Journal of Computer Science 10 (3): 530-543, 2014
ISSN: 1549-3636

doi:10.3844/jcssp.2014.530.543 Published Online 10 (3) 2014 (http://www.thescipub.com/jcs.toc)

QUALITY BASED SPEAKER VERIFICATION SYSTEMS USING FUZZY INFERENCE FUSION SCHEME

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

Received 2013-01-01; Revised 2013-02-07; Accepted 2013-12-03

ABSTRACT

Performances of single biometric speaker verification systems are outstanding in clean condition but drop significantly in noisy condition. Implementation of multibiometric systems is one of the solutions to this limitation. However, in order to ensure the performances of multibiometric systems are sustained, the optimum weight for the fusion system must be determined correctly according to the quality of current data. This study proposes the use of Fuzzy Inference System for weight inference. Two traits i.e., speech and lip are used while Support Vector Machine (SVM) is employed as the classifier in this study. The speech features are extracted using the Mel Frequency Cepstrum Coefficient (MFCC) method and the lip features are extracted using Region of Interest (ROI) method. The performances of single modal system (i.e., speech and lip) and multibiometric systems with sugeno and mamdani approaches are compared at different quality conditions in this study. Experimental results prove that the use of Fuzzy Inference System as weight inference is a very promising approach. For 15 dB SNR speech signal and 0.2 lip quality density, the GAR performances at FAR equals 0.1% for Mamdani-type, Sugeno-type, lip and speech systems are observed as 94, 95, 86 and 7%, respectively. In short, the proposed fusion scheme based on Fuzzy logic is able to maintain the performance of fusion system especially when one of the biometric sources is in noisy condition due to its capability to infer the correct fusion weight according to current data quality.

Keywords: Biometrics, Single Biometric System, Multibiometric System, Fuzzy Logic Fusion Scheme, Sugeno-type, Mamdani-type

1. INTRODUCTION

Previously, the traditional verification uses passwords, keys or smart cards which are less secure since few problems may occur due to forgotten password, duplicated keys or stolen smart cards. Nowadays, biometric data for verification systems are commercially used in data security, internet access, ATMs, network logins, credit cards and government records. More studies on biometric system have been done by researchers due to the increase of requirement of automatic information processing in many industrial fields (Chia and Ramli, 2011). Biometrics is defined as the development of statistical and mathematical methods applicable to data analysis problems in the biological sciences. Biometrics is also a technology, which uses various individual attributes of a person to verify his or her identity. Biometric characteristics can be divided into two main classes i.e., physiological and behavioral characteristics. Physiological characteristics refers to the human body such as face, fingerprints, palm print, iris, DNA, hand geometry and finger vein structure while behavioral characteristics are related to the actions of a person such as voice, keystroke dynamics, gait, typing rhythm and signature (Jain *et al*., 2004). This study implements biometric system for speaker verification systems. Speaker verification system is used to verify a person's claim from the enrollment database by using speech signal as the input data.

Single biometric systems have to face few limitations such as non-universality, noisy sensor data, large intra-

Corresponding Author: Lydia Abdul Hamid, Intelligent Biometric Research Group (IBG), School of Electrical and Electronic Engineering, Engineering Campus, Universiti Sains Malaysia, 14300, Nibong Tebal, Pulau Pinang, Malaysia

user variations and susceptibility to spoof attacks. For example, a single biometric system uses voice patterns to identify the individuals may fail to operate because of a noisy data signal captured by the system. Limitations faced by single biometric system can be overcome by applying the multibiometric system. Multibiometric system enhanced the matching accuracy of a biometric system in noisy condition as well as increases the population coverage with multiple traits (i.e., lip, iris, voice and face). Studies on multibiometrics are further discussed in Ben-Yacoub *et al*. (1999) and Pan *et al*. (2000). Besides that, multibiometric system may continuously operate even though a certain trait is unreliable due to user manipulation, sensor or software malfunctions. . However, this is only true when fusion scheme is done at the decision level where hard decision fusion for example or operator is executed. For the score level decision fusion, the multibiometric systems are at its best performance only when all traits operate in clean condition. In noisy condition, the unreliable speech signal tends to cause the system to obtain false scores for genuine and imposter signal. This problem does not occur in clean condition since both speech and lip signal gives reliable scores for genuine and imposter signal.

This study proposes the use of quality based score fusion approach to improve the performances of multibiometric systems. The quality based fusion depends on the input current condition. This method is very useful to ensure the speaker verification system is at its best performance especially in noisy condition. The quality based fusion implements the quality measure identification system to identify the quality of sample data. Researches on quality measure identification system have been discussed in Fierrez-Aguilar *et al*. (2005) and Nandakumar *et al*. (2008). In order to take full advantage of the quality based fusion approaches, this study implements the fusion mechanism for different biometric information. For this purpose, Fuzzy Inference System is developed so as to infer the optimum weight for robust and reliable multimodal biometric based security systems. The use of fuzzy logic as the fusion scheme for quality based fusion approach improves the system performances.

According to Vasuhi *et al*. (2010), the fuzzy logic decision-making is approximately the same with the human decision-making. Fuzzy design can accommodate the ambiguities of human languages and logics. It provides both an intuitive method for describing systems in human terms and automates the conversion of those system specifications into effective models. Fuzzy logic has the ability to add human-like subjective reasoning capabilities to machine intelligences as described in Prade and Dubois (1996). General block of fuzzy logic with Mamdani-type and Sugeno-type is shown in **Fig. 1**. Fuzzification is the process where each input is assigned to a lingustic variable. Degree of membership can be obtained from the lingustic variable. The degrees of membership are combined using fuzzy rules which may be expressed in terms such as “if x is A, then y is B”. The process of converting the fuzzy output based on the strength of membership is called defuzzification. Defuzzification is used in fuzzy modeling and in fuzzy logic control to convert the fuzzy outputs from the systems to crisp values.

There are two types of Fuzzy Inference System (FIS) i.e., mamdani and sugeno. A Mamdani-type FIS has fuzzy inputs and a fuzzy output. For Mamdani-type, the input is transformed into a set of linguistic variable during the fuzzification process. The Fuzzy Inference System (FIS) uses the input variables and fuzzy rule to derive a set of conclusion which will be used during the defuzzification process. A crisp number is the output of the defuzzification process (Jassbi *et al*., 2007). Mamdani-type FIS is widely accepted for capturing expert knowledge. It allows us to describe the expertise in more intuitive and human-like manner. The advantages of the Mamdani-type FIS are it have widespread acceptance, intuitive and well-suited to human inputs. However, Mamdani-type FIS entails a substantial burden.

In short, both Mamdani-type and Sugeno-type are similar in term of the fuzzification and rule evaluation process. The main different between Mamdani-type and Sugeno-type is the output of Sugeno-type is linear or constant. Besides that, Mamdani-type uses defuzzification method to extract the output while Sugeno-type uses weighted average method to extract the output. Sugeno-type FIS is computationally effective and works well with optimization and adaptive techniques, which makes it is very attractive in control problems, particularly for dynamic nonlinear systems. So that it works well with linear technique and well-suited to mathematical analysis FLT, 2010.

The first objective of this study is to analyze the performances of single modal system i.e., speech and lip at different quality conditions. Consequently, the Fuzzy Inference System is designed for weight inference. Finally, the performances of the fusion systems with weight inferred from FIS are compared to the performances of the single systems.

