

## Intelligent Voice-Based Door Access Control System Using Adaptive-Network-based Fuzzy Inference Systems (ANFIS) for Building Security

Wahyudi, Winda Astuti and Syazilawati Mohamed

Intelligent Mechatronics Systems Research Group, Department of Mechatronics Engineering  
International Islamic University Malaysia, P.O. Box, 10, 50728, Kuala Lumpur Malaysia

---

**Abstract:** Secure buildings are currently protected from unauthorized access by a variety of devices. Even though there are many kinds of devices to guarantee the system safety such as PIN pads, keys both conventional and electronic, identity cards, cryptographic and dual control procedures, the people voice can also be used. The ability to verify the identity of a speaker by analyzing speech, or speaker verification, is an attractive and relatively unobtrusive means of providing security for admission into an important or secured place. An individual's voice cannot be stolen, lost, forgotten, guessed, or impersonated with accuracy. Due to these advantages, this paper describes design and prototyping a voice-based door access control system for building security. In the proposed system, the access may be authorized simply by means of an enrolled user speaking into a microphone attached to the system. The proposed system then will decide whether to accept or reject the user's identity claim or possibly to report insufficient confidence and request additional input before making the decision. Furthermore, intelligent system approach is used to develop authorized person models based on their voice. Particularly Adaptive-Network-based Fuzzy Inference Systems is used in the proposed system to identify the authorized and unauthorized people. Experimental result confirms the effectiveness of the proposed intelligent voice-based door access control system based on the false acceptance rate and false rejection rate.

**Keywords:** Access control, voice, intelligent, artificial neural network, security

---

### INTRODUCTION

The personal safety of the population in public and private building has always been a concern in current daily life. Access control for buildings represents an important tool for protecting both building occupants and the structure itself. One of the important security systems in building security is door access control. The door access control is a physical security that assures the security of a room or building by means limiting access to that room or building to specific people and by keeping records of such accesses.

The most widespread authentication method for such system is based on smartcards. Smartcard limits room or building access to only those people who hold an allocated smart card. However, there is difficulty to prevent another person from attaining and using a legitimate person's card. The conventional smartcard can be lost, duplicated, stolen, forgotten, or impersonated with accuracy. Due to the limitations of conventional security procedures, a range of biometric verification options are currently under consideration

for security system including door access control<sup>[1,2]</sup>. In the biometrics methods, the idea is to enable automatic verification of identity by computer assessment of one or more behavioral and/or physiological characteristics of a person. Recently biometrics methods used for personal authentication utilize such features as the face, the voice, the hand shape, the finger print and the iris<sup>[1,2]</sup>. Each method has its own advantages and disadvantages based on their usability and security<sup>[3]</sup>. Among the biometrics methods, voice has the high usability characteristics which include the simplicity for the user, feeling of resistance, speed of authentication and level of false-rejection rate<sup>[4]</sup>.

In order to overcome the problems of the smartcard-based door access control, this paper introduces an intelligent voice-based door access control system for building security. The proposed intelligent voice-based access control system is a performance biometric which offers an ability to provide positive verification of identity from an individual's voice characteristics to access secure locations (e.g. office, laboratory, home). In the

---

**Corresponding Author:** Wahyudi, Intelligent Mechatronics Systems Research Group, Department of Mechatronics Engineering, International Islamic University Malaysia, PO. Box, 10, 50728, Kuala Lumpur Malaysia

proposed system features are extracted from the person voice data and then an Adaptive-Network-based Fuzzy Inference Systems (ANFIS) is used to develop models of the authorized persons based on the feature extracted from the authorized person voices.

First, the prototype of the door access control is described. Next, the speaker verification process used in the proposed system is discussed in detail. Finally, the performance of the proposed intelligent voice-based door access control is evaluated experimentally for door control access in Intelligent Mechatronics System Laboratory, Faculty of Engineering, International Islamic University Malaysia.

