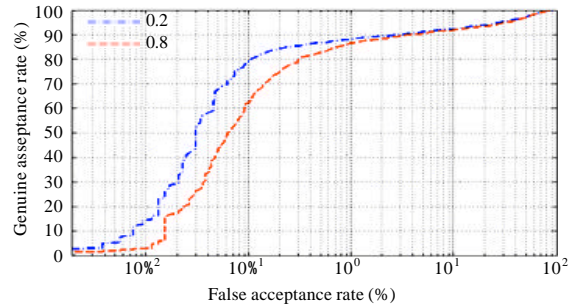


Fig. 5: Performances of MFCC-ROI-SVM system for clean audio and noisy visual for simple sum rule fusion scheme

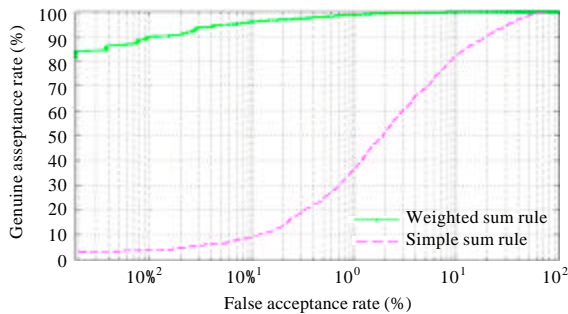


Fig. 4: EER Performances at different fusion scheme

Table 3: EER performances of MFCC-ROI-SVM for clean audio and noisy visual for simple sum rule fusion scheme

Noisy visual	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
EER	0.0676	0.1164	0.2196	1.5002	1.9075	4.4865	6.4741	7.9450	10.8288

Table 4: EER performances of MFCC-ROI-SVM for clean audio and noisy visual for weighted sum rule fusion scheme

Noisy visual	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
EER	0.0585	0.1070	0.1238	0.6954	0.8446	0.9882	1.0295	1.1289	1.1599

these systems at 1% FAR are observed as 99.8 and 35%, respectively. It also proves that the multi-modal system with weighted sum rule fusion scheme gives a good improvement in the EER performances.

Fusion of clean audio and noisy visual (simple sum rule):

Table 3 shows the performances of the MFCC-ROI-SVM systems for clean audio and noisy visual using simple sum rule fusion scheme. For this scheme, the weight is fixed to 0.5.

The system performances based on GAR and FAR for 0.2 quality densities and 0.8 quality densities are compared in Fig. 5. The GAR performances for

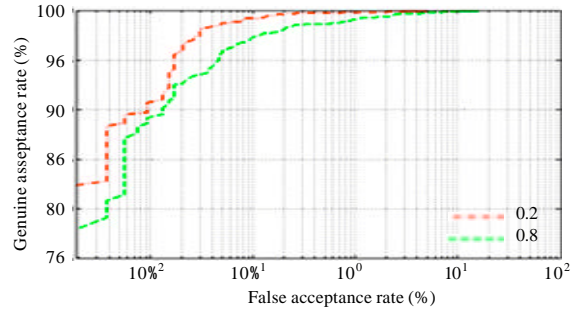


Fig. 6: Performances of MFCC-ROI-SVM system for clean audio and noisy visual for weighted sum rule fusion scheme

0.2 quality densities and 0.8 quality densities at 1% FAR are observed as 88 and 86%, respectively.

Fusion of clean audio and noisy visual (weighted sum rule):

Table 4 shows the performances of the MFCC-ROI-SVM systems for clean audio and noisy visual using weighted sum rule fusion scheme. For this method, the weight is calculated using Eq. 4.

The system performances based on GAR and FAR for 0.2 quality densities and 0.8 quality densities are compared in Fig. 6. The GAR performances for 0.2 quality densities and 0.8 quality densities at 1% FAR are observed as 99 and 97%, respectively.

For comparison, Fig. 7 shows the performances of the multi-modal system at 0.9 quality densities for simple sum rule and weighted sum rule which are 0.0676 and 0.0585%, respectively. Study shows that the GAR performances of these systems at 1% FAR are observed as 98.5 and 99%, respectively. It also proves that the multi-modal system with weighted sum rule fusion scheme gives a good improvement in the EER performances.