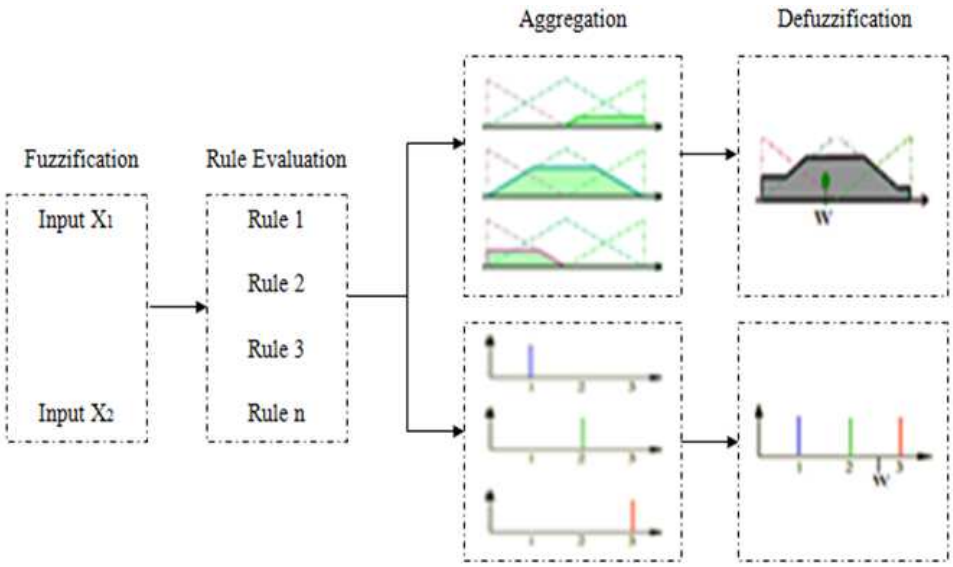


Fig. 1. Fuzzy logic with Mamdani-type and Sugeno-type

2. MATERIALS AND METHODS

Data Acquisition: In data acquisition, voice which is continuous electrical signal is converted to digital signal using a sampler and Analog-to-Digital (A/D) converter. The digitization process consists of sampling, quantization and coding. Sampling process is discussed extensively in (Rabiner and Schafer, 1978). After sampling process, the sampled signal is discrete in the time domain but still continuous in the amplitude domain. The quantization process divides the continuous amplitude range into finite subrange (Furui, 2000). Finally, the coding process is done by assigning these finite values into a sequence of codes for binary number representation.

In this study, the audio and visual data are obtained from Audio-Visual digit database (Sanderson and Paliwal, 2001). The database consists of 20 repetition of number zero from 37 different subjects. Mel Frequency Cepstrum Coefficient (MFCC) is used to obtain the features for speech modality. This study uses 12 MFCC features to form the feature vector. The data is collected in 32 kHz, 16-bit mono format. For the lip verification, the Region of Interest (ROI) of lip images are cropped and stored as JPEG files with resolution of 512×384 pixels. The ROI method to extract the lip features in this study as discussed in (Potamianos *et al*., 2000; Iyengar *et al*., 2001).

The database is divided to two sessions which are training and testing. During the enrolment process, 2220 audio data are developed for all 37 subjects. For training purposes, 740 data are used to train the system. Each subject is treated as the claimant and the other subjects as the imposters during the verification process. Therefore, the database has 40 testing data from the authentic speaker and 1440 from the imposter speaker. The visual data consists of 60 sequences of images (20 for training and 40 for testing) where each sequence consists of 10 images. In total, 22200 data are developed for all 37 subjects. Similar to speaker verification, each subject is treated as the claimant and the other subjects as the imposters during the verification process. Hence, the database has 400 testing data from the authentic lip image and 14400 from the imposter lip image.

2.1. Feature Extraction

A preemphasis of high frequencies is required to compress the signal dynamic range by flattening the spectral tilt in order to raise the SNR. The first order FIR filter is used to filtering the speech signal. The use of window function is important to minimize the signal discontinuities at the beginning and end of each frame by zeroing out the signal outside the region of interest. This study implements the Mel Frequency Cepstrum Coeficient (MFCC) processing to extract the audio features. There are few steps involved in MFCC process. First, all frames of the signal are computed using discrete Fourier transform. Next, the filter bank processing formed the spectral features at defined frequency at its exit. After that, log energy computation which consists of computing the logarithm of the square magnitude of the filter bank is performed. Finally, the mel frequency cepstrum is computed (Becchetti and Ricotti, 1999).

2.2. Classification

This study implements the Support Vector Machine (SVM) as classifier. A SVM performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories. SVM mode is a supervised learning method that generates input-output mapping functions from a set of labeled training data. The foundation of Support Vector Machines (SVM) has been developed as discussed in (Vapnik, 1995) and becomes popular and accepted nowadays due to many attractive features and promising empirical performance. Theory regarding SVM is further explained in (Gunn, 1998). In brief, decision boundary in support vector machine can be explained as presented in **Fig. 2**.

The SVM identifies the data points that are found to lie at the edge of an area in space which is a boundary from one class to another. The space between regions containing data points in different classes as being the margin between those classes. SVM is used to identify a hyperplane that separates the classes. The maximum margin between the different classes is found. An advantage of this method is that the modeling only deals with these support vectors, rather than the whole training dataset.

2.3. Fusion Scheme

A fuzzy fusion mechanism for robust and reliable multimodal biometric based security systems is developed. The use of fuzzy logic system as the fusion scheme improves the system performances. For this experiment, the fuzzy logic system consists of two inputs (speech and lip) and one output (weight). The parallel nature of the rules is one of the most important aspects in fuzzy logic (Hellmann, 2001). Initially, the input verification scores (speech and lip) are scaled to some range of score by using the min-max normalization equation as in Equation (1):

$$\hat{S}_i = \frac{s_i \min_{i=1}^{K} s_i}{\max_{i=1}^{K} s_i - \min_{i=1}^{K} s_i} \quad (1)$$

where denote the ith match score output and K is the number of the match scores available in the set (Jain *et al*., 2005).

The fuzzy logic system procedures are proposed as below (Zadeh, 1965; 1984).

Step 1: Fuzzification

In this study, there are two fuzzy models for Mamdani-type and Sugeno-type, respectively. Each model has two inputs, speech and lip and one output which is weight. **Figure 3** shows the fuzzy inference system using Mamdani-type and Sugeno-type method in Matlab Fuzzy Toolbox.

Next, the inputs are identified and the degree of each input is determined according to appropriate fuzzy sets via membership function. The membership functions are Gaussian shapes because it can covers several values in one membership. The inputs are always a crisp numerical value. For input 1 (speech), the interval is varied between [0, 40] SNR and for input 2 (lip), the interval is varied between [0, 1] quality density. The output (weight) is varied between [0, 1].

Then, the speech fuzzy set is modeled for three mfs: speech (Qlow), speech (Qmed) and speech (Qhigh) and three mfs are also modelled for the lip fuzzy set: lip (Qlow), lip (Qmed) and lip (Qhigh) as shown in **Fig. 4**. For the output fuzzy set, three mfs: weight (Qlow), weight (Qmed) and weight (Qhigh) are used. Output for Mamdani-type and Sugeno-type are as illustrated in **Fig. 5**.