### PROPOSED VOICE-BASED DOOR ACCESS CONTROL

**Proposed system description:** Figure 1 shows the schematic diagram of the proposed intelligent voice-based access control. The proposed system basically consists of three main components namely voice sensor, speaker verification system and door access control. A low-cost microphone commonly used in the computer system is used as voice sensor to record the person voice. The recorded voice is then sent to the voice-based verification system which will verify the authenticity of the person based on his/her voice.

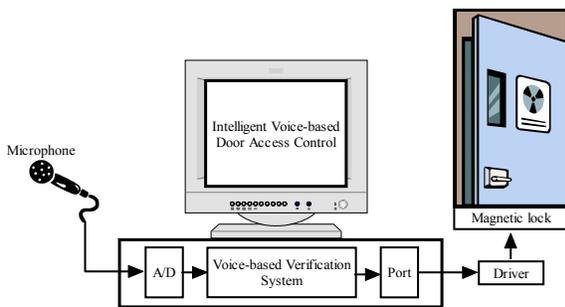


Fig. 1: Proposed intelligent voice-based door access control system

A personal computer (PC) of 1.5 MHz Pentium III processor equipped with sound card is used for speaker verification implementation. The sound card records the voice data based on the sampling frequency of 22 kHz. In this system, all of the voice data processing and speaker verification algorithms are implemented in the PC using MATLAB and its toolboxes. As a result of the voice-based verification, a decision signal which will accept or reject the access will be sent through the parallel port of the PC to the door access control.

As shown in Fig. 2, an electromagnetic lock is attached in the door for controlling the door opening

and closing. The electromagnetic lock works on 12 volts DC power supply and it is set in normally close (NC) condition. Therefore, without command signal from the verification system, the lock is always switched on and the door remains closed. In the case a person is verified by the proposed voice-based verification as an authorized user, the access is granted. The parallel port sends a signal to the electromagnetic lock driver, which is shown in Fig. 3, so that the electromagnetic lock is demagnetized. As a result the door can be opened by that authorized person for a certain period of time.



Fig. 2: Door with electromagnetic lock

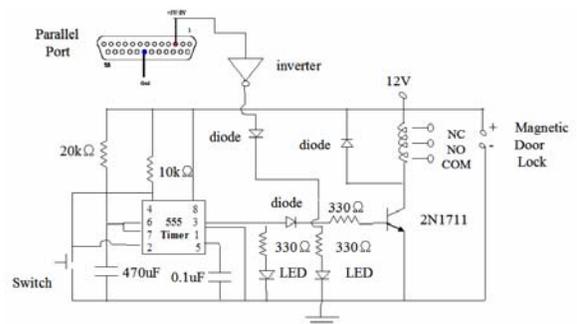


Fig. 3: Electromagnetic lock driver

The access control system in general makes four possible decisions; the authorized person is accepted, the authorized person is rejected, the unauthorized person (impostor) is accepted and the unauthorized person (impostor) is rejected. The accuracy of the access control system is then specified based on the rate in which the system makes decision to reject the authorized person and to accept the unauthorized person. The quantities to measure the rate of the access control accuracy to reject the authorized person is then called as false rejection rate (*FRR*) and that to measure

the rate of access control to accept the unauthorized person is called to as false acceptance rate (*FAR*). Mathematically, both rates are expressed as percentage using the following simple calculations<sup>[5]</sup>:

$$FRR = \frac{NFR}{NFA} \times 100\% \quad (1)$$

$$FAR = \frac{NFA}{NIA} \times 100\% \quad (2)$$

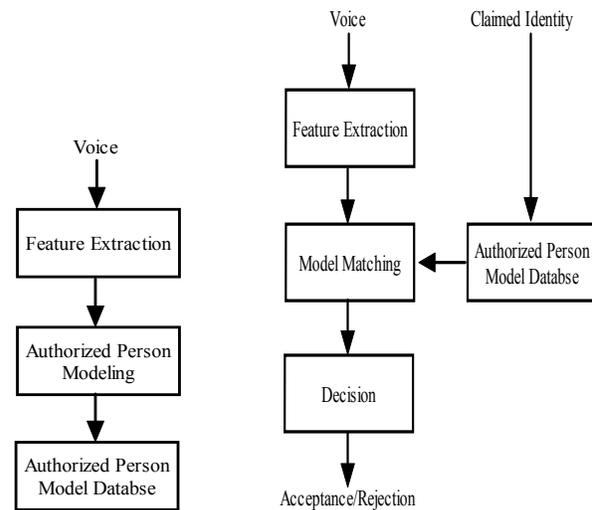
*NFR* and *NFA* are the numbers of false rejections and false acceptance respectively, while *NAA* and *NIA* are the number of the authorized person attempts and the numbers of impostor person attempts. For achieving high security of the door access control system, it is expected that the proposed system will have both low *FRR* and low *FAR*.

**Voice-based verification system:** It is well known that not only conveys a person message, voice of a person also indicates the person identity. Therefore, it can also be used in biometric system. The use of the voice for biometric measurement becomes more popular due to some reasons such as natural signal generation, convenient to process or distributed and applicable for remote access. Basically there are two kinds of voice-based recognition or speaker recognition. Speaker identification is one of the two form of speaker recognition, while speaker verification being the other one<sup>[5]</sup>. In the speaker verification system, the system decides that a person is the one who he/she claims to be. On the other hand, speaker identification decides the person among a group of persons. Speaker recognition is further divided into two categories, which are text-dependent and text-independent speaker recognitions. Text dependent speaker recognition recognizes the phrases that spoken, whereas in text-identification the speaker can utter any word.

The most appropriate method for voice-based door access control is based on the concept of speaker verification since the objective in the access control is to accept or reject a person to enter a specific building or room. Figure 4 shows the basic structure of the proposed voice-based verification system. As other methods of biometric-based security system, there are two phases in the proposed system. First phase is training or enrollment phase as shown in Fig. 4(a). In this phase the authorized persons are registered and their voices are recorded. The recorded voices are then extracted. The features extracted from the recorded voices are used the develop models of the authorized persons.

The second phase in the proposed system is testing or operational phase as depicted in Fig. 4(b). In this phase a person who wants to access the building/room

is required to enter the claimed identity and his/her voice. Furthermore, the entered voice is processed and compared with the claimed person model to verify his/her claim. In this phase, there is a decision process in which the system decides whether the feature extracted from the given voice matches with the model of the claimed person. In order to give a definite answer of access acceptance or rejection, a threshold is set. When degree of similarity between a given voice and the model is greater then threshold, the system will accept the access, otherwise the system will reject the person to access the building/room.



a. Training phase  
b. Testing (operational) phase  
Fig. 4: Basic structure of voice-based verification system

**Feature extraction:** As shown in Fig. 4, feature extraction is one of the important processes in the proposed system. Feature extraction is the process of converting the raw voice signal to feature vector which can be used for classification. Features are some quantities, which are extracted from preprocessed voice and can be used to represent the voice signal. In general, there are two types of feature extraction technique, namely; cepstral coefficient feature based and prosodic-based feature such as; fundamental frequency and formant frequency.

In this paper, Perceptual Linear Prediction (PLP) coefficients are used as feature in the proposed system. Figure 5 shows schematically extraction process of the PLP coefficient from the raw voice signal. Perceptual Linear Prediction (PLP), similar to LPC analysis, is based on the short-term spectrum of speech. In contrast to pure linear predictive analysis if speech, PLP modifies the short-term spectrum of the speech by several psychophysically based transformations. In this

method the spectrum is warped according to Bark scale. The PLP used an all-pole model to smooth the modified power spectrum. The output cepstral coefficients are then computed based on this model<sup>[6]</sup>.