Table 5: EER performances of MFCC-ROI-SVM for noisy audio and noisy visual for simple sum rule fusion scheme

Visual	Audio									
	40	35	30	25	20	15	10	5	-5	
0.1	0.1548	0.658	0.2794	0.3453	0.6554	60.755	11.6873	18.0753	26.8928	
0.2	0.2204	0.1234	0.2965	0.5499	1.2045	7.0025	13.0623	19.9897	26.9921	
0.3	0.3963	0.2609	0.3463	1.0221	2.1896	8.6552	15.3106	22.9645	27.8065	
0.4	0.5433	0.8733	0.9552	1.4518	4.0812	10.2145	17.9542	25.1905	27.9887	
0.5	0.7723	1.0689	1.4011	3.4723	6.0998	12.4259	19.3741	27.7843	28.0000	
0.6	1.0862	2.0772	3.0261	4.2239	9.8775	15.9364	22.4521	31.0024	32.8912	
0.7	2.7450	3.8917	4.6678	8.7343	11.8872	18.8448	25.9643	33.9668	36.8666	
0.8	4.2861	4.0112	9.9972	10.7733	17.8463	23.9898	29.1601	35.2345	39.3023	
0.9	8.9912	9.2558	10.6175	15.2112	19.5111	25.8099	32.8465	39.1319	44.2990	

Table 6: EER performances of MFCC- ROI-SVM for noisy audio and noisy visual for weighted sum rule fusion scheme

Visual	Audio									
	40	35	30	25	20	15	10	5	-5	
0.1	0.1487	0.2120	0.2843	0.5115	0.7461	0.7728	0.8643	0.9135	0.9797	
0.2	0.2158	0.2244	0.2997	0.5153	0.9055	1.0159	1.2115	1.3168	1.5860	
0.3	0.2168	0.2973	0.3887	0.8178	1.3326	1.9781	2.2363	3.7185	4.1310	
0.4	0.3885	0.4092	0.4129	0.9199	1.4251	2.0210	2.7196	9.7219	11.4489	
0.5	0.5837	0.6009	0.6203	1.1369	1.5400	2.2418	3.8119	17.3522	19.5561	
0.6	0.7845	0.8001	0.8808	1.2901	1.5688	3.1025	4.1740	18.2061	30.3256	
0.7	1.0501	1.0625	1.0695	1.5913	1.8738	3.6223	4.4679	34.7965	39.5755	
0.8	1.0795	1.0804	1.0933	1.7350	1.9867	4.0456	5.3097	30.0573	43.5755	
0.9	1.2491	1.2506	1.2540	1.9739	2.1434	4.5828	5.9272	34.0922	44.1676	

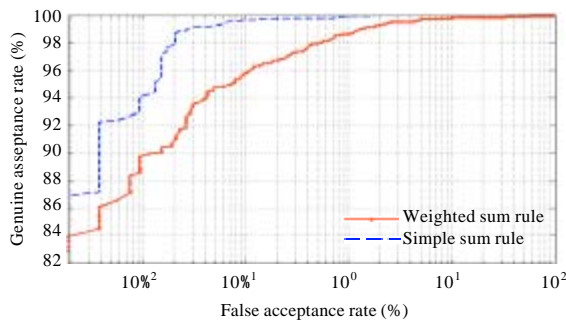


Fig. 7: EER performances at different fusion scheme

Fusion of noisy audio and noisy visual (simple sum rule):

Table 5 shows the performances of the MFCC-ROI-SVM systems for noisy audio and noisy visual using sum rule fusion scheme. For this method, the weight is fixed to 0.5.

The system performances based on GAR and FAR for 0.2 quality densities and 0.8 quality densities at 40 dB are compared in Fig. 8. The GAR performances for 0.2 quality densities and 0.8 quality densities at 1% FAR are observed as 95 and 93%, respectively.

Fusion of noisy audio and noisy visual (weighted sum rule):

Table 6 shows the performances of the MFCC-ROI-SVM systems for noisy audio and noisy visual using weighted sum rule fusion scheme. For this method, the weight is calculated using Eq. 4.