Step 2: Rule Evaluation

For this study, there are nine rules for the system. From the experiment, lip performs better than speech. Therefore, this study relies more on lip since uncertainty inputs condition are involved during the process. For example, when both speech (Qhigh) and lip (Qlow) are determined, the weight output is mapped to weight (Wmed). Rule editor is used to define the rules for each model. The rule editor for each model is shown in **Fig. 6**:

IF speech (Qlow) IF speech (Qmed) IF speech (Qhigh)
AND lip (Qhigh) AND lip (Qhigh) AND lip (Qhigh)
THEN (Wlow) THEN (Wlow) THEN (Wmed)
IF speech (Qlow) IF speech (Qmed) IF speech (Qhigh)
AND lip (Qmed) AND lip (Qmed) AND lip (Qmed)
THEN (Wlow) THEN (Wlow) THEN (Whigh)
IF speech (Qlow) IF speech (Qmed) IF speech (Qhigh)
AND lip (Qlow) AND lip (Qlow) AND lip (Qlow)
THEN (Wmed) THEN (Wmed) THEN (Whigh)

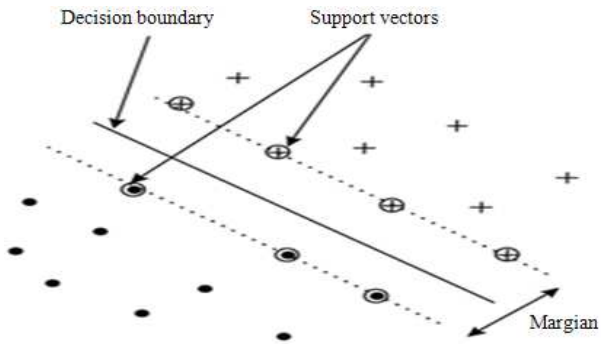


Fig. 2. Decision boundary in support vector machine

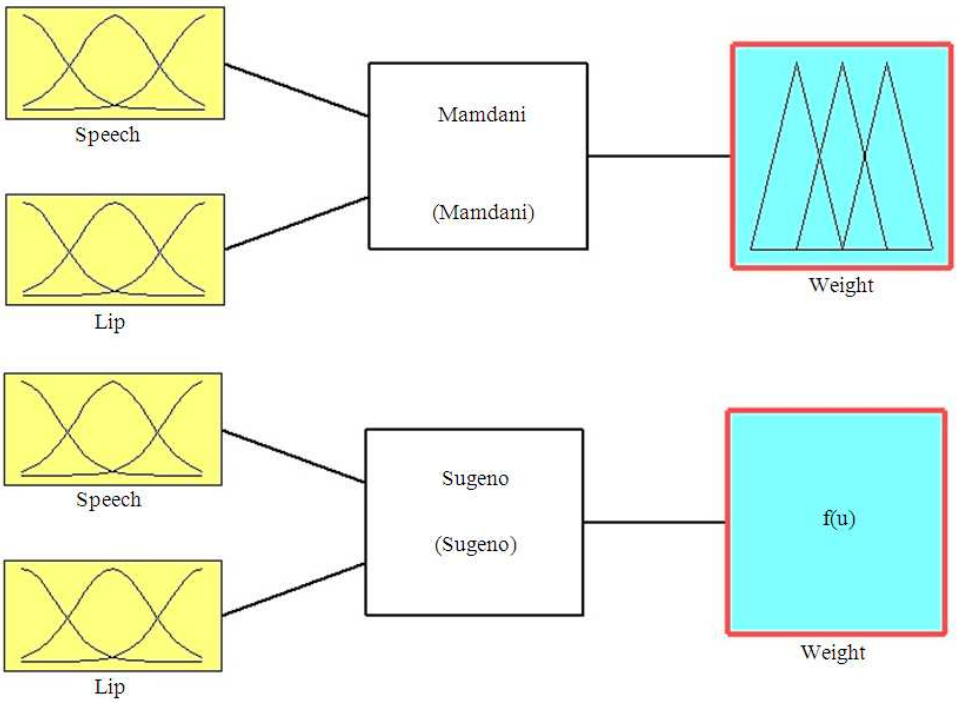


Fig. 3. Fuzzy Inference in Fuzzy Matlab Toolbox for Mamdani-type (top) and Sugeno-type (bottom)

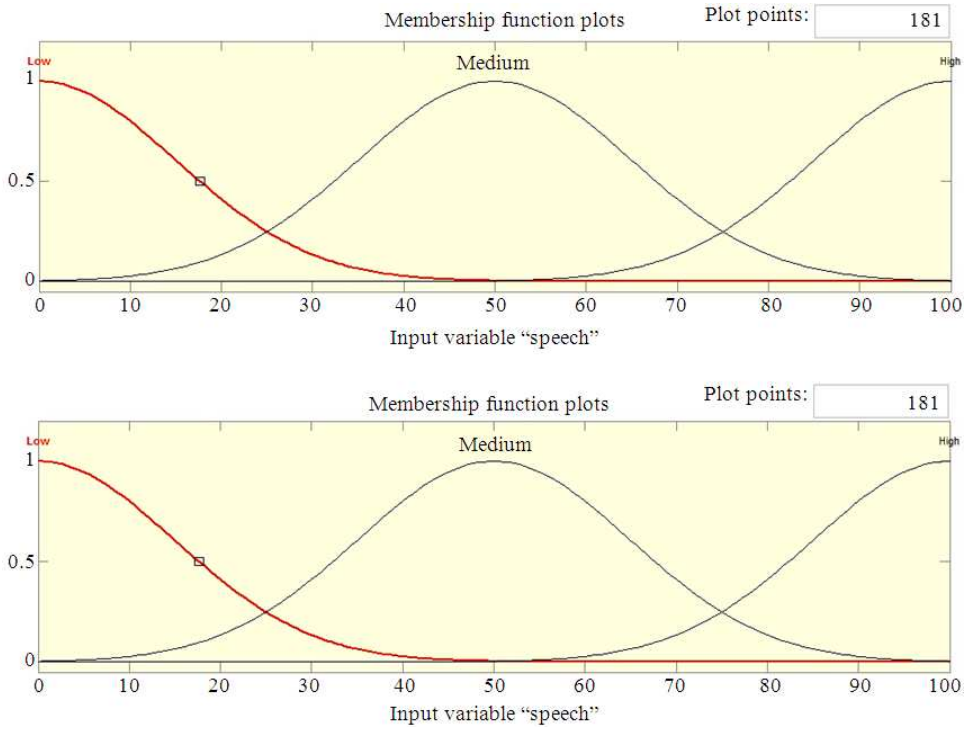


Fig. 4. Input Speech (top) and Input Lip (bottom) for Mamdani-type and Sugeno-type

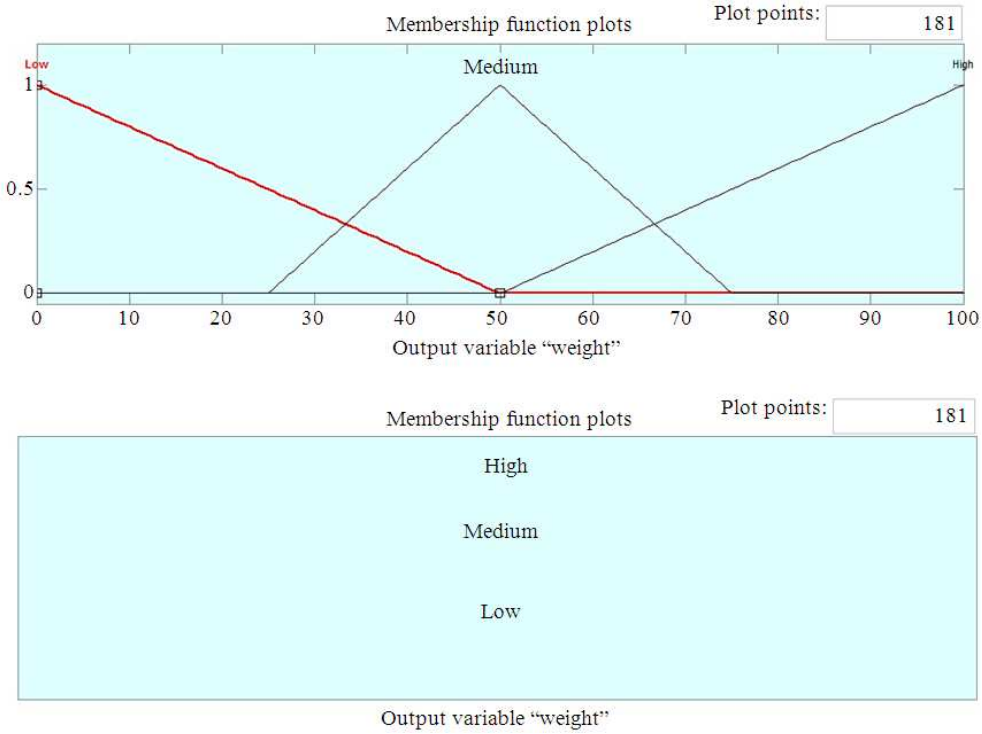


Fig. 5. Output for Mamdani-type (top) and Sugeno (bottom)