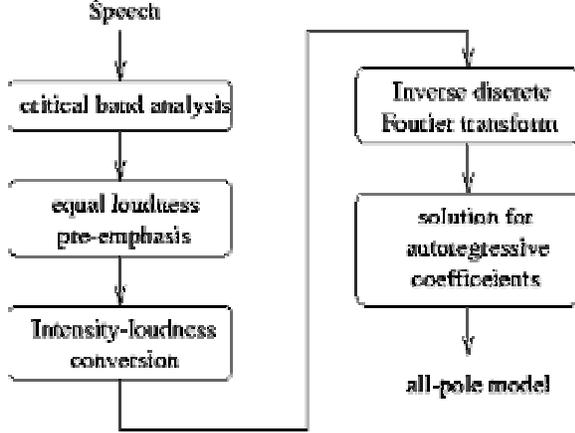


Fig. 5: PLP coefficient extraction process

In summary, the PLP coefficients are calculated based on the following steps:

**Critical band analysis:** Firstly, the voice data is framed and windowed (using available window function such as hamming window) and then it is transformed into frequency domain using the Fast Fourier Transform (FFT). Then, the obtained spectrum is warped from the radial frequency ( $\omega$ ) to the Bark frequency ( $\Omega$ ) scale using the following formula:

$$\Omega(\omega) = 6 \ln \left( \frac{\omega}{1200\pi} + \sqrt{1 + \left( \frac{\omega}{1200\pi} \right)^2} \right) \quad (3)$$

The power spectrum,  $S(\Omega)$  of the warped spectrum is then calculated. Next, the convolution operation is carried out between the warped power spectrum  $S(\Omega)$  and the power spectrum of a simulated critical-band masking curve  $\Psi(\Omega)$ , which has the following form;

$$\Psi(\Omega) = \begin{cases} 0, & \Omega < -1.3 \\ 10^{2.5(\Omega+0.5)}, & -1.3 \leq \Omega < -0.5 \\ 1, & -0.5 \leq \Omega < 0.5 \\ 10^{2.5(\Omega-0.5)}, & 0.5 \leq \Omega < 2.5 \\ 0, & \Omega > 2.5 \end{cases} \quad (4)$$

The following is the result of the convolution operation:

$$\Theta(\Omega_i) = \sum_{\Omega=-1.3}^{2.5} S(\Omega - \Omega_i) \Psi(\Omega), \quad i = 1, 2, \dots, B \quad (5)$$

where B is the number of the sample. The sampling intervals are chosen so that when the critical bands are added together it equally represents the frequency scale.

**Equal-loudness pre-emphasis:** Firstly, an equal-loudness curve is constructed. An approximation of this curve for the frequency up to 5 kHz is

$$E(\omega) = \frac{\omega^4 (\omega^2 + 56.8 \times 10^6)}{(\omega^2 + 6.3 \times 10^2)^2 (\omega^2 + 0.38 \times 10^9)} \quad (6)$$

And the following equation is the approximation for frequency higher than 5 kHz:

$$E(\omega) = \frac{\omega^4 (\omega^2 + 56.8 \times 10^6)}{(\omega^2 + 6.3 \times 10^2)^2 (\omega^2 + 0.38 \times 10^9) (\omega^6 + 9.58 \times 10^{26})} \quad (7)$$

Then, the simulated equal-loudness  $E(\omega)$  is used to pre-emphasize the sampled bark power spectrum  $\Theta(\Omega_i)$ . The following is obtained from the equal-loudness pre-emphasis:

$$L(\Omega(\omega)) = E(\omega) \Theta(\Omega(\omega)) \quad (8)$$

**Intensity-loudness power conversion:** Here, a cubic root compression of the amplitude is performed as follows:

$$\phi(\omega) = L(\Omega)^{1/3} \quad (9)$$

**Inverse discrete Fourier transform:** The power spectrum that resulted from previous amplitude compression is converted back to time domain using inverse FFT (IFFT).

**Calculation of the all-pole coefficients:** The autocorrelation signal can be used to calculate the all-pole coefficients using the well-known the Levinson-Durbin algorithm.