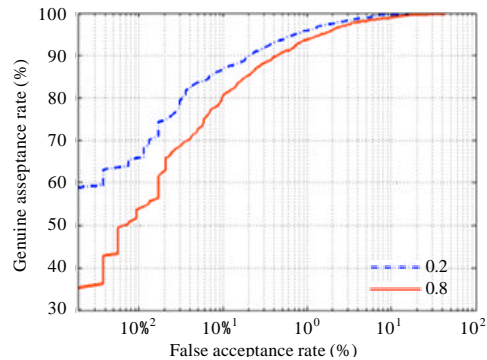


Fig. 8: Performances of MFCC-ROI-SVM system for noisy audio and noisy visual for simple sum rule fusion scheme

The system performances based on GAR and FAR for 0.2 quality densities and 0.8 quality densities at 40 dB are compared in Fig. 9. The GAR performances for 0.2 quality densities and 0.8 quality densities at 1% FAR are observed as 99 and 98%, respectively.

For comparison, Fig. 10 shows the performances of the multi-modal system at 35 dB and 0.9 quality densities for simple sum rule and weighted sum rule which are 0.1548 and 0.1487%, respectively. Study shows that the GAR performances of these systems at 1% FAR are observed as 95 and 99%, respectively. It also proves that the multi-modal system with weighted sum rule fusion scheme gives a good improvement in the EER performances.

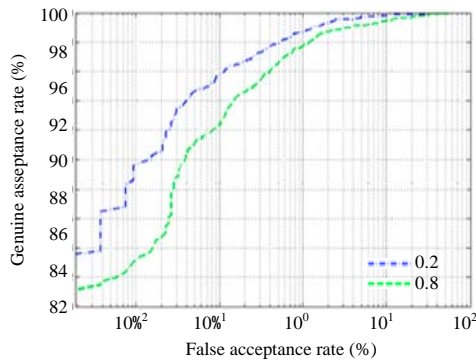


Fig. 9: Performances of MFCC-ROI-SVM system for noisy audio and noisy visual for weighted sum rule fusion scheme

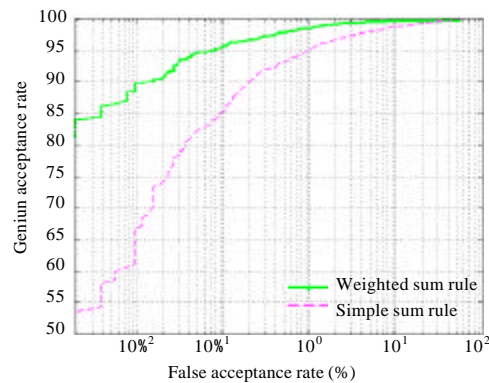


Fig. 10: EER Performance for different fusion scheme

CONCLUSION

The performances of single-modal biometric system and multi-modal biometric system are evaluated in this study. Implementation of multi-modal system with simple sum rule fusion scheme and multi-modal system with weighted sum rule fusion enhance the system performances compared to single system performance. From the results obtain, implementation of multi-modal system with weighted sum rule fusion scheme gives the best performance for the system.

ACKNOWLEDGMENT

This research is supported by the following research grants: Research University (RU) Grant, Universiti Sains Malaysia, 100/PELECT/814098 and Incentive Grant, Universiti Sains Malaysia.

REFERENCES

Al-Hadidi, M.R., 2012. Speaker identification system using autoregressive model. *Res. J. Applied Sci. Eng. Technol.*, 4: 45-50.

Cambell, J.P., 1997. Speaker recognition: A tutorial. *Proc. IEEE Conf.*, 85: 1437-1462.

Fox, N.A. and R.B. Reilly, 2004. Robust multi-modal person identification with tolerance of facial expression. *Proceedings of the IEEE International Conference on System, Man and Cybernetics*, October 10-13, 2004, Piscataway, NJ., USA., pp: 580-585.

Furui, S., 2001. *Digital Speech Processing, Synthesis and Recognition*. 2nd Edn., Marcel Dekker Inc., USA., ISBN-13: 9780824704520, Pages: 452.

Gunn, S.R., 2005. *Support vector machine for classification and regression*. Technical Report, University of Southampton, Southampton UK.