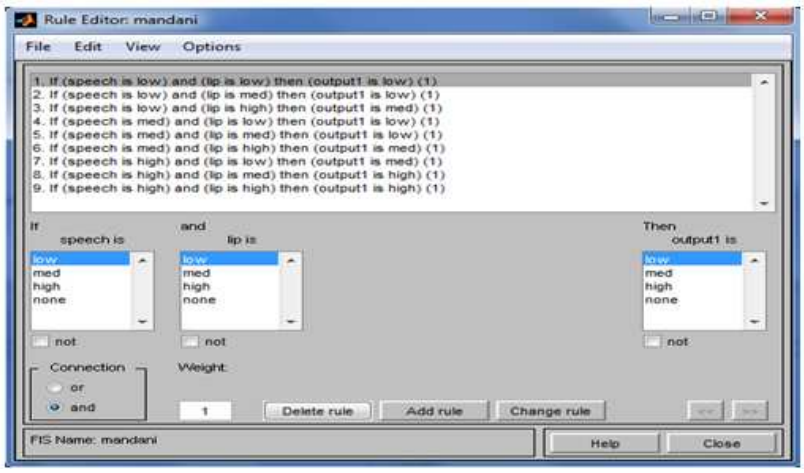


Fig. 6. Rule editor in fuzzy inference

Step 3: Aggregation

Aggregation is the process of unification of the outputs of all rules. The membership functions for all rules are scaled and combined into a single fuzzy set. The aggregation's inputs are the list of scaled membership functions and the output is one fuzzy set for each output variable. The Mamdani-type method and Sugeno-type method for aggregating the fuzzy rules and computing the output are shown in **Fig. 7 and 8**, respectively. All the rules must be combined and tested in order to make a decision.

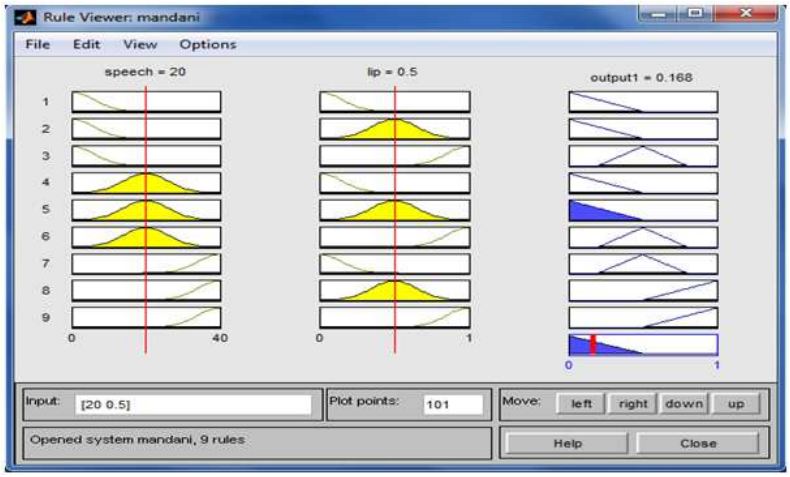


Fig. 7. Aggregation and defuzzification methods for Mamdani-type

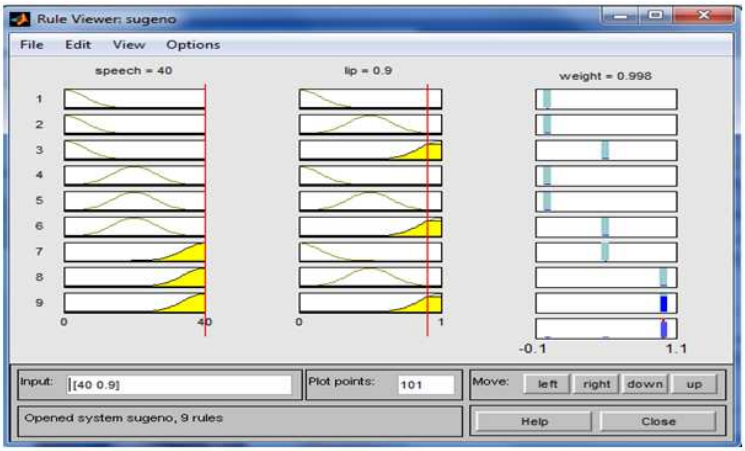


Fig. 8. Aggregation and defuzzification methods for Sugeno-type

Step 4: Defuzzification

The output of aggregation will be used as input for the defuzzification process and the output is a single number (weight). For defuzzification process, the Mamdani-type applied the centroid calculation method in order to obtain the centre of area under the curve while the Sugeno-type used the weighted average of few data points' method. The output (w) obtained from fuzzy logic system is implemented as in Equation (2) in order to calculate the fusion scores:

$$Y = wX_{speech} + (1-w)X_{lip} \quad (2)$$

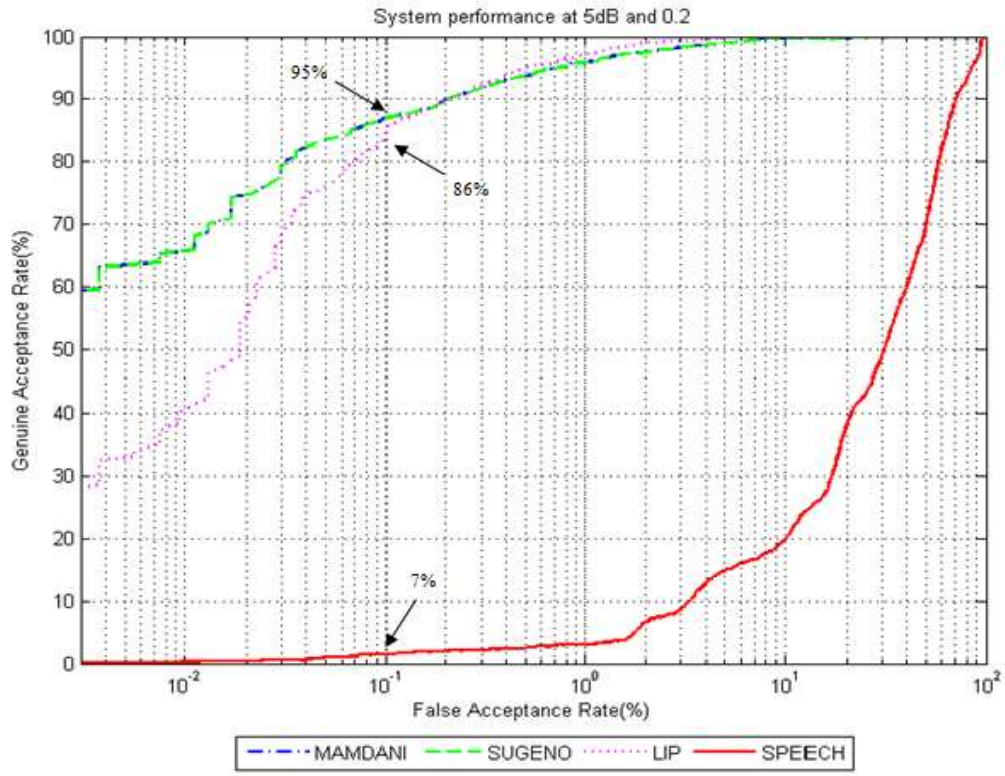
where, Y is the score and W is the weight applied to speaker's modality input data which are and respectively.