Detail discussion on the PLP coefficient can be found in<sup>[6]</sup>.

**ANFIS-based speaker model:** ANFIS proposed by Jang<sup>[7]</sup> is an architecture which functionally integrates the interpretability of a fuzzy inference system with adaptability of a neural network. Loosely speaking ANFIS is a method for tuning an existing rule base of fuzzy system with a learning algorithm based on a collection of training data found in artificial neural network. Due to the use the less tunable of parameters of fuzzy system compared with conventional artificial neural network, ANFIS is trained faster and more accurate than the conventional artificial neural network.

An ANFIS which corresponds to a Sugeno type fuzzy model of two inputs and single output is shown in Fig. 6. A rule set of first order Sugeno fuzzy system is the following form:

Rule  $i$ : If  $x$  is  $A_i$  and  $y$  is  $B_i$ , then  $f_i = p_i x + q_i y + r_i$ .

In the perspective of artificial neural network, it is a feedforward network consisting of 5 layers. Every node  $i$  in the first layer is an adaptive node with the following node function

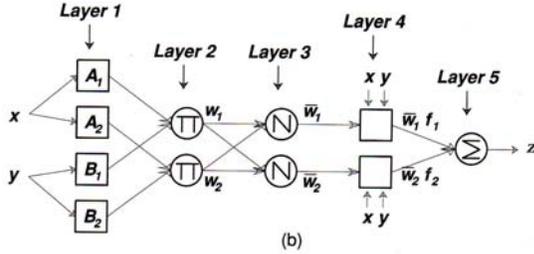


Fig. 6: ANFIS architecture<sup>[8]</sup>

$$O_{i,i} = \begin{cases} \mu_{A_i}(x), & i = 1,2 \\ \mu_{B_{i-2}}(y), & i = 3,3 \end{cases} \quad (10)$$

where  $x$  (or  $y$ ) is the input node  $i$  and  $A_i$  (or  $B_{i-2}$ ) is a linguistic label associated with this node. In other words,  $O_{i,i}$  is the membership degree of a fuzzy set  $A$  (or  $B$ ) to which the input  $x$  (or  $y$ ) is quantified. The membership function for  $A$  (or  $B$ ) can be Gaussian function, triangle membership function and others. The parameters of the membership function used in this layer are termed as premise parameters.

Second layer combines the output of the first layer so that it has the following output:

$$O_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(y) \quad (11)$$

Here each output represents the firing strength of a rule. Next layer, which is third layer, normalizes the output of the previous layer as follows;

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1,2 \quad (12)$$

In the fourth layer, the following output is calculated based on the third outputs:

$$O_{4,i} = \bar{w}_i f_i = w_i(p_i x + q_i y + r_i) \quad (13)$$

where  $f$  is function which is used in the first order Sugeno type fuzzy system. Parameters in this node ( $p_i$ ,  $q_i$  and  $r_i$ ) are referred as consequent parameters. Finally, the final output of the ANFIS is the last layer output and it is given as

$$O_{5,i} = z = \sum_i \bar{w}_i f_i \quad (14)$$

The main objective of the ANFIS design is to optimize the ANFIS parameters. There are two steps in the ANFIS design. First is design of the premise parameters and the other is consequent parameters training. There are several method proposed for designing the premise parameter such as grid partition, fuzzy c-means clustering and subtractive clustering<sup>[8]</sup>. Once the premise parameters are fixed, the consequent parameters are obtained based on the input-output

training data. A hybrid learning algorithm is a popular learning algorithm used to train the ANFIS for this purpose. In summary, the steps of building person model based on the voice data using ANFIS are as follows:

- \* Voice data collection and feature extraction of the voice data.
- \* Determining the premise parameters.
- \* Training of the ANFIS using the input pattern and desired output to obtain the consequent parameters.
- \* Validation of the trained ANFIS using training data.