Hong, L. and A.K. Jain, 1998. Integrating faces and fingerprints for personal identification. *IEEE Trans. Patt. Anal. Mach. Intell.*, 20: 1295-1307.

Hong, L., A. Jain and S. Pankanti, 1999. Can multibiometrics improve performance. *Proc. AutoID*, 99: 59-64.

Ilyas, M.Z., S.A. Samad, A. Hussain and K.A. Ishak, 2010. Improving speaker verification in noisy environments using adaptive filtering and hybrid classification technique. *Inform. Technol. J.*, 9: 107-115.

Jain, A., K. Nandakumar and A. Ross, 2005. Score normalization in multimodal biometric systems. *Pattern Recognit.*, 38: 2270-2285.

Khalaf, E., K. Daqrouq and M. Sherif, 2011. Modular arithmetic and wavelets for speaker verification. *J. Applied Sci.*, 11: 2782-2790.

Kittler, J., G. Matas, K. Jonsson and M.U.R. Sanchez, 1997. Combining evidence in personal identity verification systems. *Pattern Recog. Lett.*, 18: 845-852.

Li, Q., B.H. Juang, Q. Zhou and C.H. Lee, 2002. Automatic speaker recognition technology. *IEEE Acoust. Speech Signal Proces. Conf.*, 4: 4072-4075.

Ping, Z., D. Zhi-Ran and W. Run-Duo, 2012. Speaker recognition based on mathematical morphology. *Inform. Technol. J.*, 11: 154-159.

Ramli, D.A., S.A. Samad and A. Hussain, 2008. A UMACE filter approach to lipreading in biometric authentication system. *J. Applied Sci.*, 8: 280-287.

Ramli, D.A., S.A. Samad and A. Hussain, 2009. A multibiometric speaker authentication system with SVM audio reliability indicator. *IAENG Int. J. Comp. Sci.*, 36: 313-321.

- Reynolds, D.A., 2002. An overview of automatic speaker recognition technology. *IEEE Acoustics Speech Signal Proc.*, 4: 4072-4075.
- Ribaric, S., I. Fratric and K. Kis, 2005. A automatic verification system based on fusion of palmprint and face features. *IEEE Transact. Acoust. Speech Signal Proc.*, 4: 4072-4075.
- Ross, A. and A.K. Jain, 2004. Multimodal biometrics: An overview. *Proceedings of the 12th European Signal Processing Conference*, September 6-10, 2004, Vienna, Austria, pp: 1221-1224.
- Ross, A.A., K. Nandakumar and A.K. Jain, 2006. *Handbook of Multibiometrics*. 1st Edn., Springer, New York, USA.
- Samad, S.A., D.A. Ramli and A. Hussain, 2007. Lower face verification centered on lip using correlation filters. *Inform. Technol. J.*, 6: 1146-1151.
- Sanderson, C. and K.K. Paliwal, 2001. Noise compensation in a multi-modal verification system. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, May 7-11, 2001, Salt Lake City, UT., USA., pp: 157-160.
- Sani, M.M., K.A. Ishak and S.A. Samad, 2010. Classification using adaptive multiscale retinex and support vector machine for face recognition system. *J. Applied Sci.*, 10: 506-511.
- Scheidat, T., C. Vielhauer and J. Dittman, 2007. Single-semantic multi-instance fusion of handwriting based biometric authentication systems. *Proceedings of the IEEE International Conference on Image Processing*, September 16-October 19, 2007, San Antonio, TX., USA., pp: 393-396.
- Vijayaprasad, P., M.N. Sulaiman, N. Mustapha and R.W.O.K. Rahmat, 2010. Partial fingerprint recognition using support vector machine. *Inform. Technol. J.*, 9: 844-848
- Wan, V. and W.M. Campbell, 2000. Support vector machines for speaker verification and identification. *Proc. IEEE Signal Process. Soc. Workshop Neural Networks Signal Process.*, 2: 775-784.
- Wang, F., X. Yao and J. Han, 2007. Minimax probability machine multialgorithmic fusion for iris recognition. *Inform. Technol. J.*, 6: 1043-1049.
- Zhang, D.D., 2000. *Automated Biometrics: Technologies and Systems*. Kluwer Academic Publishers, Dordrecht.