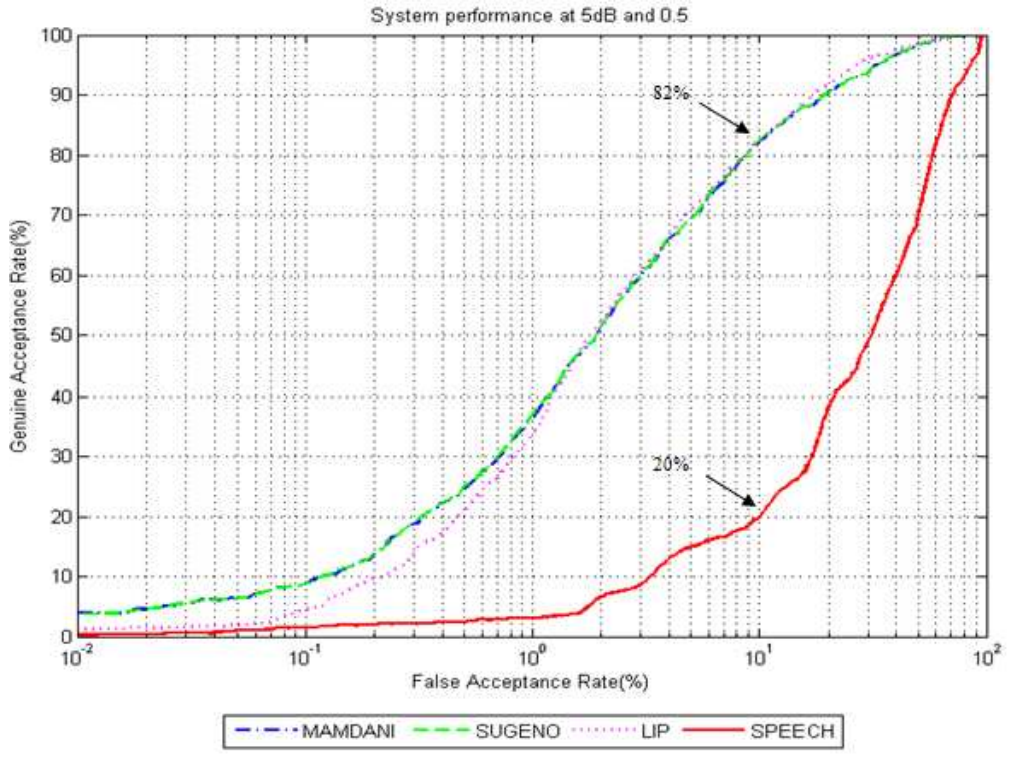
3. RESULTS

System performances for fuzzy logic fusion using Mamdani-type and Sugeno-type based on equal error rate (EER) at different levels of SNR are shown in **Table 1 and 2**, respectively. System performances based on receiver operation characteristic (ROC) showing the tradeoff between GAR and FAR percentages are then presented in **Fig. 9-11**.

Some results obtained by the single biometric and multibiometric system using Mamdani-type and Sugeno-type fusion method are also compared in terms of GAR and FAR at certain condition of speech and lip quality as illustrated in **Fig. 9-11**.

Figure 9 shows the performances of fusion systems compared to single systems at 5dB SNR with 0.2, 0.5 and 0.8 quality densities.


System performance at 5dB and 0.2
95%
86%
7%
Genuine Acceptance Rate(%)
False Acceptance Rate(%)
MAMDANI
SUGENO
LIP
SPEECH


System performance at 5dB and 0.5
82%
20%
Genuine Acceptance Rate(%)
False Acceptance Rate(%)
MAMDANI
SUGENO
LIP
SPEECH

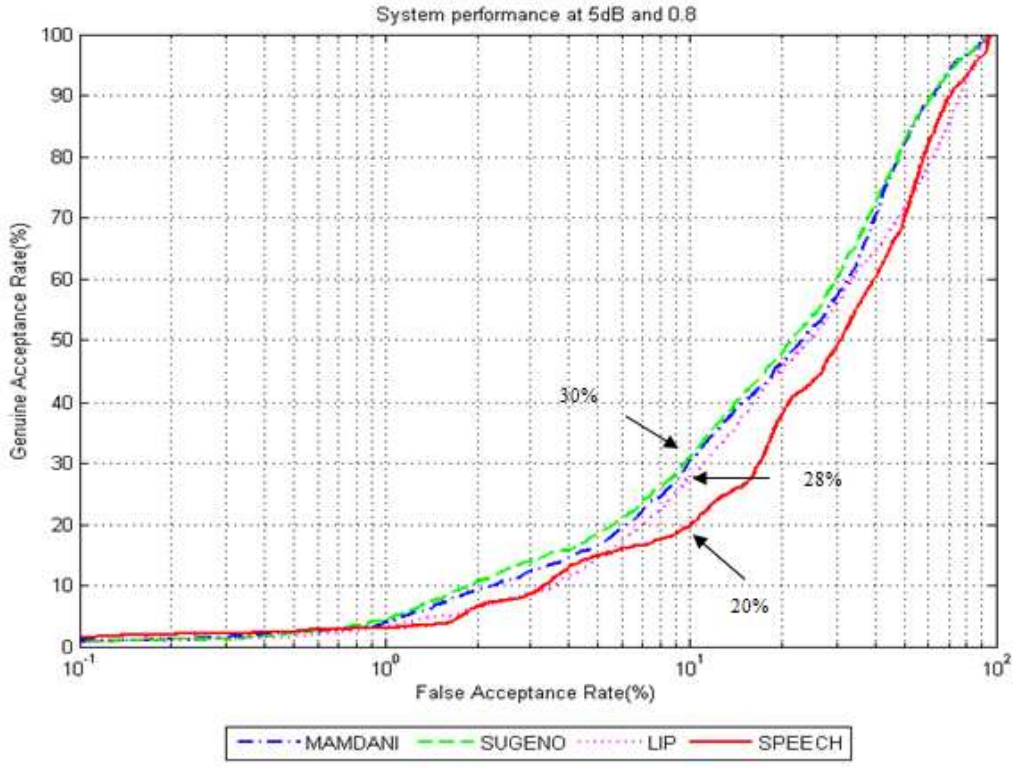
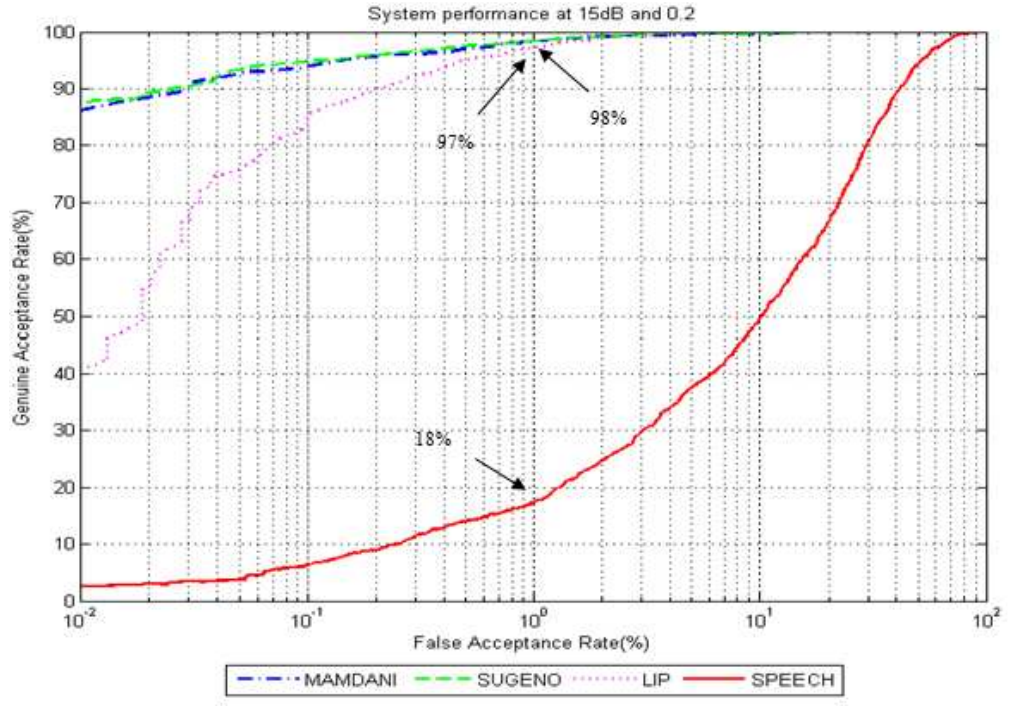