## RESULTS

**Experimental setup:** In order to evaluate the effectiveness of the proposed intelligent voice-based door access control, the proposed system is installed at Intelligent Mechatronics System Laboratory, Faculty of Engineering, International Islamic University Malaysia. Voices of nine (9) speakers from YOHO database are used in the experiment. Three (3) speakers are considered as the authorized person to access the laboratory and the other six (6) speakers are assumed as outside impostors. Each speaker, who is assumed as authorized person, has to say word ‘seven’ for 70 times where 20 voice data are used as training data and the other 50 voice data are used as testing data. This means the text-dependent speaker verification system is used in the proposed system.

The example of raw voice signal of the word ‘seven’ for an authorized person is shown in Fig. 7. To obtain the PLP coefficients, the 17 critical-band filters are used, which covers a 17 Bark frequency range. These filters are simulated by integrating the FFT spectrum of 20-ms Hamming-windowed speech segments in which the frame rate is 10-ms. Figure 8 shows the 13 PLP coefficients extracted from the voice signal shown in Fig. 7.

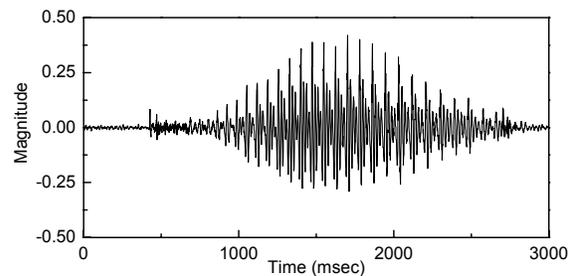


Fig. 7: Example of voice signal of authorized people

The effectiveness of the proposed method in training phase is based upon training time and the

classification rate. Moreover, the effectiveness of the proposed system in testing (operational) phase is evaluated based upon *FRR* and *FAR*. The *FAR* is

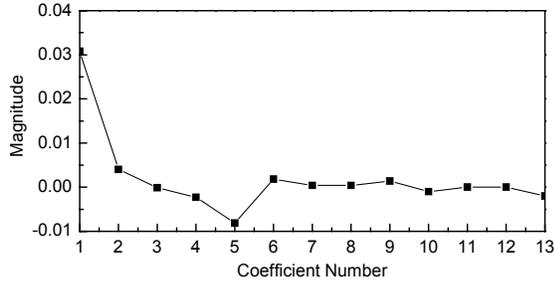


Fig. 8: Extracted 13 PLP coefficients

Table 1: Training performances

Model	Authorized Person	Training Time (sec)	Classification Rate (%)
ANFIS 1 ( $r = 0.25$ )	Person 1	51	100%
	Person 2	45	100%
	Person 3	46	100%
	Average	47	100%
ANFIS 2 ( $r = 0.50$ )	Person 1	25.8	100%
	Person 2	28.9	100%
	Person 3	27.8	100%
ANFIS 3 ( $r = 1.00$ )	Person 1	0.28	100%
	Person 2	0.27	100%
	Person 3	0.41	100%
Average	0.32	100%	

calculated based on the close set and open set. In the close test, the voice of an authorized person makes up the disguised voice to the other authorized person. On the other hand, test on the two students who are regarded as outside impostors constitute open set test.

**Training of the ANFIS-based speaker models:** The ANFIS-based speaker model is developed using Fuzzy Logic Toolbox of MATLAB. In order to allow the ANFIS learn from the input-output data available so that the consequent parameters are obtained, firstly the structure of the ANFIS has to be designed. Design of the ANFIS structure is done by determining premise parameters. Here the subtractive clustering method is used with different radius parameters. Once the premise parameters are obtained, the ANFIS model is trained by using hybrid learning algorithm for 10 iterations.

Table 1 shows the training time and the classification rate for all of the ANFIS-based speaker models for different subtractive clustering parameters. As shown in the table, all of the speaker models give perfect classification rates. There are no errors in identifying the authorized persons based on the voice data used in training phase. However, the training time is significantly different for different radius. A smaller

radius causes a longer training time. This is due to the fact that a smaller cluster radius will usually yield more, smaller clusters in the data and hence more rules. A more rules of ANFIS system result in a larger number of consequent parameters. As consequent, a longer time is needed in training process to optimize the parameters. Hence, it can be concluded a larger radius are preferable to shorten the training time.