Fig. 9. The performances of fusion systems compared to single systems at 5dB SNR with 0.2, 0.5 and 0.8 quality densities



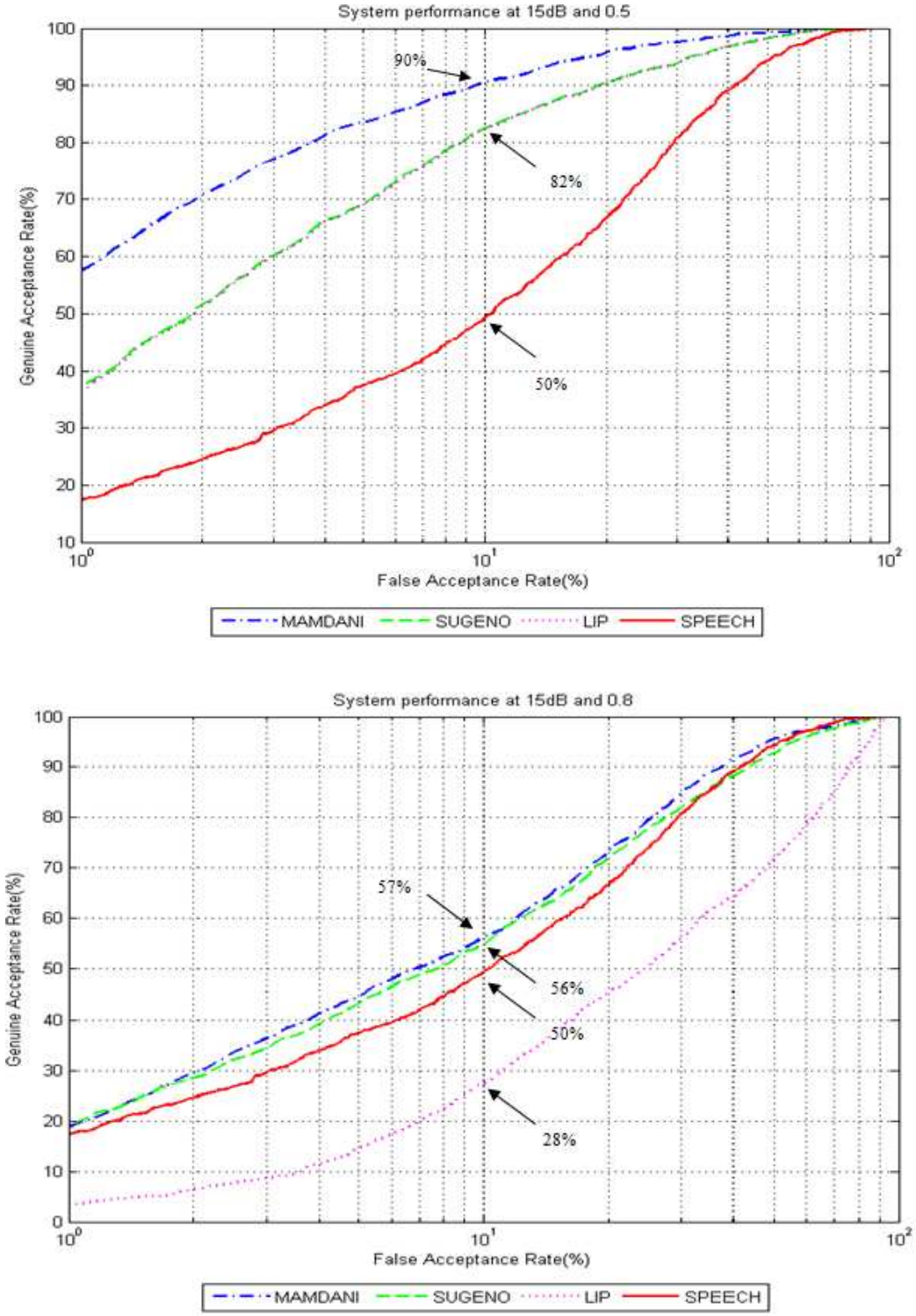
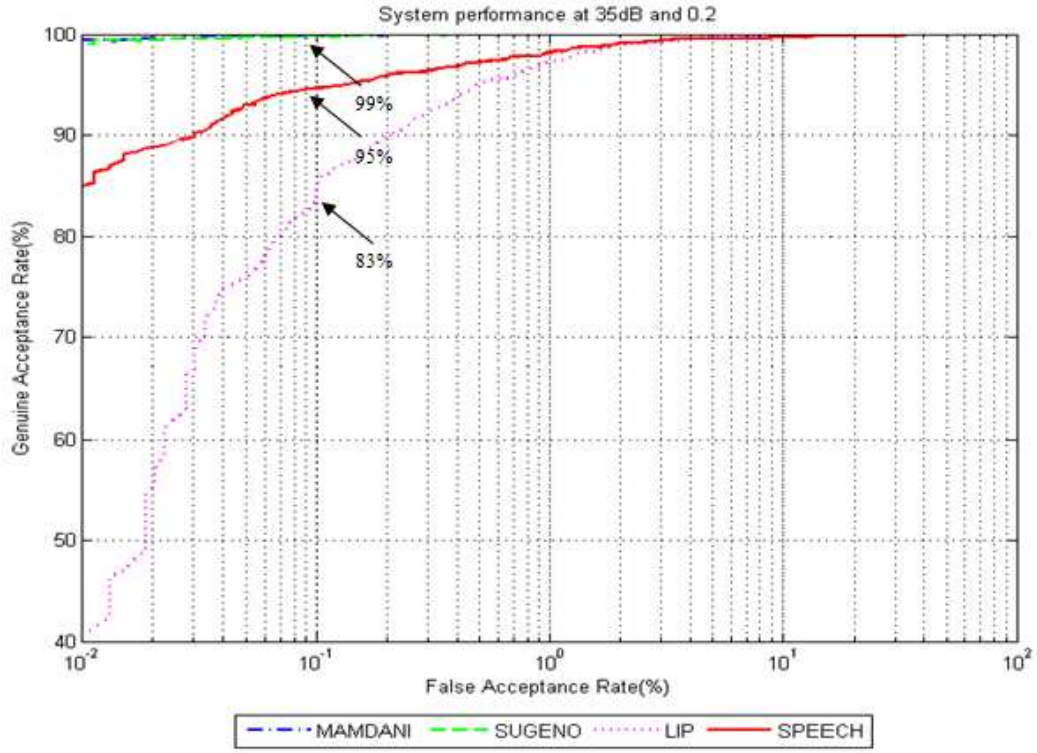
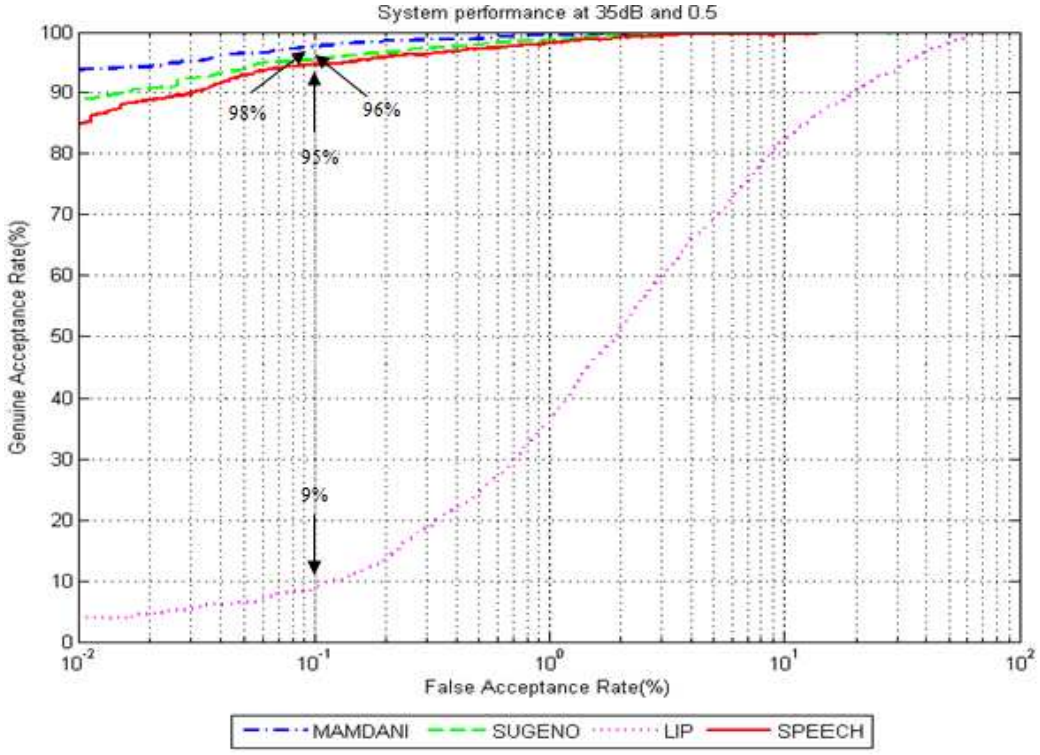


Fig. 10. The performances of fusion systems compared to single systems at 15 dB SNR with 0.2, 0.5 and 0.8 quality densities


System performance at 35dB and 0.2
Genuine Acceptance Rate(%)
False Acceptance Rate(%)
99%
95%
83%
MAMDANI
SUGENO
LIP
SPEECH


System performance at 35dB and 0.5
Genuine Acceptance Rate(%)
False Acceptance Rate(%)
98%
96%
95%
9%
MAMDANI
SUGENO
LIP
SPEECH

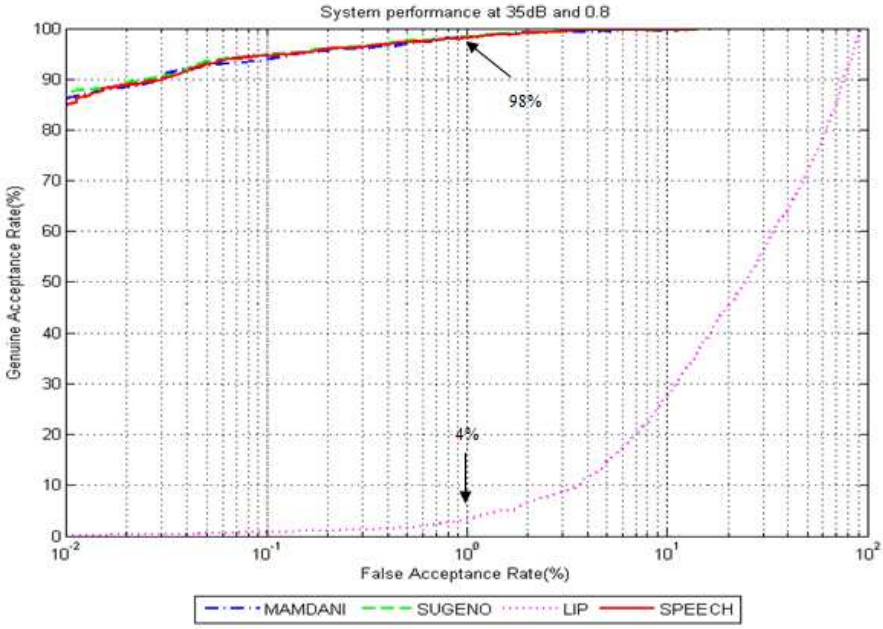


Fig. 11. The performances of fusion systems compared to single systems at 35 dB SNR with 0.2, 0.5 and 0.8 quality densities

Table 1. EER performances for fuzzy logic fusion using Mamdani-type

	Audio									
Visual	clean	40dB	35dB	30dB	25dB	20dB	15dB	10dB	5dB	-5dB
Clean	0.0428	0.0493	0.0529	0.0566	0.1036	0.2993	0.4774	0.8443	1.5429	2.1105
0.1	0.0492	0.0648	0.0601	0.0591	0.2018	0.3069	0.5818	1.1421	1.9454	2.3003
0.2	0.0511	0.0882	0.1104	0.0779	0.3904	0.7104	1.1997	2.4062	3.6421	5.7645
0.3	0.1384	0.3388	0.3010	0.3463	1.0126	1.8816	2.5723	4.0465	5.9056	10.0475
0.4	0.2056	0.6278	0.6072	0.9552	1.4251	3.8081	5.7508	7.6079	9.0465	15.0956
0.5	0.2964	0.7066	0.7423	1.4011	3.9054	5.9223	9.7147	11.9257	13.5839	20.7664
0.6	0.3119	0.7873	0.8399	3.0261	5.1242	9.1122	15.2843	17.6605	20.5227	25.1253
0.7	0.3805	0.7883	1.1562	4.6678	6.8975	10.4255	18.1961	23.7960	28.0265	29.9903
0.8	0.4377	0.7883	1.2106	5.0221	9.5918	16.8290	23.1231	28.9611	35.1328	39.5665
0.9	0.5622	0.7742	1.4884	5.6034	13.4722	19.4998	25.0901	32.2325	39.0888	43.2836

Table 2. EER performances for fuzzy logic fusion using Sugeno-type

	Audio									
Visual	clean	40dB	35dB	30dB	25dB	20dB	15dB	10dB	5dB	-5dB
Clean	0.0339	0.0489	0.0593	0.0627	0.2855	0.7642	0.9362	1.0072	1.1032	2.2117
0.1	0.0477	0.0666	0.0703	0.0976	0.8643	1.0811	1.0745	1.1924	2.0057	2.7555
0.2	0.0593	0.1342	0.1389	0.1952	1.2284	1.5907	1.6216	2.9034	3.7993	5.9015
0.3	0.3928	0.6607	0.6747	0.3987	2.8913	3.7172	3.8082	4.1225	5.1523	11.6776
0.4	0.5692	1.0801	0.9619	0.9196	6.5869	8.2226	8.3333	8.3343	8.3352	15.9945
0.5	0.6943	1.1421	1.1684	1.4310	9.4002	9.5126	10.6730	13.6806	13.6890	21.1034
0.6	0.6943	1.1355	1.1983	2.3020	10.2787	18.9921	21.5531	21.6282	21.8300	25.6724
0.7	0.8033	1.1233	1.2509	4.9278	12.6997	21.2828	23.6693	25.9741	27.6971	31.6770
0.8	0.8223	1.1515	1.2678	5.5572	13.1742	23.3183	24.8433	29.5069	36.4613	39.9001
0.9	0.8749	1.1780	1.2744	5.8708	14.9231	23.3183	26.1684	32.3931	39.1047	44.0005

When system at 5dB SNR and 0.2 quality density, GAR performances for Mamdani-type, Sugeno-type, lip and speech are evaluated as 88, 88, 83 and 2%, respectively, at 0.1% FAR. Meanwhile, at 5dB SNR and 0.5 quality density, GAR performances are observed as 82, 82, 81 and 20% for Mamdani-type, Sugeno-type, lip and speech, respectively at 10% FAR. Consequently, at 5dB SNR and 0.8 quality density, GAR performances for Mamdani-type, Sugeno-type, lip and speech equals to 30, 30, 28 and 20%, respectively at 10% FAR.