Table 2: Testing performances of ANFIS 1, Radius of 0.25

Authorized Person	<i>FRR</i> (%)	<i>FAR</i> (%)	
		Close Set	Open Set
Person 1	16	5	6
Person 2	20	9	4
Person 3	14	11	7
Overall	16	8.3	5.7

Table 2: Testing performances of ANFIS 2, Radius of 0.50

Authorized Person	<i>FRR</i> (%)	<i>FAR</i> (%)	
		Close Set	Open Set
Person 1	18	5	6
Person 2	10	7	6
Person 3	14	8	7
Overall	14	6.7	6.3

Table 3: Testing performances of ANFIS 3, Radius of 1.00

Authorized Person	<i>FRR</i> (%)	<i>FAR</i> (%)	
		Close Set	Open Set
Person 1	36	2	4
Person 2	6	14	3
Person 3	16	11	10
Overall	19.3	9	5.6

**Testing of the ANFIS-based speaker models:** Tables 2-4 show the performances of the ANFIS models when testing voice data is used. In term of both *FAR* and *FRR*, ANFIS 2 produces a better performance than the other models. Hence it can be concluded that ANFIS 2 is the best candidate as voice-based model in the proposed system. From the security point of view, ANFIS 2 is the best model for protecting the laboratory from unauthorized person (impostors) since it gives the lowest *FAR*. The overall *FAR* of the ANFIS 2 is smaller than 10%, which is good enough for common security system. In the case high level of security is needed, further improvement has to be done so that the proposed system produces a small *FAR*, which is smaller than 1 %.

However although the *FRR* of the ANFIS 2 is also the smallest, its *FRR* is larger than 10 %. Although it does not influence the level of security, a quite large value of *FRR* makes the access control system inconvenient for the authorized person. Further improvement needs to be done to improve the level of usability of the ANFIS-based model for access control system.

## CONCLUSION

This study has documented development of intelligent voice-based door access control for building security. The proposed system adopted Perceptual Linear Prediction (PLP) coefficients as the feature of the person voice and used Adaptive-Network-based Fuzzy Inference Systems (ANFIS) to develop authorized person models based on their voices. Experimental results showed that the proposed system produced a good security performance, especially it gave a good false rejection rate (*FRR*) and a good false acceptance rate (*FAR*) of the close set condition. However, further study has to be done to improve its *FRR*.

## REFERENCES

1. Kung, S.Y., M.W. Mak and S.H. Lin, 2004. Biometric Authentication: Machine Learning Approach. Prentice Hall.
2. Zhang, D.D., 2000. Automated Biometrics: Technologies and Systems. Kluwer Academic Publisher.
3. Osadciw, L., P. Varshney and K. Veeramachaneni, 2002. Improving Personal Identification Accuracy Using Multisensor Fusion for Building Access Control Application. In Proceedings the Fifth Intl. Conf. Information Fusion, pp: 1176- 1183.
4. Anonymous, 2004. Door-access-control System Based on Finger-vein Authentication. Hitachi Review. Available online at <http://www.hitachi.com/rev/archive/2004>.
5. Campbell, J.P., 1997. Speaker Recognition: a Tutorial. In Proc. IEEE., pp: 1437-1462.
6. Hermansky, H., 1990. Perceptual linear predictive (PLP) analysis for speech. J. Acoustics Society American, pp: 1738-1752.
7. Jang, J.S., 1993. Adaptive-network-based fuzzy inference system. IEEE Trans. on System, Man and Cybernetics, pp: 665-685.
8. Jang, J.S.R., C. T. Sun and E. Mizutani, 1997. Neuro-Fuzzy and Soft Computing, Prentice Hall, Upper Saddle River, NJ, USA.