Subsequently, the performances of fusion systems compared to single systems at 15dB SNR with 0.2, 0.5 and 0.8 quality densities are illustrated in **Fig. 10**. When system at 15dB SNR and 0.2 quality density, GAR performances are observed as 94, 95, 86 and 7% for Mamdani-type, Sugeno-type, lip and speech respectively, at 0.1% FAR. Meanwhile, at 5dB SNR and 0.5 quality density, GAR performances are observed as 90, 82, 82 and 50% for Mamdani-type, Sugeno-type, lip and speech, respectively at 10% FAR. At the same FAR, i.e., 10%, when system at 5dB SNR and 0.8 quality density, GAR performances for Mamdani-type, Sugeno-type, lip and speech equals to 57, 56, 28 and 50%, respectively.

Finally, the performances of fusion systems compared to single systems at 35 dB SNR with 0.2, 0.5 and 0.8 quality densities are illustrated in **Fig. 11** below. The GAR performances for Mamdani-type, Sugeno-type, speech and lip are observed as 99%, 99%, 95% and 83%, respectively at 0.1% FAR when system at 35dB SNR and 0.2 quality density. While system at 35dB SNR and 0.5 quality density, the GAR performances for Mamdani-type, Sugeno-type, speech and lip are defined as 97, 96, 95 and 10%, respectively at 0.1% FAR. GAR performances of 96, 96, 96 and 2% are then observed for Mamdani-type, Sugeno-type, speech and lip, respectively at 0.1% FAR when system at 35dB SNR and 0.8 quality density.

4. DISCUSSION

From the experimental results illustrated in **Fig. 9-11**, it is observed that fusion systems based on Mamdani-type FIS and Sugeno-type FIS are able to increase the performances of single systems i.e., speech and lip when one of the traits is in clean condition or under minor quality degradation. Fusion systems based on Sugeno-type FIS and Mamdani-type FIS are observed as the most outstanding systems compared to the other fusion schemes.

Consequently, when both of the traits are severely corrupted by noise, the performances of single system tend to decrease. However, by implementing Sugeno-type FIS and Mamdani-type FIS fusion schemes, the systems are able to maintain its performances.

5. CONCLUSION

This study concludes a multibiometric verification system that combines both speaker and lip verification using fuzzy logic with Mamdani-type and Sugeno-type. Experimental results show that Mamdani-type and Sugeno-type are quite similar in accuracy performance and much better compared to the performances of single biometric systems. As a conclusion, the limitation faced by score level fusion in multibiometric system can be overcome using the fuzzy logic system due to its capability to infer the optimum weight according to the quality of verification data.

6. ACKNOWLEDGEMENT

This research is supported by the following research grants: Research University (RU) Grant, Universiti Sains Malaysia, 100/PELECT/814098 &100/PELECT/814161and Short Term Grant 304/PELECT/60311048,Universiti Sains Malaysia.

7. REFERENCES

Becchetti, C. and L.P. Ricotti, 1999. Speech Recognition: Theory and C++ Implementation. 1st Edn., Wiley, New York, ISBN-10: 0471977306, pp: 407.

Ben-Yacoub, S., Y. Abdeljaoued and E. Mayora, 1999. Fusion of face and speech data for person identity verification. IEEE Trans. Neural Netw., 10: 1065-1074. DOI: 10.1109/72.788647

Chia, C.L. and D.A. Ramli, 2012. Comparative study on feature, score and decision level fusion schemes for robust multibiometric systems. Frontier Comput. Educ., 133: 941-948. DOI: 10.1007/978-3-642-27552-4_123

Fierrez-Aguilar, J., J. Ortega-Garcia, J. Gonzalez-Rodriguez and J. Bigun, 2005. Discriminative multimodal biometric authentication based on quality measures. Patt. Recogn., 38: 777-779. DOI: 10.1016/j.patcog.2004.11.012

Furui, S., 2000. Digital Speech Processing: Synthesis and Recognition. 2nd Edn., Dekker, New York, ISBN-10: 0824704525, pp: 476.

Gunn, S.R., 1998. Support vector machines for classification and regression. University of Southampton.

Hellmann, M., 2001. Fuzzy logic introduction," epsilon nought radar remote sensing tutorials. The Pennsylvania State University.

Iyengar, G., G. Potamianos, C. Net, T. Faruquie and A. Verm, 2001. Robust detection of visual ROI for automatic speechreading. Proceedings of the IEEE 4th Workshop on Multimedia Signal Processing, Oct. 03-05, IEEE Xplore Press, Cannes, pp: 79-84. DOI: 10.1109/MMSP.2001.962715

Jain, A., K. Nandakumar and A. Ross, 2005. Score normalization in multimodal biometric systems. Patt. Recogn., 38: 2270-2285. DOI: 10.1016/j.patcog.2005.01.012

Jain, A.K., A. Ross and S. Prabhakar, 2004. An introduction to biometric recognition. IEEE Trans. Circuits Syst. Video Technol., 14: 4-20. DOI: 10.1109/TCSVT.2003.818349

Jassbi, J., S.H. Alavi, P.J.A. Serra and R.A. Ribeiro, 2007. Transformation of a mamdani FIS to first order sugeno FIS. Proceedings of the IEEE International Fuzzy Systems Conference, Jul. 23-26, IEEE Xplore Press, London, pp: 1-6. DOI: 10.1109/FUZZY.2007.4295331

Nandakumar, K., Y. Chen, S.C. Dass and A.K. Jain, 2008. Likelihood ratio-based biometric score fusion. IEEE Trans. Patt. Anal. Mach. Intell., 30: 342-347. DOI: 10.1109/TPAMI.2007.70796

Pan, H., Z.P. Liang and Z.P. Huang, 2000. Fusing audio and visual features of speech. Proceedings of the International Conference on Image Processing, Sep. 10-13, IEEE Xplore Press, Vancouver, BC., pp: 214-217. DOI: 10.1109/ICIP.2000.899333

Potamianos, G., A. Verma, C. Neti, G. Iyengar and S. Basu, 2000. A cascade image transform for speaker independent automatic speechreading. Proceedings of the IEEE International Conference on Multimedia Expo, Jul. 30-Aug. 02, IEEE Xplore Press, New York, pp: 1097-1100. DOI: 10.1109/ICME.2000.871552

Prade, H. and D. Dubois, 1996. What are fuzzy rules and how to use them. Fuzzy Sets Syst., 84: 169-185. DOI: 10.1016/0165-0114(96)00066-8

Rabiner, L.R. and R.W. Schafer, 1978. Digital Processing of Speech Signals. 1st Edn., Prentice Hall, Englewood Cliffs, ISBN-10: 0132136031, pp: 512.

Sanderson, C. and K.K. Paliwal, 2001. Noise compensation in a multi-modal verification system. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, May 07-11, IEEE Xplore Press, Salt Lake City, UT., pp: 157-160. DOI: 10.1109/ICASSP.2001.940791

Vapnik, N., 1995. The nature of Statistical Learning Theory. 2nd Edn., Springer-Verlag GmbH, New York, ISBN-10: 0387945598, pp: 188.

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, Chennai, pp: 74-81. DOI:10.1109/ICOAC.2010.5725365

Zadeh, L.A., 1965. Fuzzy sets. Inform. Control, 8: 338-353. DOI: 10.1016/S0019-9958(65)90241-X

Zadeh, L.A., 1984. Making computers think like people. IEEE Spectrum, 8: 26-32. DOI: 10.1109/MSPEC.1984.6370431